EFFICIENT OPTICAL NETWORK-ON-CHIP DESIGN

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF SCIENCE & ENGINEERING

2018

By
Sebastian Werner
School of Computer Science

Contents

Word Count: 52547

List of Tables

List of Figures

List of Abbreviations

CMOS Complementary Metal-Oxide-Semiconductor is the technology used to construct the integrated integrated circuits out of which processors, static random access memory, as well as analog circuits are created.

CMP Chip Multiprocessors are computing chips equipped with two or more processing units ('cores'), a design that can leverage parallelism to attain higher compute performance.

DOR Dimension-order Routing is a popular routing algorithm for network topologies with multiple dimensions (e.g. a mesh has two dimensions, X and Y). A packet is first routed to the correct position within a dimension before being sent to the next dimension.

DRAM Dynamic Random Access Memory is a widely used memory technology offering high density as a memory cell merely requires with one capacitor and transistor. Unlike static memory, DRAM is volatile which is why refresh cycles are necessary to keep the stored data alive.

DWDM Dense Wavelength-division Multiplexing denotes a multiplexing technique in the optical domain that leverages the ability of optics to transmit data on several wavelengths simultaneously within the same fiber/waveguide to improve link bandwidth.

Flit In networks-on-chip, large packets are typically divided into smaller pieces – Flits (FLow control unITs) – which are subsequently disseminated through the network, thereby allowing more efficient use of resources and higher throughput.

ITRS The International Roadmap for Semiconductors consists of multiple documents composed by a group of semiconductor industry experts that outline future research directions and predictions in the semiconductor industry.

NOC Networks-on-chip represent the interconnection between modules on a chip, such as processors, caches, or memory controllers, and consists of routers and links.

SIPS Silicon Photonics denote the technology of fabricating devices capable of guiding and manipulating light using existing semiconductor fabrication techniques, which allows the co-integration of electronic and photonic devices on the same chip.

TDM Time-division multiplexing is a method that allows multiple signals to utilise a common path where each signal transmits in a separate time slot.

TSV Through-silicon Vias are high performance electrical interconnects that vertically connect several stacked dies, thereby enabling 3D integrated circuits.

VLSI Very-large-scale Integration describes the procedure of integrating a very large number – today billions – of transistors into a single chip.

Abstract

Optical on-chip data transmission enabled by silicon photonics (SiPs) is considered a promising candidate for future on-chip communication as the high-bandwidth, low-latency, and relatively-distance-independent nature of photonics overcomes the technological limitations of the electrical interconnects currently used in Chip Multiprocessors (CMPs). Present SiP technologies, however, impose static power overheads that often eliminate their performance and dynamic power benefits. Consequently, many research efforts have been focusing on both the technology and the architectural level to solve this static power problem and to unleash the full potential of SiPs. This thesis proposes and evaluates novel architectural approaches to allow for a more power-efficient utilisation of optical links in networks-on-chip (NoCs).

First, it carries out a thorough review and **analysis of the state-of-the-art** optical NoC (ONoC) architectures and a comparison of current electrical and optical interconnect technologies in terms of latency and power consumption. Then, it proposes 'Amon', **a novel all-optical NoC** based on wavelength-selective routing that decreases static optical power by exhibiting a topology with lower path losses and fewer wavelength switches. Moreover, Amon features an improved destination-reservation mechanism and backend modifications to further improve performance and power efficiency. Afterwards, it introduces 'Lego', **a novel hybrid NoC** combining electrical and optical interconnects in an architecture in which high quantities of low-bandwidth optical links provide high bisection bandwidth with reduced power consumption due to lower overall optical losses. A distance-based routing mechanism ensures that optical links are used for large enough distances to hide their serialisation delay and an electrical NoC for local traffic otherwise. Finally, it presents **a novel subchannel scheduling scheme and arbitration mechanisms for shared optical buses** to improve bandwidth utilisation and power efficiency compared to the state of the art by scheduling contending nodes both in time slots and subchannels.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright

Acknowledgements

First and foremost, I would like to thank my supervisor Javier Navaridas Palma for his guidance and support in the last three years. My research has greatly benefited from his critical feedback, keen insights, and enthusiasm. Thank you for always having an open ear, discussing my ideas for countless hours, and giving me mental support when I needed it. By the same token, I would like to thank my co-supervisor Mikel Luján for always keeping me up-to-date with ongoing research and conferences, connecting me with the right people to get support, and giving me guidance when I needed it. Thank you both for teaching me how to do research and for giving me the opportunity to pursuit a PhD in the first place.

I would also like to thank Vasilis Pavlidis for taking his time to give me very helpful feedback on my first year report.

In general, I would like to thank everyone at the APT group for always having a positive attitude, trying to be helpful and supportive whenever possible, and making our group a great environment both inside and outside the lab. In particular, I would like to thank Yaman Cakmakci, Will Toms, Guillermo Callaghan, and Andrey Rodchenko for their support when dealing with the servers and simulation tools, John Mawer for helping with all kinds of miscellaneous issues, and Guillermo Callaghan, Athanasios Stratikopoulos, Richard Neill, Yaman Cakmakci, and Will Toms for their help reviewing this thesis.

Finally and most importantly, I thank my family, my girlfriend, and my friends back in home Germany for their love and support. I am truly blessed and grateful for having you in my life and would not be where I am today without you. Thank you for believing in me.

Chapter 1

Introduction

1.1 Motivation

While technology scaling provides designers with billions of transistors on a single chip, power consumption constraints prevent the continued scaling of single processor performance. In order to maintain performance scaling while coping with power constraints and design complexity, an industry-wide shift towards CMPs has occurred, with tens or hundreds of cores on a single chip. Processor designs are therefore increasingly turning into communication-centric systems in which the NoC is a decisive determinant of power and performance. In fact, the NoC can consume up to 30% of the total power budget in modern CMPs [PDB14]. As this problem is expected to exacerbate with increasing NoC sizes, many research efforts have investigated and proposed novel NoC architectures to improve power efficiency and scalability.

One of the main reasons for the NoC's high energy consumption is the inherent technological limitation of electrical interconnects to scale energy and delay at the same rates as transistors. These technology trends are illustrated in Figures 1.1 and 1.2, which outline the significant gaps between these two components for shrinking feature sizes. Although repeaters can speed up signals on electrical wires, this measure also increases power consumption. The latency of electrical interconnects is thus limited by the power budget and may prohibit further performance and power scaling of CMPs by increasing the number of cores. It is commonly expected that electrical interconnects alone will not be able to satisfy power and performance demands of future many-core systems [ON12]. Consequently, sooner or later, they have to be augmented, or even replaced, by more advanced interconnect technologies capable of delivering higher bandwidth within lower power budgets.

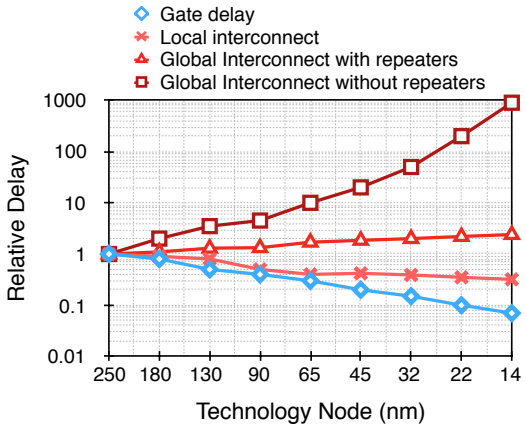


Figure 1.1: Delay Scaling [Com14]

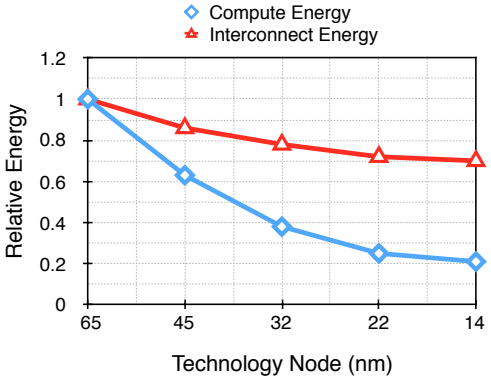


Figure 1.2: Energy Scaling [Bor13]

Breakthroughs in SiPs have made CMOS compatible components available that enable the integration of optical data transmission on the same chip as electronic components, thereby allowing optical links to be used for chip-to-chip, chip-to-DRAM, or on-chip communication. Computing systems could now leverage the benefits of transmitting data optically, which are high bandwidth density, signal propagation of light in silicon, and low-energy data transmission. In addition, as opposed to electrical interconnects, optical link length has a low impact on latency and energy as no additional circuitry is required to drive a signal [BCB+14]. All these properties make optical data transmission a promising candidate for future on-chip communication with increasing demands for scalability, bandwidth, and power efficiency.

However, significant challenges arise when implementing optical links in NoCs, both due to the immaturity of SiP devices and the inherent technological requirements of optics. Providing bandwidth to the NoC currently comes at the cost of considerable static power overheads. These overheads can become large enough to lead to inefficient designs that cancel out all the benefits of optical data transmission. Moreover, since chips work electrically, data conversion must take place from the optical to electrical domain and vice versa, which imposes additional circuitry, latency, and energy. In addition, NoC proposals must be practical, feasible, and should be benign to VLSI implementation with a clear concept of layout and packaging constraints.

Growing research efforts dedicated to optical NoCs (ONoCs) have been focusing on both the technology level and on the architectural level in recent years, with promising results showing that both fields are crucial for power-efficient ONoC designs and thus essential to merit its widespread commercial adoption. The thesis at hand contributes to the field of ONoC architectures by identifying benefits and shortcomings of the state

of the art and by exploring novel approaches to implement optical links in NoCs.
Generally, SiPs open up a whole new field of exciting opportunities in the on-chip interconnect domain. Exploring novel ideas on how to efficiently integrate optical links into the on-chip communication fabric is challenging, but the benefits can be significant. Recent NoC proposals that implement optical interconnects have shown large improvements in power efficiency compared to early studies in this field, and numerous research groups have been investigating new architectures based on improved or novel SiP devices.
This thesis proposes novel architectural approaches of ONoCs, which require a detailed understanding of the communication demands of CMPs, the benefits and drawbacks of both optical and conventional electrical interconnects, and what is feasible and practical. In addition, it considers the most recent advances in SiPs, their impact on ONoC designs, and opportunities for future architectural approaches.

1.2 Thesis Contributions

This thesis investigates novel architectural approaches to utilise optical interconnects in NoCs in more power-efficient ways. In particular, it makes the following contributions:

- Provides a review of all relevant NoC proposals since the advent of ONoCs and classifies them based on the taken approaches to implement optical links for on-chip communication. In doing so, it discusses their main findings, limitations, and pitfalls, identifies key research areas, and hints at possible future work.

- Improves the state of the art of wavelength-routed all-optical NoCs by proposing a novel topology 'Amon' that allows for fewer path losses and microrings, and in turn reduced static power. In addition, novel modifications of the destination-reservation mechanism performed on the control network offer improved latency and throughput. Compared to state-of-the-art proposals, Amon reduces power consumption by 21% and the destination-reservation mechanism improves latency by 75% on realistic workloads. In addition, an evaluation of decreasing the number of injection channels into the NoC (i) as well as increasing the number of ejection channels (ii) is conducted in terms of power and performance. Results show that these approaches halve static power (i) and eliminate the susceptibility of ONoCs to traffic hotspots without noticeable area overheads (ii).

- Proposes a novel approach to combine electrical and optical links in a NoC topology to reduce power consumption while achieving the same performance goals. Studying the relation between laser power and link bandwidth on optical buses reveals an exponential relationship between these two metrics, mainly caused by SiP device losses. Implementing a higher quantity of low-bandwidth optical links in a topology provides similar bisection bandwidth at lower power consumption. To mitigate the serialisation delay imposed by low-bandwidth optical links, a distance-based routing approach is proposed in which optical links are only used for larger distances, and an electrical network is used for local traffic. This approach can achieve up to 3.25$\times$ higher power efficiency on synthetic traffic compared to alternative approaches while imposing insignificant latency overheads on realistic workloads. Area overheads can be up to 80% compared to a baseline electrical mesh, but does not seem to impede an efficient layout or design feasibility.

- Proposes a novel bandwidth sharing mechanism that offers higher throughput on shared optical buses – a key building block for ONoCs – without incurring noticeable power overheads. Rather than scheduling requesting nodes sequentially on the bus, the possibility to tune microrings individually is leveraged to allow multiple requesters to transmit data on the bus both in parallel and sequentially by utilising time slots *and* subchannels. Both a centralised and distributed arbitration mechanisms for subchannel scheduling is developed. Although increasing complexity of the arbitration mechanism, subchannel scheduling more than doubles throughput compared to the state-of-the-art sequential scheme LumiNOC without power overheads. Merely packet latency for low injection rates is increased (10-30% depending on bus bandwidth).

- Evaluates the design options of performing bus arbitration on the same bus as data transmission by re-using the transmission medium versus performing bus arbitration on a separate control bus. Simulation results suggest that, although a parallel bus introduces additional resource overheads, the throughput gains of performing bus arbitration in parallel to data transmission outweigh these overheads significantly. In fact, power efficiency is doubled when a parallel control bus is used in combination with subchannel scheduling.

1.3 Publications

Much of the work in this thesis has appeared (or will appear) in the following publications:

- S. Werner, J. Navaridas, and M. Luján, *Amon: An Advanced Mesh-like Optical NoC*, IEEE 23rd Annual Symposium on High-Performance Interconnects (HOTI), 2015

- S. Werner, J. Navaridas, and M. Luján, *A Survey on Design Approaches to Circumvent Permanent Faults in Networks-on-Chip*, ACM Computing Surveys (CSUR), 2016

- S. Werner, J. Navaridas, and M. Luján, *A Survey on Optical Network-on-Chip Architectures*, To appear in: ACM Computing Surveys (CSUR), 2017.

- S. Werner, J. Navaridas, and M. Luján, *Designing Low-Power, Low-Latency Networks-on-Chip by Optimally Combining Electrical and Optical Links*, IEEE 23rd International Symposium on High-Performance Computer Architecture (HPCA), 2017.

- S. Werner, J. Navaridas, and M. Luján, *Subchannel Scheduling for Shared Optical On-chip Buses*, To appear in: IEEE 25th Annual Symposium on High-Performance Interconnects (HOTI), 2017.

In addition, based on the work presented in this thesis, following papers are in progress to be submitted as extensions to the publications listed above:

- S. Werner, J. Navaridas, and M. Luján, *Advanced Backend Modifications and Destination-reservation Mechanisms to Improve Power-efficiency in Wavelength-routed Optical NoCs*, Planned for submission to: OSA Journal of Optical Communications and Networking (JOCN).

- S. Werner, J. Navaridas, and M. Luján, *Assessing Parallel Bus Arbitration for Shared Optical Buses with Subchannel Scheduling*, Planned for submission to: OSA Journal of Optical Communications and Networking (JOCN).

1.4 Outline

Chapter 2 provides a technology review of SiP components and discusses basic optical buses – the backbone of any higher-order NoC topology – along with a discussion on design challenges.
Chapter 3 reviews the state of the art of ONoC proposals in the literature. In particular, it studies all-optical NoCs (i.e. NoCs communicating optically only), hybrid NoCs that combine electrical and optical links in the topology, and bandwidth sharing techniques that aim to maximise bandwidth utilisation. Besides, it provides a discussion of available simulation platforms and modelling tools, and a brief overview of other important active research areas in the domain of ONoCs.
Chapter 4 discusses Amon, the contribution to the realm of all-optical NoC architectures along with its topology, routing algorithm, switch design, and laser power distribution network. Subsequently, it presents the proposed destination-reservation mechanisms and backend modifications.
Chapter 5 evaluates a novel distance-based approach to combine electrical and optical links in a NoC topology, which advocates using low-bandwidth optical links for low-power long-distance communication, and electrical links for local communication to outbalance the serialisation overheads of the low-bandwidth optical links. The novel topology 'Lego' demonstrates the efficiency of this approach.
Chapter 6 analyses the suitability of shared optical buses as on-chip interconnects for current SiP device technologies and future projections, followed by a description of subchannel scheduling and the arbitration techniques implementing it. Subsequently, it provides a discussion on performing bus arbitration on the same bus as data transmission versus on a parallel control bus, as well as a study evaluating the bus proposals within a realistic NoC.
Finally, Chapter 7 concludes the thesis contributions and outlines opportunities for future work.

Chapter 2

Silicon Photonics: Technology Review

2.1 Introduction

The pace at which SiP devices and materials have been evolving – and the various scales at which they can be deployed – has sparked much excitement in the scientific community; however, these attributes also make it increasingly difficult to keep up with the state of the art of this technology and to identify which devices are the most suitable and promising at which scale. For instance, while some devices may be ideal for chip-to-chip communication, they may be unsuitable for on-chip communication where design constraints differ. This section discusses the SiP devices that are currently considered the most suitable for the on-chip domain and explains the concept of optical on-chip data transmission along with optical buses – the backbone of higher-order ONoC topologies. Subsequently, it summarises design challenges of implementing ONoCs, and analyses how they compare to electrical interconnects.

2.2 Devices

SiP devices enable the integration of optical data transmission on-chip. This thesis focuses on optical NoCs based on *microring resonators* (MR)[BDHVV$^+$12], which are currently the SiP devices that offer the highest bandwidth density and energy efficiency [BRNB16]. Additionally, their compact footprint (ring radii can be as small as 2-3 μm [NFA11] [RBP$^+$17]) allows for large-scale integration. Aside from MRs, other SiP devices to guide and manipulate light have also been successfully demonstrated – most notably Mach-Zehnder Interferometers (MZIs) [LLR$^+$07], broadband ring resonators (BRR) [BCB$^+$14], directly modulated vertical-cavity surface-emitting lasers

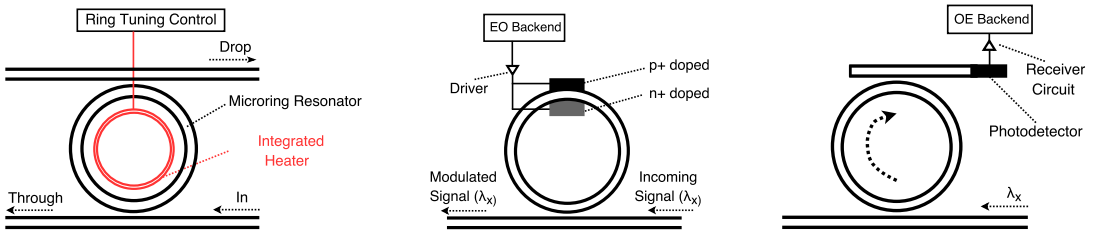
(VCSELs) [WML+13], or arrayed waveguide grating routers (AWGRs) [KII+03] – but are currently considered less suitable for intra-die, on-chip communication due to several limitations, such as impractical area footprint (AWGR [GPAY17]), incompatibility with wavelength-division multiplexing (VCSELs [HB14]), long reconfiguration times (BRR [HJH14]), or lack of wavelength selectivity (MZIs [RBP+17]) – all key properties that will be explained in more detail in the following.
MRs form the basis of modulators, switches, and wavelength-selective filters, which are the key components to encode/decode optical data and guide light across the chip. Each MR is designed and dimensioned to respond to one particular wavelength, referred to as *resonance wavelength*. MRs are susceptible to manufacturing mismatches and temperature variations, which can shift their resonance wavelength and make them non-functional. To prevent that, each MR is equipped with an integrated heater that controls its resonance wavelength by changing the ambient temperature [GLM+11].

2.2.1 Wavelength-selective Filters

Figure 2.1a depicts an example layout of a MR (as widely found in the scientific literature [BCB+14]), which is typically placed next to waveguides – an optical wire carrying optical signals. Light travels through the waveguide from the 'In' port to the 'Through' port. If the MR is *in resonance* with the wavelength of the optical signal, it filters the signal and 'drops' it to the 'Drop' port. If *off resonance* with the optical signal's wavelength, the signal will simply pass the MR through the 'Through' port without getting filtered. Figure 2.2 illustrates how MRs can function as switches to steer optical signals from one waveguide to another. It is also possible to drop multiple wavelengths by implementing multiple MR filters (*filter banks*) for switching, as shown in Figure 2.3, where each filter responds to a different wavelength. This functionality allows to steer optical signals through an optical network and forms the basis of any MR based switching topology.

2.2.2 Electrical to Optical Data Conversion

Modulators perform the electrical-to-optical (EO) data conversion by encoding electrical bits onto optical signals. Figure 2.1b illustrates a simplified layout of a MR-based modulator. An optical signal on wavelength λ_x entering from the 'In' port is first captured by the MR (if in resonance). Then, bits are typically modulated onto the optical signal by using simple ON/OFF keying, which was shown to enable modulation



(a) Basic MR layout with integrated heater

(b) MR modulator

(c) MR filter at the receiver

Figure 2.1: Microring Resonator: Utilisation Scenarios

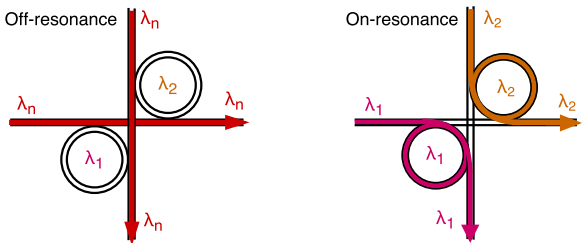


Figure 2.2: Wavelength switch: optical signals in resonance with the MR filters are captured and dropped while off-resonance signals remain unaffected.

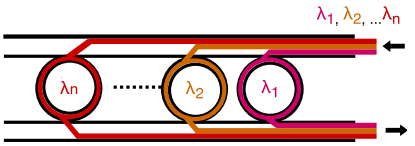


Figure 2.3: MR filter bank for switching multiple wavelengths

speeds of 10 Gb/s or higher at low energy consumption [DLF+09]. Although more advanced modulation techniques exist (e.g. Quadrature Amplitude Modulation), they require more complex transmitter and receiver circuitries and are therefore often regarded as less suitable for the on-chip domain [GPAY17]. Modulation is performed by a P-I-N diode that injects/depletes charge carriers to shift the MR's resonance wavelengths in/out, thus enabling ON/OFF keying [LBCB10]. Finally, the modulated signal will then exit the MR through its 'Through port'.

Electrical data (i.e. bits) is typically first encoded and conditioned (e.g. for error correction purposes) prior to modulation [BCB+14]. Serialisation circuitry might be necessary if the core frequency and modulation speed differ. Common modulation data rates are 10 Gb/s, and MRs enabling up to 40 Gb/s have been demonstrated [GTER11]. Core frequencies in CMPs or NoC routers, on the other hand, are normally lower to achieve higher power efficiency (typically less than 5 GHz). In such cases, a serialiser is used to up-convert the data rate (e.g. by combining multiple input wires) [BCB+14]. Finally, a driver circuit controls the SiP modulator as it typically has different electrical requirements than the CMOS circuitry of the digital logic [BCB+14].

2.2.3 Optical to Electrical Data Conversion

MRs are also deployed at the receiver side. In order to receive data, an optical signal must be fed into a photodetector which converts photons into electrical currents, thus performing optical-to-electrical (OE) data conversion [BCB+14]. Photodetectors typically respond to a wide light spectrum and are not wavelength-selective. Therefore, in order to receive data modulated on a certain wavelength, photodetectors are placed on a waveguide at the Drop port of a MR that selectively filters only the desired wavelength (see Figure 2.1c). Current photodetector technologies output electrical currents that are below the level necessary to drive voltages for operating digital logic, which is why amplifiers are needed [BCB+14]; however, Heck et al. [HB14] suggest that transistor capacitances might be small enough to be directly driven by a photodetector for technologies below 22 nm, which would eliminate the need for amplifiers. The following steps of deserialisation and decoding mirror the functionality of the serialiser and encoder at the sender side, respectively. Although necessary, these conversion steps – both at the sender and receiver side – do not introduce considerable latency (tens of picoseconds [KH12]). Larger serialisation degrees, however, have an impact on energy consumption of the serialisation circuitry in the electrical backends, particularly if link utilisation is high [GPAY17], and should thus be considered carefully.

2.2.4 A Basic Optical Link

Figure 2.4 illustrates the SiP building blocks necessary to perform optical on-chip data transmission between a sender and receiver [SCK+12]. Figure 2.5 summarises the required steps to generate and receive an optical signal as described by Bergman et al. [BCB+14]. A laser source, typically off-chip for current technologies, emits light at different wavelengths ($\lambda_1..\lambda_n$) which is guided into the chip in an optical fiber and coupled into the on-chip waveguide. A waveguide can accommodate multiple wavelengths (with an estimated practical limit of $<64\lambda$ within one waveguide for current technologies [PSDLL11]), which allows to transmit data in parallel on different wavelengths on the same waveguide. This property is referred to as dense wavelength-division multiplexing (DWDM), and is the reason for the superior bandwidth density of optical links. Since each MR responds to one particular wavelength, n modulators are required to transmit data on n wavelengths, often referred to as *modulator bank* [BCB+14]. The modulated wavelengths traverse the waveguide until they are extracted at the receiver, which requires a filter bank of n wavelength-selective filters that guide the wavelengths

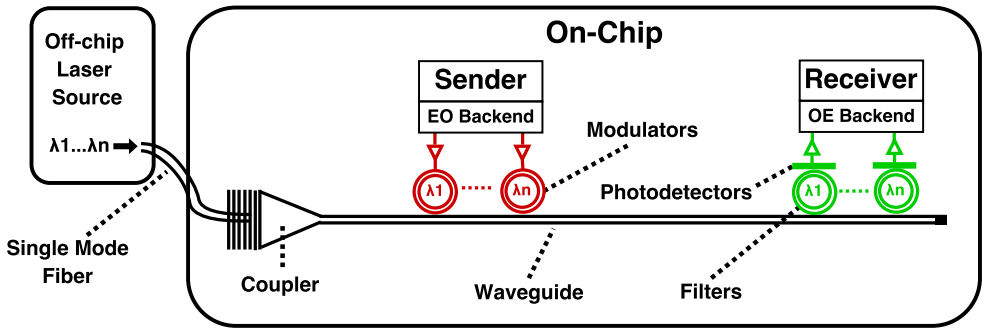


Figure 2.4: A Basic Optical Link for Data Transmission [SCK+12]

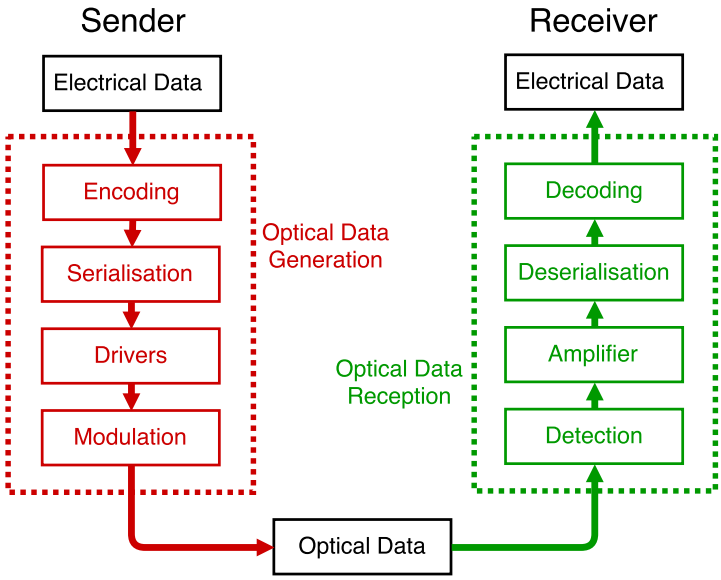


Figure 2.5: Optical Transmission Steps [BCB+14]

to a photodetector for OE data conversion each. Link bandwidth is determined by the number of wavelengths on the link and the modulation rate. For instance, 64λ and 10 Gb/s modulators/detectors achieve a link bandwidth of 640 Gb/s (or 128 bits/cycle at a core frequency of 5 GHz).

2.3 Optical Buses

While Figure 2.4 illustrates a simple optical link with one sender-receiver pair, modern CMPs with core counts in the tens or hundreds require more sophisticated communication infrastructures, such as crossbars, buses, or NoCs. Luckily, being able to accommodate a number of different wavelengths on the same waveguide provides a number of intriguing design opportunities.

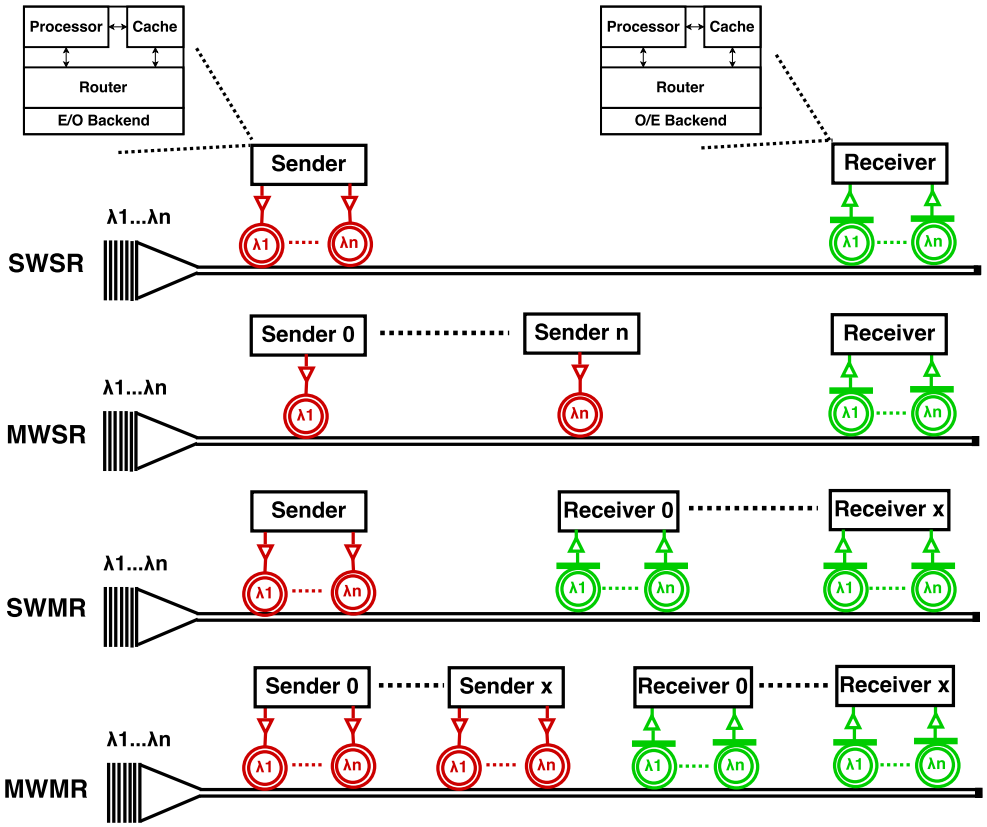


Figure 2.6: Basic Optical Bus Architectures [BCB+14]

2.3.1 Basic Optical Buses

Figure 2.6 depicts the basic optical bus architectures [BCB+14], each of them entailing different benefits and trade-offs. The Single-Writer-Single-Reader (SWSR) bus is a basic optical link between a sender and a receiver on which, as described above, the sender modulates its data on n wavelengths in parallel using DWDM.

On a Multiple-Writer-Single-Reader (MWSR) bus, each sender modulates its data on a dedicated, non-overlapping subset of wavelengths, allowing multiple senders to transmit data simultaneously to a receiver on the same waveguide. With n wavelengths provided by the laser source, up to n senders could send data to the receiver, each on its own wavelength. In a contention-free crossbar consisting of MWSR buses, each receiver would have a designated waveguide.

The Single-Writer-Multiple-Reader (SWMR) bus allows one sender to broadcast data on the entire optical bandwidth to all senders attached to the bus simultaneously. This provides a compact and efficient broadcast network without the requirement to split the bandwidth between multiple senders (as in the MWSR bus).

2.3.2 Control Network Assisted Optical Buses

It is widely accepted that implementing an entire NoC with the previously discussed bus designs only would require a high number of optical links which is impractical and leads to high power consumption. SWSR buses merely enable point-to-point connections. Implementing topologies with SWSRs buses would require $(N-1) \times N$ buses and in turn high laser power, which Section 2.4 will discuss in more detail. The main drawback of MWSR buses is that the available bandwidth is split between the senders which reduces the bandwidth-per-sender, or a considerable number of wavelengths has to be provided to offer the same bandwidth as in an SWSR bus. However, as discussed earlier, the number of wavelengths within a waveguide for current technologies has practical limits ($\sim$64), and high quantities of wavelengths within a single waveguide increases laser power significantly due to high optical losses. SWMR buses improve power efficiency by allowing efficient data communication to multiple receivers simultaneously, without having to split bandwidth between the senders (as there is only one); however, in SWMR buses, the laser source has to drive all receivers of the bus simultaneously at all times, thereby requiring more output power to drive all photodetectors (see next section for more details).

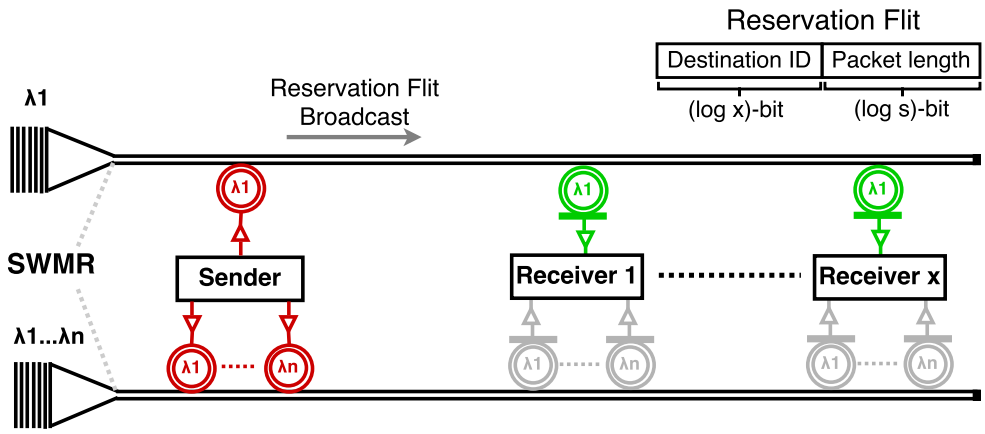
Two previously proposed designs tackle these problems, namely the Multiple-Writer-Multiple-Reader (MWMR) bus [LBGP14] [BP14], and the reservation-assisted SWMR (R-SWMR) bus [PKK$^+$09]. Both proposals take advantage of the possibility of turning on and off MRs by using the integrated heaters to shift the MR's resonance levels. In the MWMR (see Figure 2.6), also referred to as 'shared optical bus', multiple senders and receivers are connected to the same waveguide and have MRs to send and receive on the entire available optical bandwidth. As simultaneously transmitting nodes would overwrite each others' data, bus arbitration has to be performed prior to data transmission, just as in traditional electrical on-chip buses. Bus arbitration can be performed in an either distributed or centralised fashion, either on the same bus on which data transmission takes place [LBGP14], or on a separate bus [KSKK15], and can be performed either electrically or optically. All of these design choices come along with design benefits and trade-offs, which is currently an active research area for ONoCs. After arbitration, only the nodes that take part in the communication tune in their MRs, while all other nodes detune theirs to prevent interfering with the data transmission. The performance results of Li et al. [LBGP14] suggest that the latency overheads of this time-division-multiplexing (TDM) approach are acceptable in the on-chip domain. In addition, the power savings are large since 1) wavelength sharing

allows for fewer total wavelengths in a NoC and in turn laser power and 2) the number of receivers the laser has to drive is always one [LBGP14].
The R-SWMR bus addresses the problem of high laser power in SWMR links introduced by the large number of receivers that the laser source has to drive. The ideal number of receivers for minimal laser power is one. To ensure that this is the case at all times, the R-SWMR is supplemented with a separate, low-bandwidth SWMR bus on which the sender broadcasts a reservation packet prior to data transmission to inform all connected nodes about the prospective destination. Figures 2.7a and 2.7b show the R-SWMR bus during the destination reservation and data transmission phase, respectively. The reservation packet contains the destination address to notify the nodes about the next destination and a packet length indicator from which destinations can extract the duration of the data transmission. Initially, all nodes have their MR filters detuned on the data bus. Upon reception of the reservation packet, the destination tunes in its MR filters while all other nodes keep theirs detuned. After data reception, the destination detunes its MRs again. Thanks to this mechanism, the laser power on the data bus is significantly reduced as only one receiver is driven by the laser source at any time.
The latency and power overheads of using a separate SWMR bus prior to data transmission are small. Only very little bandwidth is required on the reservation bus to enable reservation of destinations at minimum latency since the reservation packets are small: $(log\ x)$ bits are required to decode the destination ID, and $(log\ s)$ bits for encoding the packet lengths, where x denotes the number of destinations on the bus and s the number of packets lengths supported by the NoC (which is typically low [1]). Most ONoCs assume 10 Gb/s modulation speeds and 5 GHz core clock rate, which would lead to $(log\ x + log\ s)/2$ wavelengths required to modulate the reservation packet in just one core clock cycle.

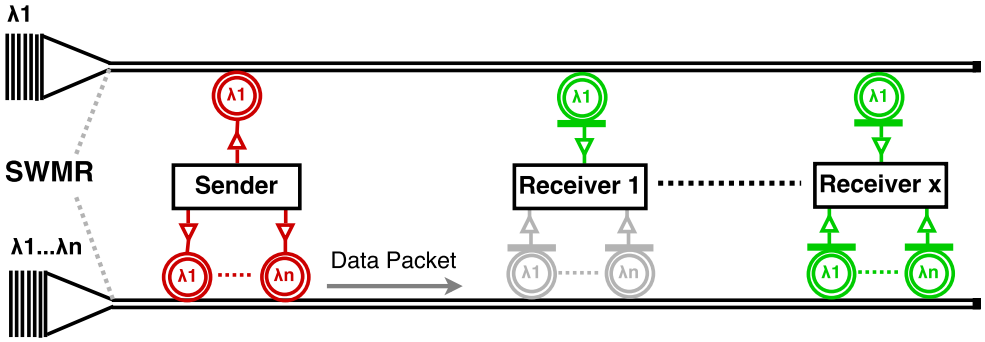
2.4 Design Challenges and Technological Implications

A shift away from electrical to optical interconnects requires the latter to outperform the former significantly to justify the costs and risks typically associated with adopting a new technology. Unfortunately, aside from all its benefits, ONoCs pose a number of critical design challenges that can have a high impact on their power efficiency.

[1] most CPU architectures have two packet sizes, one for cache line transfers and one for control and coherence traffic [LBGP14] [HGK10]



(a) Reservation-assisted SWMR bus during destination reservation phase.



(b) Reservation-assisted SWMR bus during data transmission phase.

Figure 2.7: Reservation-assisted SWMR bus with x receivers and s number of supported packet lengths. The optical bandwidth required on the reservation bus to allow minimal, one-cycle data modulation of the reservation packet is (log x + log s)/2, assuming modulation speeds of twice the core data rate (e.g. 5 GHz and 10 Gb/s).

Therefore, numerous research groups have been addressing these issues both on the technology and architectural level.

In order to design NoCs that utilise SiPs efficiently, a detailed knowledge of the power requirements of SiP devices and their performance metrics is essential. This section outlines the benefits and trade-offs of optical on-chip data transmission that designers must consider to make efficient use of its high-bandwidth capabilities. As conventional electrical interconnects are the competing technology, we compare optical to electrical links throughout this section.

2.4.1 Power Consumption

Electrical wires consume both static (leakage) and dynamic power. Although leakage power has become increasingly important due to the high integration densities of shrinking technology nodes [PDB14], increasing bandwidth demands, link lengths, and network sizes mean dynamic power is still a major contributor to the total power consumption [SCK+12]. Optical links, on the other hand, are static power dominated. Therefore, once a path is established and static power paid for, data transmission is very low energy and relatively-distance-independent [BCB+14].
The static power required at the laser source and for MR heating consumes the vast majority in ONoCs, and can hurt the power efficiency of ONoCs significantly, especially for applications that exhibit low NoC utilisation. A detailed analysis of what impacts laser and MR heating power and how it can be kept at a minimum are thus essential to design power-efficient ONoCs.

MR Heating Power

As discussed earlier, MRs respond to one particular wavelength based on their geometry and ambient temperature. MR heating (or 'tuning' / 'trimming') is required to mitigate temperature variations and post-manufacturing geometric mismatches, which can cause the resonant wavelength of MRs to shift to incorrect levels. Integrated heaters can shift/control the MR's resonant wavelength 'towards the red' through heating or 'towards the blue' through current injection [NFA11]. Since MRs form the basis of most ONoCs, appropriate tuning is necessary to ensure correct network functionality. This section reviews state-of-the-art MR tuning techniques and its impact on recently proposed ONoC architectures.
Assumptions with regard to MR heating power vary significantly across studies in the scientific literature, partly due to varying assumptions of the utilised MR tuning technique and assumptions on temperature variations. The vast majority of recent ONoC proposals, however, estimates heating power by multiplying a fixed assumed heating power per MR by the total number of MRs in the NoC (i.a. [PKM10] [KH12] [HJH14]). These studies assume a temperature range of 20 K and 1 μW/K tuning power per MR, i.e. in total 20 μW/MR. Other studies assume 16 μW/K per MR with a temperature range of 10 K [CAJ15]. To reduce MR heating, instead of shifting a MR's resonant wavelength to its original wavelength channel, Georgas et al. [GLM+11] propose to

shift it merely to the next closest channel and perform bit-reordering to maintain correct functionality. Athermal MR devices capable of maintaining correct functionality in the face of temperature variations (e.g. with cladding materials) have also been demonstrated [GCL13], most recently even with CMOS-compatible fabrication processes [FSB+15]. Although currently exhibiting a fairly large area footprint of 25 μm radius [FSB+15] (vs. $\sim$3 μm radius for regular MRs [NFA11]), advances in these devices are very exciting as they eliminate the need for MR heating altogether.

Nitta et al. [NFA11] demonstrated that tuning using current injection can easily lead to instabilities and thermal runaways. Utilising heating only is thus often considered the more practical approach [DH15a] and has been assumed in previous studies [JBK+09]; however, it has also been acknowledged that, without a mechanism to tune MRs back towards the blue, MRs must be designed to operate at temperatures higher than the heat dissipated from the electronic layer could ever lift them [NFA11]. Although that could lead to designs in which MRs must be constantly heated above the chip's temperature, rather simple techniques were shown to reduce the total heating power significantly: for instance, Parka [DH15a] proposed placing an insulation layer in a 3D stacked chip between the electrical and photonic die to isolate the MRs from the heat dissipated on the electrical layer. This approach shows highly promising results as it can lower the heating power by $\sim$4$\times$ and $\sim$5$\times$ for two different die cooling techniques.

Note that, although the vast majority of recent publications on ONoCs assume a fixed value for heating per MR (in particular 20 μW/MR), Nitta et al. [NFA11] revealed that accurately determining MR heating power is actually more complex than assuming a perfectly linear relationship between MR count and heating power, and is also significantly impacted by other factors, such as die area, ambient temperature of the chip, and the rate at which heat can be transferred outside the chip. In the absence of fully-integrated power/thermal simulation environments, however, assuming a fixed per-MR tuning power is considered to provide reasonable estimates [NFA11]. Besides, note that technological aspects, such as the thermal conductivity of the material surrounding the MR and its thermal tuning coefficient, influence MR tuning power, too. Nevertheless, since each MR requires heating, it is reasonable to assume that there will always be a relationship between the MR count of the NoC and heating power. ONoC designs should thus aim to keep the number of MRs as low as possible. Sun et al. [S+15] provide more details on the technology level of MR heating techniques.

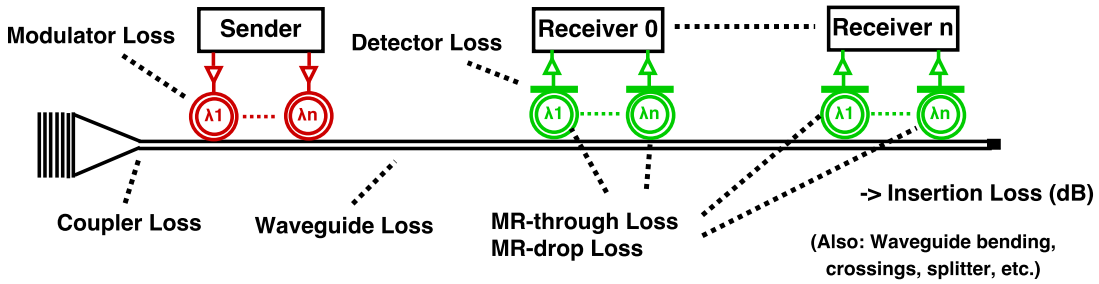


Figure 2.8: Optical Path Losses on a SWMR Bus

Laser Power

Numerous factors have a direct impact on the power required at the laser source (P_{laser}), which can be calculated with Equation 2.1 [LBGP14].

$$P_{laser} = N_{wv} \times L_e \times P_{sense} \times 10^{IL_{max}/10} \tag{2.1}$$

N_{wv} denotes the number of wavelengths that the laser source must provide. L_e represents the wall-plug efficiency of the laser, i.e. the ratio between the electrical input power and the optical output power of a laser. P_{sense} signifies the optical power required at the photodetector to correctly detect photons and convert them into electrons. Current devices exhibit values between 8 μW and 20 μW [ZPL+11, MNM+12]. If a laser source has to drive more than one photodetector on a link, the laser power required at each photodetector is added up [SCK+12]. As P_{sense} and L_e are technology dependent, novel architectures cannot reduce the impact of these metrics. Novel NoC designs, however, can improve the total optical path losses (IL_{max}) and N_{wv} significantly. In particular, IL_{max} and the employed laser technology deserve particular consideration since a detailed knowledge of them is essential to implement power-efficient and technologically practical NoCs.

Optical Path Losses Optical losses on the path from the laser source to the photodetector degrade the optical signal and require the laser to provide sufficient output power to mitigate these losses and to drive the photodetector at satisfactory bit error rates. Laser power must thus be provided based on the path that causes the highest insertion loss (i.e. IL_{max}) . Figure 2.8 depicts the loss incurred by different SiP devices on an optical signal on a SWSR bus. Coupling losses degrade the optical signal when it is coupled from an off-chip fiber into an on-chip waveguide. Losses are also incurred

at the modulator, for traversing the waveguide, for passing through a MR without getting dropped ('MR-through' loss), for getting dropped by a MR ('MR-drop' loss), and at the photodetector.

As described in the previous section, receivers require MR filters to filter optical signals and drop them onto the waveguide leading to the photodetector. Dropping a wavelength introduces significant path losses and should be considered carefully. While indispensable at the receiver side, the amount of dropping a wavelength due to switching in the network fabric may be reduced by smart network designs. MR-through losses are much lower per MR than MR-drop losses for current technologies (50-100× [OMS+12, GMS+14, LSZP14]); however, they can become significant in bus architectures that require placing a large number of MRs adjacent to a waveguide. For instance, considering the SWMR bus in Figure 2.8, for wavelength λ_n to reach *receiver x*, it has to pass $(n \times x)$ MR filters. The number of wavelengths and receivers thus considerably contributes to the total path losses in this case, and designers must be aware of the devices on optical paths and their impact on IL_{max}.

Losses are also introduced by SiP devices when guiding optical signals through the network. *Waveguide propagation* losses denote the losses per mm, which are exacerbated with increasing core counts and die sizes. *Waveguide bending* may be required in the physical layout to route waveguides across the chip. *Waveguide crossings* are difficult to avoid and must also be carefully considered, although recent research demonstrates significant technological improvements [LSZP14]. *Optical splitters* are used to distribute optical signals over a number of waveguides. Splitter losses – although low in absolute value – can accumulate and become increasingly critical in NoCs that require high splitting degrees, e.g. when a large number of links exist, but only few laser sources can be coupled into the chip due to packaging constraints. Aside from optical device losses, *non-linear effects* in optics also contribute to the total loss and increase along with the amount of optical power injected into a waveguide [LBGP14].

There seems to be little consensus between the publications in the scientific literature since the assumed loss values of the SiP devices vary significantly across different studies and are often a mix between projections/speculations and demonstrated devices [DH15b]; however, some of the loss values can have a decisive impact on the efficiency of an ONoC and could potentially make less favourable designs with one technology assumption more favourable with other assumptions. For instance, while some studies assume 0.3 dB/mm waveguide propagation loss [GMS+14] [CAJ15],

others use 0.0271 dB/mm [BS11] [OOTR+17], which is an order of magnitude difference. Both waveguide technologies exist, have been successfully demonstrated, and could be deployed in future ONoCs. This applies to many different devices loss parameters: for example, MR-drop losses of 1.5-0.5 dB can be found in literature [LBGP14] [OOTR+17] [HJH14], so can MR-through losses of 0.01-0.0001 dB [JBK+09] [LBGP14] [CAJ15]. Therefore, we strongly believe that the most useful approach is to evaluate ONoC proposals with both aggressive and conservative technological assumptions to identify the impact on the power efficiency of a topology, or even perform design sweeps across a range of device parameters.

On-chip vs. Off-chip Lasers Advances in on-chip laser technologies based on Germanium [CACP+12] and/or Indium phosphide [FPC+06][PB06] enabled to integrate DWDM-compatible lasers with compact footprint into the photonic die. This enables to batch process lasers along with the photonic die and eliminates the need for laser source coupling, which reduces both coupling loss and packaging costs. In addition, since integrated on chip, it is possible to switch lasers on and off within nanoseconds, which would pave the way for adaptive laser control mechanisms that can have tremendous potential to reduce laser power (up to 92% [DH15b]). Combining this promising technology with advances ONoC architectures is discussed in more detail in Section 7.4 on future work.

Although imposing higher packaging costs and coupling losses, off-chip lasers benefit from higher temperature stability which results in higher wall-plug efficiencies. For instance, while state-of-the-art on-chip laser technologies exhibit laser efficiencies of a maximum of 15% [KNK+13], for off-chip lasers they could be as high as 30% [HB14]. The technology that ultimately offers the higher power efficiency depends on both the actual coupling losses and the wall-plug efficiencies. The vast majority of studies currently assume an off-chip laser source as they are currently more mature, provide significantly higher manufacturing yield than on-chip lasers, and can be replaced easily if defective [DH15b]. In addition, many studies do not count laser power of the off-chip laser towards the processor power budget (and thermal design power); however, Heck et al. [HB14] note that laser power should still be considered carefully since it does have an impact on the total overall system efficiency.

In either case, designers should be aware that the number of lasers that can be coupled into the chip is low since it is either limited by packaging constraints and cost

(off-chip), or by area, layout, and temperature constraints (on-chip) [CZC+14]. Implementing a large number of optical links would thus require the light to be distributed across the chip using splitters, which introduces additional losses. In addition, given the direct dependency of laser power on the number of wavelengths and IL_{max}, ONoCs should be designed so that the total number of wavelengths required in the NoC is low, and paths should be designed with device losses in mind to minimise IL_{max}. Although devices are constantly evolving, there is currently no clear roadmap for SiPs regarding device losses. Therefore, it is reasonable to assume that optical losses will remain a crucial issue in the near future and must be addressed rigorously in the NoC design and layout.

Dynamic Power Consumption

Global electrical wires have become increasingly energy-consuming in many-core architectures [And14], as they require repeaters, regenerators or buffers to provide satisfactory signal integrity and latency, with increasing energy consumption for longer link lengths. Figure 2.9 plots the energy required to transmit a 64-bit packet over an electrical and optical link with increasing link length, modelled with DSENT [SCK+12] with a 22 nm technology. Both links have a link bandwidth of 64 bits/cycle and are clocked at 5 GHz with 10 Gb/s modulators on the optical link. Optical links consume energy in the backend circuitries and for modulation/detection. For short distances, the electrical link is more energy-efficient as it does not require EO and OE conversions; however, for link length $>$ 0.5 mm, the relatively-distance-independent energy consumption of optical data transmission dominates electrical links. From a dynamic energy perspective, it is therefore beneficial to utilise electrical links for short distances. For instance, in a 64-core chip, tile widths/lengths are often between 1-2 mm [BSP+16, VTL+16], meaning that only communication to direct neighbours should be electrical in this case. A trend towards growing core counts and die sizes would make optical data transmission increasingly beneficial in terms of energy/dynamic power, particularly for communication between nodes located at large distances to each other. Note that the exact values of Figure 2.9 on the optical link can vary for different SiP devices if higher losses are assumed; however, the relative trend, i.e. low distance dependency of optical data transmission, would not change. In addition, router traversal of a 64-bit packet in 22 nm at 5 GHz requires $\sim$2 pJ, which is similar to the energy needed to traverse an electrical link of 1.3 mm - further emphasising the significance that electrical links have on total energy consumption of NoCs.

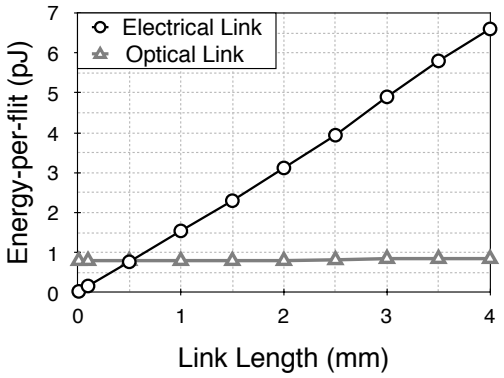


Figure 2.9: Dynamic Energy for Transmitting a 64-bit Packet

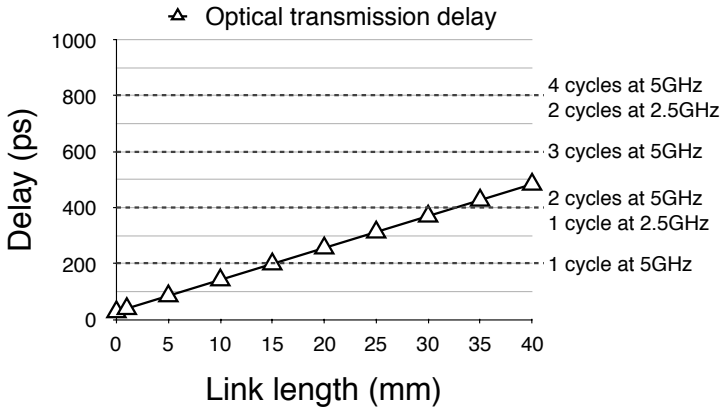


Figure 2.10: Transmission delay vs. link length

2.4.2 Latency and Throughput

According to Ho et al. [Ho06], electrical signal propagation takes 131 ps/mm in an optimally repeated wire at 22 nm. At 5 GHz, one hop over an electrical link in a NoC is therefore commonly accepted to take one clock cycle (note that this is also subject to clock frequency, layout, final link lengths, etc.). Optical links, on the other hand, require EO and OE conversions and signal propagation delay in the waveguide (t_{prop}). Signal propagation of light in silicon waveguides, however, has been identified to be 10.45 ps/mm (based on models utilising International Roadmap for Semiconductors (ITRS) predictions [HCC+06]), which is particularly beneficial for long-distance communication, especially because optical links, as opposed to electrical links, do not require repeaters/pipelining to drive and/or speed-up the signal. Figure 2.10 plots the transmission delay on an optical link vs. the link length. In addition to waveguide propagation delay, latency includes the delay for the EO backend (9.5 ps), modulator (14.2 ps), detector (0.22 ps), and OE backend (4.0 ps) [CCH+07]. We observe that

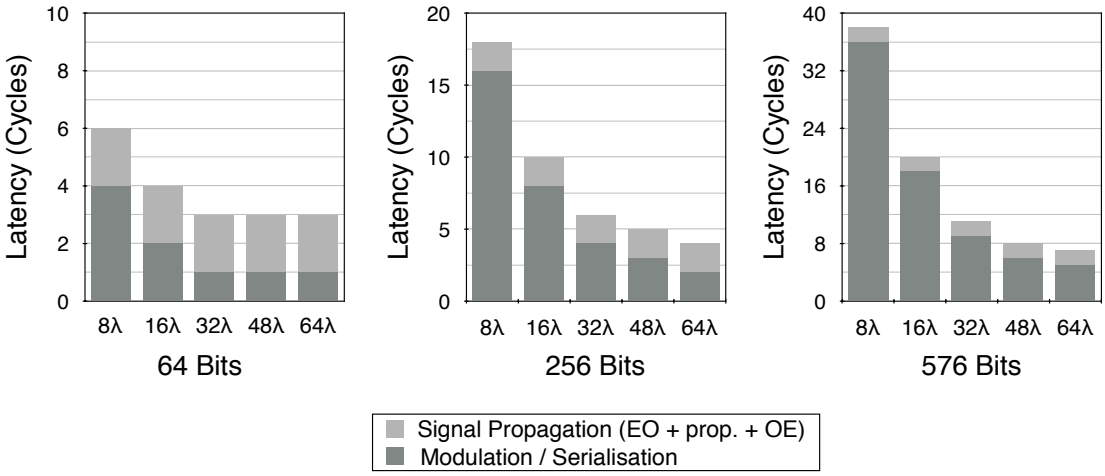


Figure 2.11: Latency vs. optical bandwidth on a typical optical link. For simplicity: propagation and OE delay take 1 clock cycle each. We assume 5 GHz core clock rate and 10 Gb/s link data rate, i.e. a serialisation degree of 2.

distances/link lengths on chip have very little impact on the overall latency of optical links, and long distances can be traversed within one core clock cycle.

The major contributor to latency is data modulation, i.e. the time it takes to serialise a packet based on the available bandwidth and link data rate. This is outlined in Figure 2.11, which lists the impact on the delay of different packet sizes common in the on-chip domain, with different numbers of wavelengths typically required in ONoCs. We assume a link propagation delay and delay through the backend circuitries of one cycle for simplicity, and a typical link data rate of 10 Gb/s (modulators/detectors) and 5 GHz core clock frequency. With this configuration, two bits can be modulated on one wavelength in one core clock cycle – leading to one core clock cycle modulation delay of a 64-bit Packet with 32λ. These values are an important guideline in order to ideally trade-off power required for increasing number of wavelengths and latency. For instance, increasing link bandwidth from 16λ to 32λ decreases latency only by one clock cycle to transmit a 64-bit packet, but more than doubles laser power. Bandwidths lower than 8λ introduce too much latency in relation to the power benefits. This illustrates that designers should always carefully balance optical bandwidth and laser power based on the actual bandwidth demands.

In order to minimise packet latencies, these delays must be carefully compared to the electrical delay. Although electrical links do not need EO and OE conversions, the only energy-efficient way of reaching distant cores is through several hops in a topology, which introduces router delay and contention in the NoC. Router traversal delay

depends on the clock frequency, and high clock frequencies of 5 GHz may need up to 5 pipeline stages (e.g. Intel's TeraFLOPS design [VHR+07]). If we assume aggressively pipelined routers that can be traversed in two clock cycles (assuming enough link bandwidth), one hop would take 3 cycles. While this delay adds up for each additional hop to reach a destination, hardly any delay is added on optical links when the distance increases (assuming direct connections). Optical links are thus the preferred choice in terms of latency and energy if distances are large enough.

2.4.3 Physical Layout and Integration

Manufacturing SiPs is still a niche market and usually imposes tight constraints. The costs associated with laser source coupling in the chip packaging process limits the number of laser sources available for a NoC, which has to be considered by designers when proposing a NoC topology. For instance, designing topologies that require a large number of links that need to be provided with light leads to large amounts of splitting and in turn higher insertion loss. If a NoC topology requires different wavelengths in different links, either one laser source for every wavelength set is necessary or wavelengths need to be distributed using MR filters, which incurs additional loss.
The impact of the number of wavelengths in a NoC on the required laser power may also have another undesired side effect: the maximum attainable output power of current multi-wavelength laser is limited, which means that one laser source may not be able to supply the entire NoC with light at sufficient power levels. Also, based on the coupling point of the laser and NoC layout, potentially large distances must be traversed on chip to provide a waveguide with light. Moreover, the more laser sources are required by the system, the higher the cost. It is therefore essential for the efficiency of a design that a detailed analysis of the physical layout/floorplan is conducted when proposing a (logical) topology.
The placement and spacing of SiP components is important to mitigate crosstalk noise. Although advances in materials and device technologies lead to compact SiP components, the spacing required between components makes their placement and layout non-trivial. If every tile has to be provided with modulators and receivers for optical communication, sufficient space must be available to place and interface these devices. Recent work assumes 5 μm clearance between MRs to avoid crosstalk [LBGP14], which imposes a tighter limit on the number of MRs that can be provided to each tile with common tile dimensions of 1-2 mm. As mentioned before, integrating optical components on-chip is widely envisioned to be implemented by placing them on

a separate layer using 3D integration; however, while this decreases the interferences between the SiP and CMOS components, the spacing considerations still exist since each MR is driven by a through-silicon via (TSV), which can have a diameter of 10 μm [SZZ+14] [DH15a].
Apart from spacing issues, routing waveguides should avoid excessive waveguide crossings since they increase IL_{max}. Recent studies have compared a number of different ONoC designs and revealed that minimising waveguide crossings in the layout can lead to longer waveguide lengths and in turn propagation losses [RGBB13]. In addition, it is important to study how logical topologies can be mapped to a physical layout as the number of unavoidable waveguide crossings also depends on the topology.
Finding the ideal physical layout also depends on the utilised device technologies, in particular their loss values: for instance, for technologies with high losses for waveguide crossings and low waveguide propagation losses, it may be more efficient to implement longer waveguides if that allows to minimise the number of waveguide crossings. Given the young age of ONoCs, many past proposals required designers to find an efficient/ideal physical layout for a given topology manually. However, recent years have seen a rise of numerous automatic synthesis and layout tools that assist designers to explore the design space, to minimise IL_{max}, and in turn significantly improve ONoC designs [BRSB13].

2.5 Summary

All in all, designing optical NoCs requires a very careful apportioning of the optical resources in order to keep static power low and to result in feasible and layout-friendly designs. The fact that static power can grow sharply as the number of optical links and number of wavelengths is increased means designers need to find architectures that make the most out of the available resources. The number of wavelengths in a NoC does not only increase laser power but also MR heating power since each MR responds to one particular wavelength. To reduce static power, one design objective should be to efficiently utilise the available optical bandwidth. Although optical data transmission offers low latency over wide distances, decreasing bandwidth to reduce static power leads to serialisation latencies that should be considered carefully to obtain performance goals. Finally, the number of laser sources and their output power for current technologies is limited in on-chip networks. NoC topologies should thus have a clear notion of how light is distributed across the chip and the incurred loss overheads.

Chapter 3

Optical Network-on-Chip Design: State of the Art

3.1 Introduction

The scientific literature is replete with studies aiming to explore the most efficient use of optical links in the on-chip communication fabric. Novel architectural approaches are constantly appearing and demonstrate that advanced NoC designs can be decisive for power efficiency. Given the vast design space and opportunities that SiPs have to offer to designers, novel architectural approaches are evolving to obtain higher efficiency than the state of the art. The biggest challenge tackled by all designs is to make efficient use of optical resources to keep static power at acceptable levels while meeting performance goals. A large number of different approaches have been proposed, which can roughly be categorised into 1) all-optical NoCs utilising optical data transmission only, 2) hybrid NoCs that utilise both electrical and optical links, and 3) bandwidth sharing and bus arbitration techniques that leverage TDM. This chapter introduces recent proposals within these categories and points out to what fields the thesis at hand contributes. In addition, it discusses currently available simulators and modelling tools for ONoCs used by the community and their capabilities and shortcomings. Finally, this chapter presents the interested reader with recent studies dealing with SiPs on larger scales than on-chip interconnects, research in design synthesis tools, and adaptive laser control mechanisms, which are, aside from NoC architectures, fundamental to the widespread adoption of ONoCs in the future.

3.2 All-optical Networks-on-Chip

3.2.1 Motivation

With increasing core counts, die sizes, distances on chip, and a lack of technology scaling of electrical interconnects, the question arises whether we will soon reach a point at which a complete technology shift from electrical to all-optical NoCs, i.e. NoCs performing optical data transmission only, could offer the highest power efficiency. The benefits of performing data transmission all-optical are clear: low energy even for large distances, high-speed optical data transmission, and high bandwidth density through DWDM. At the same time, electrical NoCs are very distance sensitive as a large number of hops, i.e. link and router traversals, may be necessary to reach a destination, leading to additional latency and energy for each hop, network congestion, buffer requirements at intermediate nodes, and poor scalability for increasing number of nodes. To make all-optical NoCs superior to electrical NoCs, designs must exhibit low static power by decreasing the number of wavelengths, path lengths, and MR requirements. Several different approaches have been proposed in recent years. This section will review previously proposed ideas and analyse their strengths and weaknesses.

3.2.2 All-optical NoC Proposals

Contention-free Wavelength-routed Optical NoCs

Contention-free, all-to-all communication requires an optical crossbar, which could simply be implemented by using SWMR or MWSR buses with one bus assigned to each sender (SWMR) or receiver (MWSR); however, this scales the waveguide count linearly with the number of nodes, causing large numbers of MRs and optical links, and in turn unacceptable power overheads.
Wavelength-routed optical NoCs (WRONoCs) overcome this issue by implementing wavelength-selective filters to route optical signals through the network according to their wavelength. A sender selects a wavelength for modulation based on the destination it wants to address. Wavelengths are re-used across all senders, and each sender uses a different wavelength to address a destination. An N-node WRONoC thus requires N different wavelengths in the NoC – one for each destination – to enable each sender to transmit data to each receiver. To avoid data corruption of two optical signals sent on the same wavelengths on the same waveguide, switches based on MR filters are implemented to provide collision-free paths between any source-destination pair.

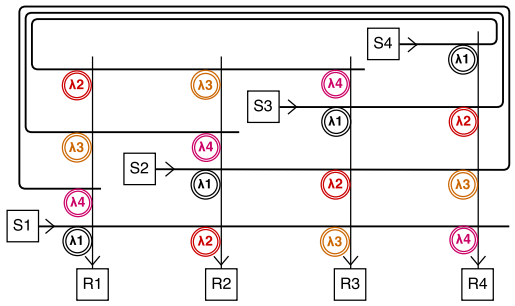


Figure 3.1: Folded Crossbar [RGBB13]

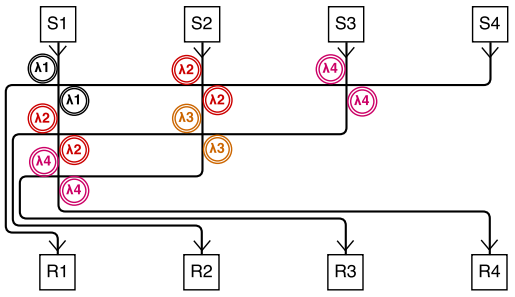


Figure 3.2: Snake [RGBB13]

The (wavelength-based) routing algorithm is thus embedded in the MR switches and thus deterministic.

Several WRONoC topologies have been proposed, such as the λ-Router [BGB+07], the folded crossbar or Snake [RGBB13]. Figures 3.1 and 3.2 illustrate the latter two. MR filters are placed as shown in the figures to drop wavelengths so that the optical signals are forwarded to the correct destinations. Senders will choose the wavelength assigned to the destination for data modulation, which will then be routed through the network to the destination. Multiple source-destination pairs will communicate on the same wavelengths, which are all guaranteed collision-free paths through the NoC. Each sender needs to have modulators to be able to address each destination. Realistically, more than one wavelength is required to provide the NoC with sufficient throughput. For instance, just one wavelength per sender would result in a link bandwidth of 2 bits/cycle for 5 GHz core frequency and 10 Gb/s modulators. Therefore, a set of wavelengths (λ-set) is assigned to each node for data transmission.

Since every node requires $(N-1) \times \lambda$ modulators, the number of MRs required just for modulation is $(N-1) \times N \times \lambda$ in an all-to-all contention-free WRONoC, and that is not factoring in the MRs for performing wavelength-selective routing (depends on topology) and filtering at the receiver ($(N-1) \times \lambda$). In addition, a laser (or multiple lasers) must provide $N \times \lambda$ wavelengths to the NoC. The contention-free switching of WRONoC crossbars, therefore, comes at poor scalability with regard to laser and MR heating power. Although the number of wavelengths in these NoCs could be reduced by increasing the number of waveguides (spatial-division multiplexing), this approach is limited by layout constraints and area overheads and does not fundamentally solve the power consumption and scalability issue. Therefore, these kinds of all-optical NoCs are only practical for CMPs of smaller scales (16 cores and less [OOTR+17]).

Ye et al. [YXH+13] propose a WRONoC for 3D mesh-based ONoCs that allows for a regular topology and light-weight, non-blocking optical routers with dimension-order

routing as known from electrical NoCs; however, just like the previous designs, its topology consists of all-to-all optical switches, which lead to large MR requirements, and relying on a high amount of wavelength switching considerably contributes to the overall path losses and in turn laser power. Aurora [LQJ+15] also provides a mesh structure and thus has similar drawbacks.

Ramini et al. [RGBB13] studied whether topologies that use switching elements (i.e. MR filters) to perform routing are absolutely necessary, or whether ring topologies that rely on spatial division multiplexing (SDM), i.e. communication spread across several different waveguides, rather than MR switching can provide higher efficiency, particularly as their design imposes fewer waveguide crossings. Their results reveal that optical ring topologies, although simpler, offer poor scalability with regard to waveguide lengths (and in turn propagation losses) by relying on SDM for routing and have high connectivity requirements, which translate to higher power as the number of nodes increases. Note that their study is based on waveguide propagation loss of 0.15 dB/mm; however, since then, waveguides with significantly lower loss have been demonstrated (0.0271 dB/mm [BS11]), which may make ring topologies more favourable. A comparison of these two types of topologies with more advanced technology parameters has not been published yet, but would certainly be interesting as ring topologies require no MRs for switching and in turn less MR heating.

All in all, contention-free WRONoCs suffer from very limited scalability due to high MR tuning power which renders these approaches infeasible for larger number of cores. However, contention-free operation is often unnecessary to satisfy on-chip communication demands, and deploying some sort of resource sharing can help to alleviate the power consumption and scalability issue of WRONoCs. This approach has been evaluated by a number of recent proposals that utilise different forms of resource sharing, and will be discussed in the following.

Control Network Based Wavelength-routed Optical NoCs

WRONoCs discussed in the previous section have high demands in both laser and MR heating power with poor scalability as they provide all-to-all contention-free paths through the entire NoC and need one dedicated wavelength set for each destination. A number of proposals identified these limitations and presented numerous novel ideas to improve the scalability of WRONOCs by lowering the number of wavelengths and MRs in the NoC [LBTO+11] [KAH11] [KH12] [HJH14].

The basis of these approaches is a separate control network on which nodes must check

for availability of a destination and reserve it prior to data transmission by exchanging request (REQ) and acknowledgement (ACK) packets. This effectively reduces the number of receivers at the destination nodes from $(N-1) \times \lambda$ to just λ, since only one source can transmit to a destination at any given time. This allows collision-free operation without the need for providing all-to-all contention-free paths.
In addition, the number of wavelengths of contention-free WRONoCs were identified to be impractically high in terms of laser source requirements and laser power. Therefore, a splitting of the address space has been proposed, i.e. multiple destinations share the same λ-set as their address. This can lead to situations in which different sender-receiver pairs communicate on the same wavelength simultaneously. To prevent data collision, more sophisticated switching topologies are required to provide collision-free paths under any circumstances. However, the total number of MRs is actually decreased compared to crossbar WRONoCs since in total fewer wavelengths must be routed through the NoC.
Based on these techniques, several studies further improved these types of WRONoCs by decreasing the number of λ-sets necessary for addressing, effectively reducing the total number of wavelengths in the NoC which in turn decreases laser power and/or the number of lasers coupled into the chip. CoNoC [KH12](Figure 3.3) uses $N/2$ λ-sets for addressing, i.e. two nodes share the same λ-set address. As this can lead to situations in which two sender-receiver pairs communicate on the same λ-set ('collision'), their topology is equipped with 'cross-links' that are selected by the sender based on its distance to the receiver and provide collision-free paths through the NoC. QuT [HJH14] further improves CoNoC by proposing a collision-free topology and routing algorithm with both 'cross'- and 'bypass'-links that allow for collision-free communication with four destinations sharing the same λ-set for addressing, thereby reducing the number of λ-sets to $N/4$. Figures 3.3 and 3.4 show the topologies of these designs and routing examples (enabled by MR filters) that provide collision-free paths when two senders modulate on the same λ-sets to transmit to destinations that share the same λ-set. This illustrates how MRs can be utilised to design sophisticated ONoC architectures that improve overall design efficiency significantly.

3.2.3 Summary

All in all, while some studies suggest that for smaller network sizes ($<$ 16 nodes) contention-less WRONoCs are more power efficient than arbitrated ONoCs (mainly because arbitration overheads are more significant in smaller NoCs)[RTB14], they are

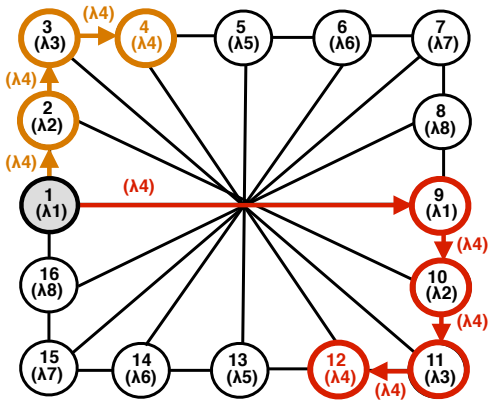


Figure 3.3: CoNoC: a ring topology complemented with cross links provides paths for two optical signals on the same wavelength to traverse the NoC simultaneously without collision.

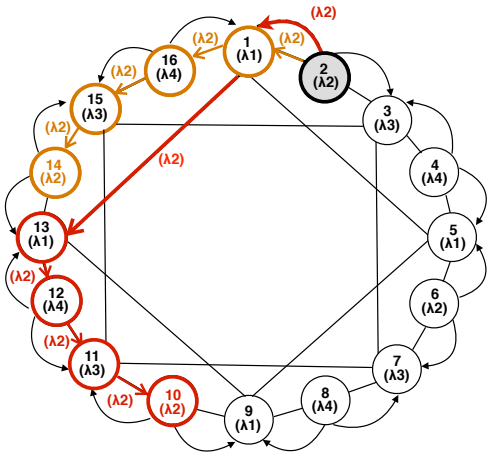


Figure 3.4: QuT: both cross and bypass links added to a ring topology provide paths for up to four signals on the same wavelength to traverse the NoC simultaneously without collision.

not suitable for higher number of nodes as they scale poorly in terms of MR count and number of wavelengths and make inefficient use of the available bandwidth. However, arbitrated WRONoCs based on control networks were shown to significantly reduce both the number of wavelengths and MRs, making them a more suitable candidate for NoCs of larger scale. In fact, the most recent study (QuT) outperforms a large number of alternative state-of-the-art ONoCs in terms of power consumption. Although state-of-the-art designs offer topologies capable of decreasing laser power and MR heating significantly, they still require too much static power overall and should be improved by more advanced designs.

In addition, although rigorous evaluations under synthetic traffic have been conducted, none of the discussed control network based proposals was evaluated with realistic traffic workloads. Emerging multi-threaded applications in CMPs, however, often exhibit structural and transient hotspots, i.e. some nodes receive or transmit a disproportionally large amount of packets either over the entire course of execution (*structural hotspot*), or temporarily in some of the application phases (*transient hotspot*) (as identified by Gratz et al. [GK10]). These hotspots could have a large impact on control network based WRONoCs in which a destination can only receive data from one sender at a time, potentially leading to high contention and in turn large performance degradations. A simulation study stressing these NoCs with realistic traffic would be important

to identify the impact of realistic workloads on these NoCs.
Another shortcoming of the previous proposal is that they do not explicitly assume a laser power distribution network (LPDN) – a term introduced by Ortín-Obón et al. [OOTR+17] denoting the network that distributes the light for data transmission to each source – and report laser power of different switching topologies only for the highest optical loss path [KAH11][KH12][HJH14]. However, each node in the NoC must be provided with light to transmit data to every other node which would require one dedicated laser source at each node. This approach is impractical as laser source coupling is a major cost factor in the packaging process. Most recent studies thus explicitly assume and analyse a LPDN [OOTR+17] and reveal that the LPDN adds to the total loss and requires a detailed analysis, especially for NoCs of larger sizes in which light must be distributed to a high number of nodes. Besides, although laser technologies are currently improving at a fast pace, the output power per laser is limited. Assessing the total power required per laser is thus important to identify what is technologically feasible, and what the laser output power demands of a WRONoC topology are. For instance, if the laser power demands of a WRONoC are too high to be served by just one laser, multiple lasers must be coupled into the chip which in turn increases design cost.
Section 4 tackles these issues by proposing novel extensions to QuT's destination-reservation mechanism which allows to parallelise the control packet exchange on the control network to ongoing data transmission on the data network, thereby significantly improving performance. In addition, 'Amon' is introduced, a novel WRONoC topology tailored to a mesh-based layout to reduce path losses and simplify layout, with fewer MRs necessary to perform routing compared to QuT and novel architectural modifications to the switch backends to reduce both the impact of traffic hotspots and power consumption. Moreover, a detailed analysis and example layout of a LPDN for Amon is presented.

3.3 Hybrid Networks-on-Chip

3.3.1 Motivation

A number of recent studies have revealed that discarding electrical interconnects in NoCs altogether actually leads to unnecessary inefficiencies regarding both performance and power for the current state of electronic and SiP technologies. This is attributed to the fact that there are situations in which electrical interconnects are, despite their shortcomings, more efficient than their optical counterparts. There are numerous reasons to assume that the first step of integrating optical data transmission on-chip will probably be to utilise a NoC that features both electrical and optical links in its topology. These benefits include not only power and latency, but also design cost and integration challenges.

The deciding factor to make emerging technologies more pervasive in commercial products is cost. While the fabrication of electrical interconnects is a mature process, SiP processes are considered to lag roughly ten years behind (recent prototypes use a 45 nm technology which dates back to 2007 [OMS$^+$12, GMS$^+$14, LSZP14]). In addition, since SiPs are a niche market with low volumes, their production imposes additional cost overheads and may only be justified by high-end applications. Therefore, unless optical data transmission for very short distances is significantly beneficial to electrical interconnects, and these benefits are urgently required by current applications, there will be very little motivation for engineers to take the risk of moving to a new technology that entails bigger feature sizes, more challenging integration, and design cost. Since electrical interconnects can still satisfy current demands for short distances, replacing them with optical links that can only provide marginal benefits for short distances seems unlikely to happen. Supplementing short-distance electrical links with optical links for larger distances seems a more natural step forward.

Another important consideration is that electrical interconnects consume significant amounts of dynamic power consumption. Optical interconnects, on the other hand, have very low dynamic power but large static power demands for MR heating and at the laser source. While static power overheads can be mitigated for communication-intensive workloads, it becomes more significant for applications that feature frequent periods of idleness. Unfortunately, this is the case for a number of application domains, such as scientific computing with compute-intensive execution phases [DH14], or even in server computing (utilisation rates can be around 30% [BH07]). In order to be suitable to a wide range of application domains, static power in NoCs should

therefore be kept to a minimum. This can be achieved by offloading traffic to electrical interconnects, i.e. utilising both technologies in a hybrid NoC design. The following section discusses the proposals in literature dedicated to revealing the most efficient way of doing so.

3.3.2 Hybrid NoC Proposals

Network Topologies Implementing Both Electrical and Optical Links

Many hybrid NoC proposals revolve around the idea of implementing both an electrical and optical network and utilise them in a distance-based fashion in which electrical links are utilised for short, and optical links for long distances. To do this efficiently, a number of nodes are typically grouped into clusters. Intra-cluster communication is executed over an electrical network, whereas inter-cluster communication requires the sender to first send the packet over the optical network to the cluster in which the destination resides. Once the destination cluster is reached, the local electrical network is used to forward the packet to its destination within the cluster.

Meteor [BP14] utilises this approach and evaluates such a NoC for varying cluster sizes. Based on their results, the most efficient way is to cluster 16 nodes (4×4 submeshes) for a 64-node NoC with a 8×8 layout. This is illustrated in Figure 3.5. Gateway routers for accessing the optical NoC for inter-cluster communication constitute the access points of each cluster to the optical network, which consists of four MWMR buses of 64λ bandwidth each. Each gateway router has access to all four of these buses.

Atac [KMP+10](Figure 3.6) proposes both a 64-node and 1024-node version. In the former, the cluster size is one, i.e. each node has access to the optical network, which is a global crossbar with 32 SWMR buses of 64λ bandwidth each. Rather than dealing with fixed clusters, nodes transmit data on the optical network if the destination is further than four hops away. Otherwise, the electrical NoC is used. Providing all-to-all global optical communication, however, is highly inefficiently due to large static power overheads. Atac's 1024-node version comes closer to the previous approach as it utilises the same network as in the 64-node case, but clusters 16 nodes at each access point. Their original proposal performs intra-cluster communication over a 2D electrical mesh; however, they refine this in a follow-up study by replacing the mesh with a more efficient star network [KSC+12].

Firefly [PKK+09], as shown in Figure 3.7 for 64 nodes, concentrates four cores at each

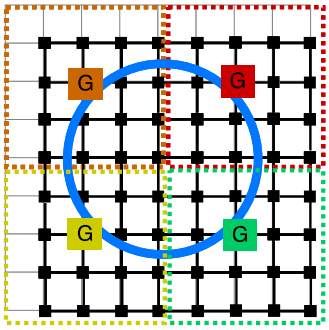


Figure 3.5: Meteor

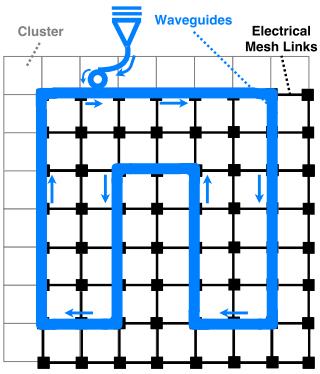


Figure 3.6: Atac

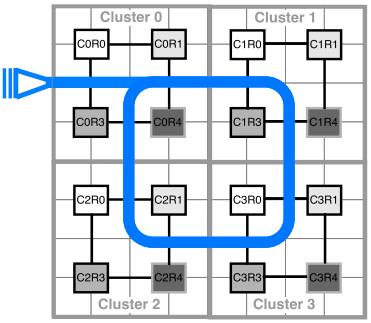


Figure 3.7: Firefly

router, and defines fixed clusters containing four routers each. Intra-cluster communication is executed over an electrical 2D mesh. Each router in a cluster has a dual in every other cluster, with which they are connected optically (e.g. C0R0, C1R0, C2R0, and C3R0) and together form what the authors refer to as 'Assembly'. For inter-cluster communication, packets are therefore sent over the optical network to the router in the Assembly that resides in the same cluster as the destination. From that point, packets are forwarded to the destination on the electrical mesh. Optical links connecting the nodes of an Assembly are implemented as R-SWMR buses, as introduced earlier. The formation of Assemblies decreases the number of nodes that form a crossbar and thereby reduces the total number of MRs. Electrical links are efficiently utilised for short distances. Concentrating four nodes at each router requires higher bandwidth on the links to avoid early saturation. This increases bandwidth requirements on the optical links and, in turn, laser power. Their evaluation results show that the traffic pattern has a large impact on performance, with localized patterns being more benign.

ORNoC [LBTO+11] deploys a similar topology as Firefly in the sense that cores are grouped in clusters and perform intra-cluster communication on an electrical mesh network. Inter-cluster communication takes place through the optical network ORNoC, consisting of an optical ring on which wavelength assignment is performed automatically for contention-free use of shared optical resources. Each cluster contains a gateway router through which all nodes of a cluster can access the optical NoC. Packets are routed to/from the gateway over the electrical intra-cluster mesh when a destination resides in a different cluster.

HOME [MYW+10] is a hierarchical hybrid NoC that utilises a packet-switched electrical mesh network for local intra-cluster communication, and a circuit-switched optical network for inter-cluster communication. Four nodes are clustered at one HOME

router through which both the electrical and optical networks are interfaced.
Phastlane [CKA09] routes packets on optical links not based on distance, but based on packet size. It combines a packet-switched mesh network with an optical, contention-free crossbar on which it transmits cache lines over several hops in one cycle. The optical crossbar utilises a simple, predecoded source routing approach.

Combining Electrical Routers with Optical Links

A number of proposals use electrical interconnects only as local links to connect nodes to their input router, utilise optical links for global interconnects, and intermediate electrical routers to implement the routing functionality.
Joshi et al. [JBK+09] propose a three-stage SiP Clos (PClos) that uses point-to-point optical links for low-energy, long-distance data transmission between the Clos stages. Both the routers and the links between the cores and routers are electrical – connections between routers are optical. Clos networks have high path diversity and show constant performance across all traffic patterns; however, each message has to pass through all router stages, which is inefficient for applications that leverage locality of cores or perform near data processing [BCM+14]. BLOCON [KC11] is a buffer-less implementation of PClos that features a scheduling algorithm and path allocation scheme for managing routing in the Clos. It lowers latency and improves throughput, but also has higher MR heater and laser power compared to PClos.
A similar approach as PClos is taken in PROPEL [MK10], which also utilises optical links for data transmission between intermediate routers. In a mesh-like layout, each node can send and receive data to/from every other node in the same row and column, where a MWSR bus is dedicated to each node. If a destination does not reside in the same row or column as the sender, XY-routing is performed: the packet is first routed to the router in the same row that resides in the same column as the destination node, and is subsequently forwarded to the destination over the column link connecting the intermediate router and the destination. Four cores are clustered at each node, and 16λ bandwidth is provided between any two routers. Since PROPEL basically implements optical crossbars in each X and Y direction, scaling it up directly would lead to large optical resource requirements. Consequently, the authors propose E-PROPEL, a 256-node solution that clusters four 64-node PROPELs which provides more efficient bandwidth scalability.
MPNOC [ZL10] concentrates four cores at each router and implements four clusters of 64 nodes (or 16 routers). MPNOC utilises a 3D approach where 16 decomposed

optical crossbar slices are placed on a separate optical layer each to minimise the number of waveguide crossings. Each slice is thereby a 16×16 crossbar that connects all tiles from one cluster to another (inter-cluster communication), or all tiles from same cluster (intra-cluster communication). Crossbars are composed of MWSR buses, one for each receiver.

PHiCIT [RJS15] contrasts with the previous proposals in the sense that it divides the NoC into equally-sized clusters but uses a 2D electrical mesh for inter-cluster communication and optical crossbars for local, intra-cluster communication. The authors argue that inter-cluster communication has low activity during application execution, however, these few messages demand high throughput as they are required for main memory, task migration, or internal synchronisation traffic. For high throughput, electrical links are cheaper in the sense that they provide higher throughput for much less data-independent power, which makes them more suitable for these types of traffic patterns. Clusters, on the other hand, are organised by computation complexity, communication requirements, and functional relationship of IP cores, leading to much higher traffic demands, making optical crossbars more suitable, particularly for small cluster sizes. Their design is superior to a baseline electrical mesh and an optical mesh NoC, based on their evaluation results. However, both of these NoCs were shown to be inefficient, and PHiCIT seems to be tailored to particular traffic patterns and microarchitecture, which may limit the suitability of this approach to systems where the application domain is known a priori (e.g. like in embedded systems).

3.3.3 Summary

Most hybrid NoC designs proposed in the scientific literature identified that electrical links are more suitable for short-distance communication, and optical links for longer distances. Optical links are often deployed to decrease the diameter (i.e. average hop count) of a topology as long-distance connections are easier to implement than with electrical links and do not incur distance-related energy overheads. In addition, many designs try to attain high bandwidth utilisation of the optical links to de-emphasise their static power overheads. Often, multiple nodes are clustered around an optical link which they use through some sort of hub router. Chapter 5 contributes to the research area of hybrid NoCs by proposing a novel distance-based approach of combining electrical and low-bandwidth optical links in a topology that aims to utilise both interconnect technologies in a way that they ideally balance out each other's drawbacks. The effectiveness of this approach is demonstrated with a novel NoC architecture.

3.4 Bandwidth Sharing and Arbitration Techniques

3.4.1 Motivation

Next to novel approaches to improve all-optical and hybrid architectures on the NoC-level, the static power overheads associated with optical bandwidth scaling requires an explicit focus on maximising bandwidth utilisation of optical buses to achieve high power efficiency. Bandwidth sharing has been identified by many as an efficient approach to doing so: multiple nodes can send on the same optical bandwidth on the same waveguide and use this bandwidth in a TDM fashion to maximise link utilisation. An example of this approach is the shared optical bus introduced in Section 2.3.2. The number of nodes sharing optical bandwidth, as well as the total bandwidth to be shared, is very sensitive to optical losses and power consumption. Therefore, recent proposals in literature investigated efficient ways of sharing bandwidth between nodes. Sharing optical bandwidth and utilising it in a TDM fashion requires arbitration prior to data transmission to ensure correct, uncorrupted delivery of data. Latency and energy imposed by arbitration can have a considerable impact on the overall performance and efficiency. Arbitration techniques that impose as little overhead as possible and that are suitable to the on-chip domain have therefore received attention by the research community, too.

3.4.2 Bandwidth Sharing and Arbitration Proposals

Token Ring Arbitration

Token ring arbitration with optical tokens has been studied by a number of proposals [VSM+08][VBSL09][PKM10][MKL12]. Corona [VSM+08] has an optical crossbar on which a node is permitted to send data by contending for the bus on an optical token-ring arbitration network, implemented as a MWSR bus. The authors further deepened this study with a detailed analysis of a channel-based and a slot-based distributed token-based arbitration, and their suitability for optics [VBSL09]. Their schemes vary priorities dynamically to ensure fairness. In FlexiShare [PKM10], a reduced number of channels are globally shared for which arbitration is performed with separate channels and buffers, leading to slight additional power and area overheads. Their token-stream mechanism for channel arbitration and credit distribution, however, is highly efficient as it halves the amount of utilised channels compared to a conventional crossbar under balanced, distributed traffic. R-3PO [MKL12] utilises a token-based control network

to handle accesses in a 3D NoC with an optical crossbar. Generally, the main issue with token ring arbitration is that it leads to an increase in latency with increasing number of nodes due to longer waiting times for receiving a token, thus providing limited scalability. A number of alternative arbitration schemes have been proposed to alleviate this problem.

Alternative Distributed Arbitration Schemes

Featherweight [PKM11] is a light-weight arbitration scheme with QoS support that implements a feedback-controlled, adaptive source throttling scheme to asymptotically approach weighted max-min fairness among all nodes. It provides large power reductions while providing freedom from starvation with negligible throughput loss.
In 'Channel Borrowing' [XYM12], each channel is allocated to an owner node, but can also be utilised by a few other nodes during idle time. Each node has a statically assigned channel to avoid starvation and can borrow an additional idle channel to boost bandwidth and improve network utilisation. The authors propose a selection policy for choosing a channel to be borrowed that enables low probability of conflict, and a distributed arbitration mechanism to resolve contention of multiple nodes wishing to borrow the same channel.
'Wavelength Stealing' [ZKS+13] enables opportunistic channel sharing without incurring any arbitration overheads by implementing collision recovery. Similar to 'Channel Borrowing', each node has one dedicated channel to each destination on which service is always guaranteed, essentially implementing a point-to-point network. In addition, senders can steal access to channels owned by other senders to that same destination they want to transmit data to, enabled by placing additional modulator MRs along the shared waveguides. Service on the stolen channels is not guaranteed and is performed arbitration-free: owners of the channels are not notified about the 'theft' and collisions that arise from it are corrected at the destination node using erasure coding.
GASOLIN [LYM15] proposes pipelined distributed global arbitration for MWMR crossbars that allows the arbitration process to be parallelised and simplifies arbiter design. The distributed arbiters implemented at each node share global request information to identify free channels and maximise bus utilisation. Compared to token-based arbitration, GASOLIN reduces the number of channels by 50%.
SUOR [WXY+14] implements a bidirectional ring waveguide that is divided into multiple non-overlapping sections that can be utilised independently, thereby supporting multiple transactions simultaneously. Their hybrid control network consists of agents

– one for each node – through which nodes can access the ring waveguide. Agents communicate with processing nodes optically with low delay and share information with each other over short electrical wires with high connectivity.

LumiNOC [LBGP14] implements a shared optical bus on which wavelengths are used for both data transmission and arbitration (referred to as 'in-band' arbitration), thereby saving the overheads of a separate arbitration network. Their buses have an arbitration phase prior to the data transmission phase, and form a double-back waveguide on which sending nodes will also receive the packet they were transmitting. This waveguide is important in their arbitration phase: all nodes are synchronised at the beginning of the arbitration phase, and every node on the bus is assigned to one unique subset of wavelengths for receiving. For arbitration, nodes modulate an arbitration flag containing the destination address, source address, and packet size indicator on every other node's wavelength set. The source address fields are 1-hot encoded, i.e. if a node wants to use the bus it sets the bit corresponding to its address to '1' in the source address field. By the time all arbitration flags have been received by all nodes, the source address field will be analysed to see whether there is only one node who wants to use the bus or whether there has been a collision (i.e. >1 bit set to '1' in the source address field). In case there is only one sender, each node knows who the destination is (encoded in the destination address field), and the packet length from which they can infer the duration of data transmission. Nodes will use this information to either 1) detune their MR filters if they are not a receiver or 2) tune in their MR filters for the duration of transmission. If data collision occurs, all contending nodes enter a dynamic scheduling phase in which all senders are scheduled sequentially on the entire optical bandwidth in which they first transmit an abbreviate arbitration packet to inform its destination to tune in followed by the actual data transmission. After data transmission, all nodes tune in their MR filters and enter the arbitration phase again. Results show that this approach leads to large power savings and good throughput.

Kakoulli et al. [KSKK15] used the same arbitration approach as LumiNOC, but perform arbitration on a parallel control bus. Although this increases resource requirements and power consumption, it was shown to significantly improve throughput and latency. In fact, their results indicate that the fairly small power overheads of a parallel control bus are justified by the large performance gains, suggesting that parallel bus arbitration is the overall more power-efficient design approach in terms of performance-per-Watt. If resources are not tightly constrained, adding a separate control bus is a superior solution to in-band arbitration.

3.4.3 Summary

Designs that enable an efficient use of optical bandwidth allow a NoC to achieve the same performance goals while requiring less total bandwidth, which has a large impact on static optical power and is, therefore, a key research area for efficient NoC design. An arbitration mechanism should ideally enable low latency overheads while enabling an efficient use of the available bandwidth, and various intriguing approaches have been proposed in the recent literature. This thesis contributes to this research area by proposing a novel bandwidth sharing and arbitration technique that offers large throughput benefits to shared optical buses by allowing bus utilisation on both time slots and subchannels. In addition, it evaluates the benefits of performing bus arbitration on a separate bus in addition to a data transmission bus. While many of the discussed proposals investigate arbitration techniques on the NoC level, the study in this thesis is dedicated to optical buses as they are typically the backbone of higher-order ONoC topologies. Therefore, advances on the buses would carry over to improve a NoC's efficiency as a whole. These proposals are presented in Chapter 6.

3.5 Performance Simulation and Power Modelling

Either new electrical NoC simulation infrastructures or extensions to existing ones are necessary to study and evaluate NoCs with SiPs in a meaningful way. Unfortunately, given the fairly young age of ONoCs, there is a lack of sophisticated simulation tools, particularly open-source tools, that would allow to evaluate NoC proposals in terms of power, performance, and area. However, the absence of such tools has been acknowledged by the community and numerous efforts have been conducted to implement open-source simulation infrastructures for both electronic and SiP NoCs [SCK+12] [CHB+10] [RBW+16]. This section will review the proposed simulators in recent years and discuss their capabilities and limitations.

Gem5 [BBB+11] is one of the most widely used simulators in the community since it is cycle-accurate and can simulate complete CMPs ranging from the microarchitectural level up to allowing full-system simulations. Many of the ONoC proposals discussed in this chapter utilise a TDM approach. Therefore, it must be possible to implement contention-resolution schemes, which requires event-based, cycle-accurate simulators to model a NoC accurately. There have been ONoC studies that utilise Gem5 and compute energy requirements of the SiP components through analytical models and technology assumptions [CAJ15] [GPABY16] [GPAY17]. Laer et al. [VLJW13] aimed to

extend Gem5 to support TDM based WRONoCs; however, while their extension was reported successful for up to 16 cores, higher core counts led to unreliable results for which the cause could not be identified due to the high complexity of Gem5 (according to the authors). Any efforts to extend Gem5 for TDM based ONoCs have thus been dropped. Besides, Gem5 is capable to simulate systems up to 64 cores and thus not suitable for studies that aim to investigate NoCs of larger scales.
Other studies evaluating ONoCs based on TDM use in-house simulators (e.g. ocin_tsim [Pra10] [LBGP14] or Phoenixsim [HJH14][CHB+10][RBW+16]) that have not been made available to the public. LioeSim [MYH+14] was another effort to design a platform capable of simulating both electrical and optical NoCs, as well as their interaction. For realistic traffic studies, they support the use of workload traces of real applications, which is a reasonable design trade-off between accuracy and simulation time. However, according to the authors, this project has been terminated, too, and no official version has been released to the public (to the best of our knowledge). Without the availability of such simulation infrastructures, the most reasonable and accurate approach seems to be to utilise analytical power and energy models from demonstrated SiP devices and backend circuitries and to integrate them into cycle-accurate simulators for traditional electrical NoCs – an approach that has been taken by a large number of studies to estimate dynamic, leakage, laser, and MR heating power (i.a. [RGF+14] [DPS+14][DH15b][CAJ15][GPABY16][GPAY17]).
Deploying fixed energy values for SiP devices and circuitries gives a good estimate of the potential of ONoCs in the future. To obtain higher accuracy, however, modelling tools capable of capturing the interaction between transistor technology nodes, SiP technologies, and data rates, as well as power optimisations and trade-offs between SiP devices and driver specifications are needed [SCK+12]. For that purpose, Sun et al. [SCK+12] proposed DSENT, which is the first (and state-of-the-art) NoC modelling tool for energy, power, and area estimations for both electrical and optical NoCs. DSENT is open-source and was shown to provide the highest accuracy and newest technology nodes amongst all available NoC modelling tools [SCK+12]. However, it also has limitations: while capable of accurately modelling all basic bus architectures (see Chapter 2), it cannot model more WRONoC topologies or optical switches. In addition, it is merely a modelling tool and not a network simulator, i.e. on its own, DSENT is not able to extract dynamic power estimates for workload traces. Nevertheless, energy values of the electrical and SiP components can be obtained with DSENT and manually integrated into a network simulator to estimate dynamic power.

The same applies to laser power and MR heating power extracted with DSENT.
TDM and contention-resolution may not always be required in ONoCs (e.g. hybrid NoCs based on buses or contention-free switching topologies). In these cases, it is appropriate to utilise simulators that are not cycle-accurate, and instead use different synchronisation strategies to trade-off accuracy with simulation speed. Graphite is such a simulator and offers a large-scale CMP simulation infrastructure. Most importantly, Graphite integrates DSENT into its simulation framework, which offers accurate dynamic power estimations while allowing for both synthetic and realistic workload simulation (e.g. SPLASH-2 [WOT+95] and PARSEC benchmarks [BKSL08]). If cycle-accurate simulation is not required, Graphite in combination with DSENT most likely represents the most accurate and practical open-source simulation infrastructure for conducting ONoC studies.

3.5.1 Simulation Tools Used in This Thesis

Chapter 4 presents the novel WRONoC switching topology Amon based on destination-reservation, i.e. requires contention-resolution at the destination nodes and in turn cycle-accurate simulation to provide accurate performance estimates. The state-of-the-art control network based WRONoC (i.e. QuT [HJH14]) utilises Phoenixsim, which is not open-source (as discussed earlier). Their Phoenixsim version is based on the widely deployed OMNet++ discrete-event simulation platform [Var99]. We, therefore, opted to use HNOCS [BIZCK12], which is an event-driven, cycle-accurate NoC simulator also based on OMNet++, and used analytical models to estimate laser, MR heating, and dynamic power. This approach has also been taken by QuT and thus allows for a fair comparison.
Chapter 5 discusses a novel approach of combining electrical and optical links in a hybrid NoC design. This proposal, and the alternative NoCs it is compared to, do not require TDM and can be modelled accurately without cycle-accurate simulations. Therefore, the simulation study in this chapter was conducted with Graphite, which is ideal for this purpose and allows to evaluate all NoCs within a CMP simulation infrastructure with accurate power estimations with DSENT.
Finally, Chapter 6 discusses different bus arbitration and bandwidth sharing approaches, which also require contention-resolution. Since there is no need to rely on analytical models to have fair comparisons (like in Chapter 4), the study in this chapter utilises DSENT for all power estimations as it has been argued that it provides higher accuracy than fixed analytical models [SCK+12]. Performance simulations were conducted with

HNOCS since a cycle-accurate, event-based simulator is required to accurately model bus arbitration.

3.6 Other Research in the Realm of ONoCs

The emergence of SiPs has led to a large number of exciting research challenges and opportunities, and is by no means restricted to network architecture. Although this dissertation focuses on NoCs, SiPs is applicable to all scales from NoCs to interconnecting components in 2.5D integrated circuits [TZ14, YGS16, GPABY16], processor-to-DRAM [BJO+09, BSK+10, UMC+10], chip-to-chip [DPS+14], and intra-rack communication in data centres [LLK+10, CHTB11, PKD+10]. A lot of research on the technology-side is dedicated to bringing down laser power and MR heater power. Laser power is targeted by developing off-chip lasers with higher wall-plug efficiencies and coupling devices with lower coupler losses [BCB+14].

On-chip lasers [LSCA+10] have gained much attention since they would eliminate coupling losses altogether, decrease packaging costs, and allow for adaptive laser sources that can quickly be switched on and off based on traffic demands, allowing for adaptive bandwidth scaling [KSC+12, LBLO+14, PTDS15, KK16]. Particularly the latter property has gained much attention due to its large potential of decreasing laser power, and a number of proposals have thus targeted efficient laser control schemes to adapt optical bandwidth dynamically to the NoC demands [DH14, DH16b, DH16a]. In fact, power results of studies targeting adaptive lasers are so promising that we believe that ONoC architectures should be designed so that they can easily be extended to incorporate adaptive laser mechanisms.

Technologies decreasing MR heater power requirements include novel heating techniques [S+15], the utilisation of insulation layers in 3D integrated circuits [DH15a], and athermal devices that are robust to temperature variations [GCL13] [FSB+15].

An increasing number of studies focus on design automation and place&route tools for WRONoCs [BRBS16][PRG+16][OORVYB16][OOTR+17] which were shown to provide significant improvements to IL_{max} and thus laser power.

Chapter 4

Assessing All-optical Network-on-Chip Design

4.1 Introduction

As discussed in Section 3.2.2, WRONoCs (wavelength-routed optical NoCs) based on MR filters and wavelength-selective routing are currently considered the only technologically sensible approach to enable all-optical on-chip communication. In particular, control network based WRONoCs represent the most power-efficient and scalable architecture as they reduce the number of receivers per destination to one, require fewer MRs for wavelength-routing, and decrease the total number of wavelengths in the NoC by reusing wavelength-addresses across multiple destinations. Aside from these benefits, a number of outstanding challenges regarding the switching topology, destination-reservation mechanism, and the laser power distribution network must be addressed to make control network based WRONoCs the preferred choice for future on-chip communication.

A WRONoC's switching topology determines how optical signals are routed through the NoC and has a decisive impact on power consumption. Recently proposed topologies provide large power reductions compared to previous designs [KH12][HJH14]; however, state-of-the-art WRONoCs still require significant amounts of MR heating and laser power which must be further lowered to make them a power-efficient and practical design solution. In particular, novel topologies are necessary that further minimise optical path losses, the number of MRs for wavelength-routing, and the total number of wavelengths in the NoC.

The destination-reservation performed on the control network prior to data transmission is required as each destination has only one ejection channel, and largely determines latency and throughput – particularly for workloads exhibiting traffic hotspots. State-of-the-art approaches [KH12][HJH14] make senders back-off and retransmit control packets if a destination is occupied, which can lead to unnecessary waiting times that seem inefficient for NoCs that typically require low-latency communication. Evaluating the performance improvements of advanced destination-reservation mechanisms for realistic application workloads is particularly important as this has been neglected by the previous studies that propose the state-of-the-art designs.
As discussed in Section 3.2.2, these proposals also neglect to study a LPDN (laser power distribution network) and report laser power only for the highest optical path loss and assume one laser provided to each node. Laser source coupling, however, is a critical cost factor in the packaging process and the number of laser sources is typically low. To propose WRONoCs that are more practical (and realistic), a separate LPDN should be considered and analysed along with its path losses and layout.
Throughout this study we observed that the injection and ejection backends of the switches typically used in WRONoCs are the main limiting points of the design as a single ejection channel drastically reduces throughput, while having multiple injection channels increases the power budget significantly without providing any substantial benefits. Exploring novel architectures modifying the switch backends could therefore improve the overall power efficiency of WRONoCs.
This chapter investigates all of these research questions and makes the following novel contributions:

- 'Amon', a novel control network based WRONoC including all design aspects, i.e. a VLSI friendly grid-like switching topology, routing algorithm, practical switch design, and a detailed analysis of the LPDN and its impact on the overall NoC. Amon reduces laser power by 21% and MR heating power by 16% compared to the state-of-the-art proposal QuT [HJH14] for 64 nodes.

- A novel destination-reservation mechanism that improves the one proposed in QuT by simplifying the transmission protocol *and* parallelising data and control transmissions to maximise throughput. This proposal improves throughput by 40% on synthetic workloads and latency by 50% on PARSEC workloads, while saving 45% dynamic power by discarding negative acknowledgements and packet retransmissions.

- A backend extension that adds ejection channels to each switch to allow nodes to receive optical signals from each incoming link simultaneously rather than from just one source at a time. This approach resolves the susceptibility of control network based WRONoCs to traffic hotspots and reduces packet latency by 50% on PARSEC traces (on average). Power overheads of these extensions are negligible, even if added homogeneously to each node in the NoC ($< 0.1\%$).

- A mechanism in the switch backends of Amon that leverages MR tuning to dynamically select the injection waveguide prior to data transmission, which reduces the number of injection channels from four to one and, in turn, achieves power reductions by 43% and 60% for conservative and aggressive SiP technology parameters, respectively. Although increasing packet latency by 20% (on average for PARSEC traces), latency is still much lower than in QuT and a 2D Mesh (75% on average).

- The proposed destination-reservation mechanism and the two backend modifications of the injection and ejection channels are investigated for Amon, but are applicable and equally significant to other WRONoC architectures.

4.2 AMON: An Advanced Mesh-like Optical NoC

Amon is an all-optical WRONoC design that consists of a data network and a control network for data transmission and destination reservation, respectively (like the NoCs discussed in Section 3.2.2). Once a destination is reserved, data is transferred through the network according to a deterministic routing algorithm which is implemented in hardware with MR filters that will forward optical signals to their destinations. Data reservation is necessary because each node in Amon has only one ejection channel, i.e. can only receive data from one sender at a time. Amon adopts many efficient attributes of QuT – the state-of-the-art control network based WRONoC design – but offers a topology that requires fewer MRs, reduces path lengths, and features a more efficient destination-reservation mechanism. This section first introduces Amon's data network, and along with that its routing algorithm and switch microarchitecture, followed by a discussion on the laser power distribution network and control mechanisms.

4.2.1 Data Network

Logical Topology

Senders modulate data packets on the wavelength (λ) or on a set of multiple λ (λ-set) that is assigned to the destination that shall be addressed. The modulated λ-set will then be routed through the network and filtered at the destination. The λ-set size determines the number of λs on which senders can modulate their data (e.g. 8λ, 16λ, etc.), and in turn link bandwidth on the data network. Like in QuT, the number of λ-sets for addressing is split into $N/4$, i.e. four destinations in the NoC are addressed by the same λ-set, in order to maintain QuT's design efficiency regarding laser power and number of required laser sources.

Figure 4.1 shows the topology of the data network for 48 nodes. Amon consists of four *Submeshes* in which each node has a unique λ-set address within its Submesh. λ-sets for addressing are re-used across the Submeshes. All links are bidirectional (implemented with two separate waveguides) unless indicated by arrows. As four nodes share the same λ-set address, up to four nodes could be transmitting data modulated on the same λ-set simultaneously through the network.

Optical signals in NoCs like Amon would collide if nodes transmitted data simultaneously on the same λs on the same path or if their paths were crossing. In Amon, collision-free paths are always guaranteed as there is only one λ-set address per Submesh, and physically separated waveguides to send data between Submeshes. For that purpose, Amon relies on the following different links:

Mesh links (shown in black): are the links used for forwarding optical signals within a Submesh.

Intermesh links (shown in blue): are used for data transmission to nodes in other Submeshes. They use separate waveguides parallel to the Submesh links in the sender's Submesh, and logically merge into the Submesh links of the destination's Submesh.

Amon's data network can be scaled very flexibly as it does not require a square mesh. Its only limitation is that all Submeshes must have the same size due to row and column connections through the Intermesh links. Therefore, the 48-node Amon in Figure 4.1 could also be composed by 3×4, 6×2 or 2×6 Submeshes. The following section will illustrate how the different links in Amon are used to implement Amon's collision-free routing algorithm.

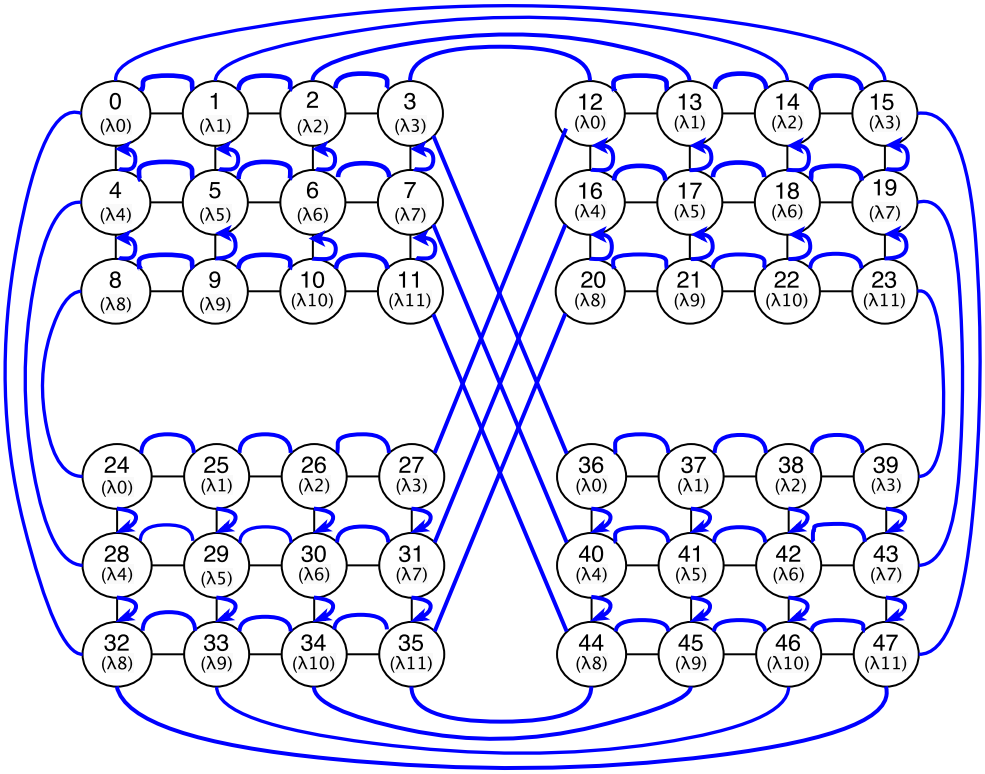


Figure 4.1: Data network topology for a 48-Node Amon with 4×3 Submeshes.

Wavelength Routing

Amon uses MRs to route optical signals according to their wavelength. Therefore, all routing paths in Amon are *predefined* and effectively perform deterministic routing. Data is injected into the network based on the relative position between the sender and receiver. From a sender's perspective, a destination can either be

In the same Submesh: packets will be injected into the local Submesh and routed using static dimension-order routing (DOR).

In a different Submesh: packets will be injected into the Intermesh links leading to the destination's Submesh. Once in the desired Submesh, the local Submesh is used as above.

This simple behaviour is directly implemented in hardware (see the router architecture in the next section) and requires very simple computation at injection time: λ-set = *destination_id* % (N/4) and Submesh = *destination_id* / (N/4). With typical power of 2 number of nodes, this can be further simplified to bit operations (the 2 most significant bits define the Submesh and the remaining bits define the λ-set). Figure 4.2 shows routing examples on a 64-node Amon:

(i) **Red path** – Node 43 (N_{43}) sends to Node 10 (N_{10}): N_{43} injects data into its Intermesh link to the west and on the λ-set λ_{10} to address N_{10}. Once the optical signal reaches N_{10}'s Submesh, the MRs at N_6 will route the signals down in Y-direction to N_{10}. Node 10 has MR filters that respond to λ_{10} at its ejection channel to eject the optical signal from the network into its photodetector.
(ii) **Green path** – Node 0 (N_0) sends to Node 63 (N_{63}): N_0 modules data on λ_{15} to address N_{63} on the Intermesh link leading to N_{63}'s Submesh, i.e. the Intermesh link to the east. Again, once the Submesh is reached, the optical signal is forwarded in DOR fashion, leading it to N_{63} where it will finally be ejected.
(iii) **Pink path** – Node 20 (N_{20}) sends to Node 31 (N_{31}): routing within the Submeshes is as explained in the previous examples: N_{20} injects its optical signal on λ_{15} on the Submesh link to the east. N_{23} implements MR filters to drop the optical signal to the 2D Mesh link to the south where it finally will be ejected by N_{31}.
As we will show later in more detail when we discuss Amon's physical layout, switching optical signals with MR filters significantly adds to IL_{max}, and thus laser power. Therefore, one goal is to minimise the amount of switching required between any source-destination pair. We achieve this by implementing DOR as follows: in (i), the Intermesh link leading from the south-west to the north-west Submesh does not stop at N_4: it goes straight through the Submesh, all the way to N_7. This way, there only needs to be MRs switching the signal either up or down, depending on the λ-set. For instance, N_6 implements MR filters to route λ_{10} down to N_{10}. This means that Intermesh links from a Submesh are extended to be the Submesh links in another one (see Section 4.2.1 for more details). The logical distinction is made for the sake of simplicity.

Router Microarchitecture

This section deals with the different types of MRs required to enable the routing algorithm and give a close-up of two representative switch designs for illustration. Each switch in Amon has three different types of MR filters that serve different purposes:
Injection MRs (I) enable nodes to inject modulated optical data signals into the network. A node needs injection MRs to send data to nodes residing in the different Submeshes, as well as in its own Submesh. Therefore, it must be possible to inject data into the Intermesh links and the Submesh links. Each node thus needs three injection MRs for each Intermesh link, and two-to-four injection MRs for Submesh links depending on its location within its Submesh (i.e. corner vs. middle node).

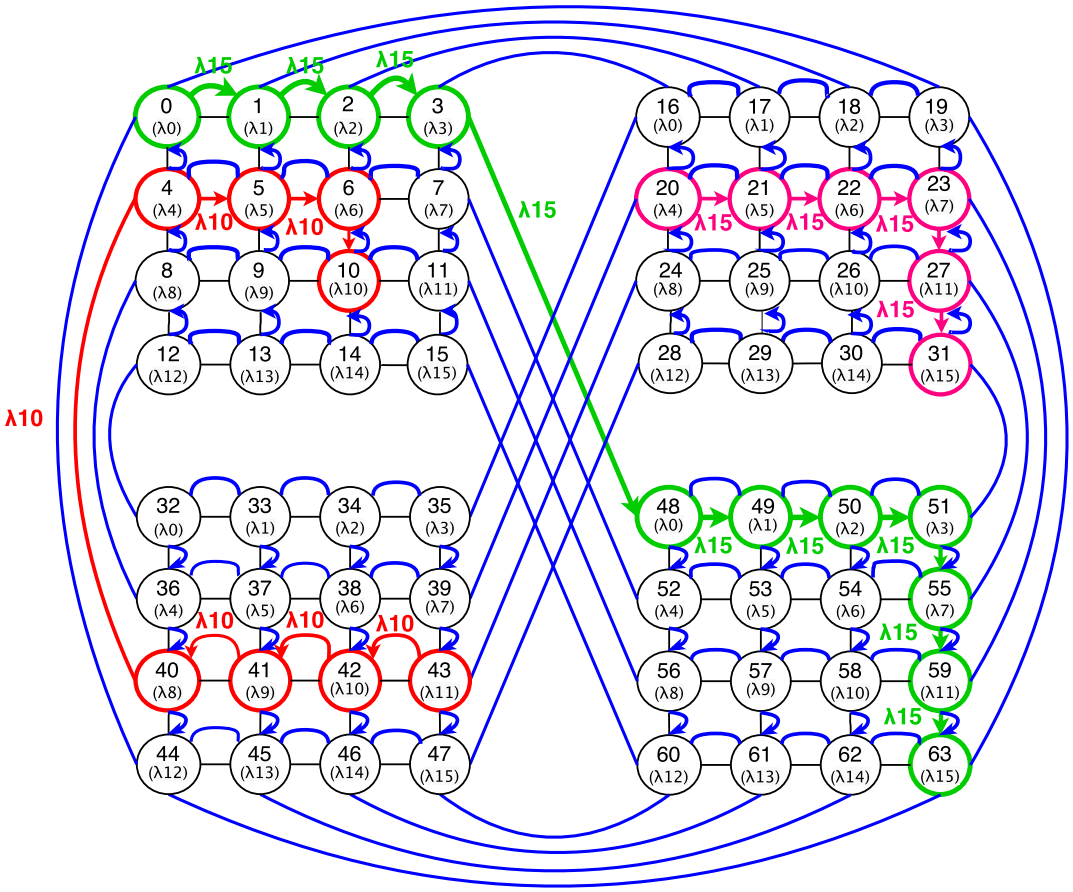


Figure 4.2: Wavelength Routing Examples in Amon

Ejection MRs (Ej) allow to eject the optical signals of a node's λ-set address. Ejection MRs are placed on the Submesh links entering from each cardinal direction (if applicable) to eject the signals from each possible direction.

Switching MRs (S) perform the actual routing through the network by providing the necessary turns for the optical signals as described in the routing algorithm. Therefore, they are strategically placed between waveguides to drop optical signals from one waveguide to another.

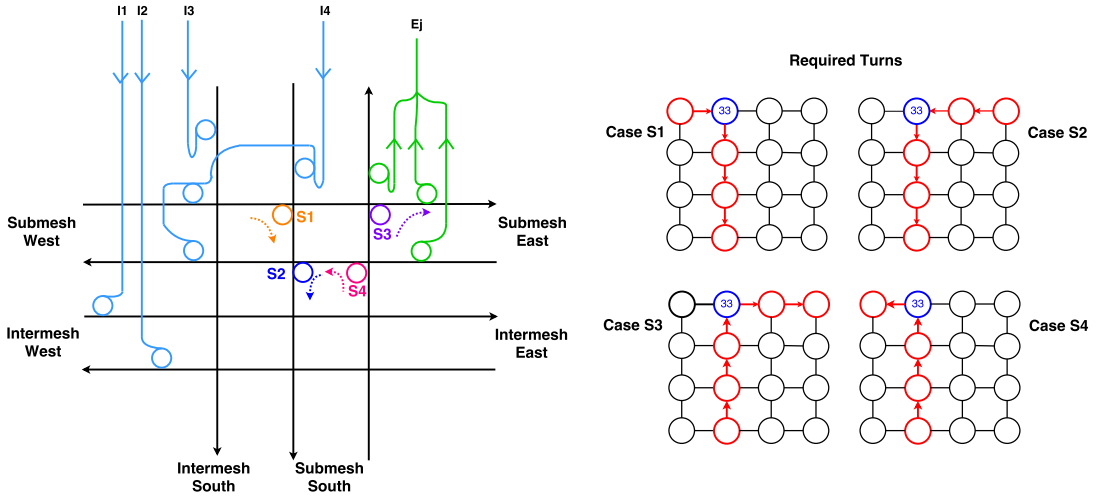
Example Switches There are two different basic switch designs in Amon:

i) Switches located in the bottom row of the top Submeshes (e.g. Figure 4.2 Nodes

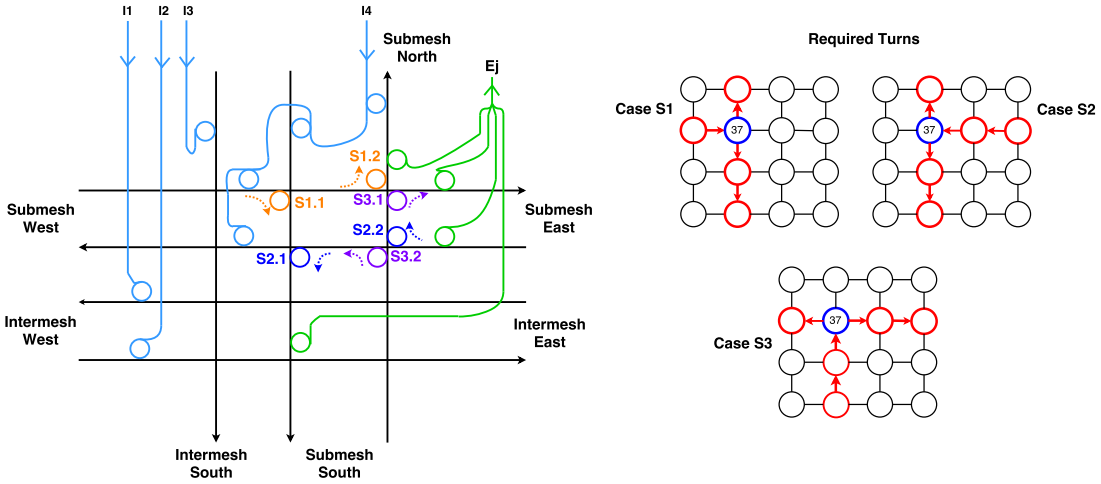
12/13/14/15 and 28/29/30/31) and the top row of the bottom Submeshes (e.g. Figure 4.2 Nodes 32/33/34/35 and 48/49/50/51). These switches have incoming and outgoing links on three of the four cardinal directions.
ii) Every other switch has incoming and outgoing links from every cardinal direction.
The design of each switch can be inferred based on the following two switch designs by adjusting the MR filters based on the location within a Submesh. Since Amon's topology is symmetric, the switching in each Submesh is the same, just mirrored.

Switch 33. Figure 4.3a illustrates a close-up of Switch 33 in a 64-node Amon example, which is located in the top row of the Submesh in the south-west (see Figure 4.2). The injection MRs are drawn in blue. As described above, Node 33 must be able to inject its modulated optical signals into each of the Intermesh links (I1, I2, I3) and Submesh links, and thus places MRs on the corresponding waveguides. In addition, ejection MRs must be placed on each link entering from each cardinal direction. Since Switch 33 resides in the top row of the Submesh, it only needs three ejection MRs as there is no incoming link from the north. S1, S2, and S3 represent the MRs that provide the turns necessary in this switch, which are illustrated at the right-hand side in Figure 4.3a:
S1 provides the turns of optical signals entering the switch from the west that are modulated on λ-sets of destinations located below Switch 33, i.e. λ_5, λ_9, and λ_{13} for Node 37, 41, and 45, respectively.
S2 provides the turns of optical signals coming into the switch from the east destined for nodes located south of Switch 33 (same λ-sets as in S1).
S3 provides the turns for optical signals entering the switch from the south destined for nodes located east of Switch 33, which are modulated on λ_2 and λ_3 for Node 34 and 35, respectively. *S4* provides the turns for the optical signals incoming from the south and destined for Node 32, i.e. λ_1.

Switch 37. Figure 4.3b illustrates a close-up of Switch 37 in a 64-node Amon. Similar to Switch 33, injection MRs must be placed on all Intermesh and Submesh waveguides so that Node 37 can sent data to each node in the network. Ejection MRs are placed on the Submesh waveguides of each cardinal direction. Switch 37 is located in the middle of the Submesh and must, therefore, provide more turns than Switch 33 (as it is not located in the top row). The following MRs are necessary to implement the switching:
S1.1 provides the turns for case S1, i.e. signals entering from the west for nodes located south to Node 37, i.e. it filters λ_9, and λ_{13} for Node 41 and 45, respectively.



(a) Switch 33 (in Submesh South-West)



(b) Switch 37 (in Submesh South-West)

Figure 4.3: MR Switching in a 64-node Amon

S1.2 provides the turns for case S1, i.e. signals entering from the west for the node located north to Node 37, i.e. it filters λ_1 for Node 33.

S2.1 provides the turns for case S2, i.e. signals entering from the east for the nodes located south to Node 37, i.e. it filters λ_9, and λ_{13} for Node 41 and 45, respectively.

S2.2 provides the turns for case S2, i.e. signals entering from the east for the node located north to Node 37, i.e. it filters λ_1 for Node 33.

S3.1 provides the turns for case S3, i.e. signals entering from the south for the nodes located east to Node 37, i.e. it filters λ_6 and λ_7 for Node 38 and 39, respectively.

S3.2 provides the turns for case S3, i.e. signals entering from the south for the node

located west to Node 37, i.e. it filters λ_4 for Node 36.

Which λ-sets are filtered by these MRs thus depends on the turns that must be provided and the location of the switch in the Submesh. The same applies to the injection MRs: each injection MR that provides the injection to nodes located in another Submesh must be able to filter all λ-sets in that Submesh (i.e. N/4) since the node must be able to address each destination in that Submesh. For injection within the Submesh, however, the number of injection MRs depends on the relative location of the switch. For instance, in Switch 37, the MR placed on the Submesh waveguide to the east for I4 also only needs to filter the λ-set of the nodes located east to Node 37, which is just Node 36 (i.e. λ_4).
In total, each node needs injection MRs for the λ-set of each node in the NoC apart from itself (i.e. $(\lambda \times (N-1))$). Ejection MRs have to filter out the λ-set representing the node's address, and therefore only λ MRs are required per ejection point.

Layout

Figure 4.4 illustrates an example layout of Amon's data network for one Submesh. Since Amon is symmetric, the whole layout can simply be obtained by placing the same waveguides for every other Submesh. Intermesh links (in red) from all other Submeshes must be routed to the destination Submesh, in which they will logically become Submesh links (although physically being the same waveguide). We illustrate this by changing the colour of the waveguide from red to black. In addition to the Intermesh waveguides, two more waveguides in the middle columns of the Submesh must be routed from the north to the south since there is no Intermesh link entering the Submesh from the north that could provide the Submesh waveguide to route packets to the south.
The Submesh links in the left and right most columns leading from the north to south are provided by the Intermesh links entering the Submesh from the west and east, respectively, that originated from the top Submeshes. For instance, the path from Node 19 to 44 consists of just one waveguide, namely the Intermesh link originating in Node 19. When that waveguide reaches Node 32, a turn must be provided to the south of the Submesh, leading to Node 44. This is more efficient than dropping the optical signal from the Intermesh link to a separate waveguide leading the signal to the south since dropping a wavelength introduces considerably more loss than a bend (up to 100× [HJH14]). Moreover, at Node 32, the only turns a signal coming from the east could

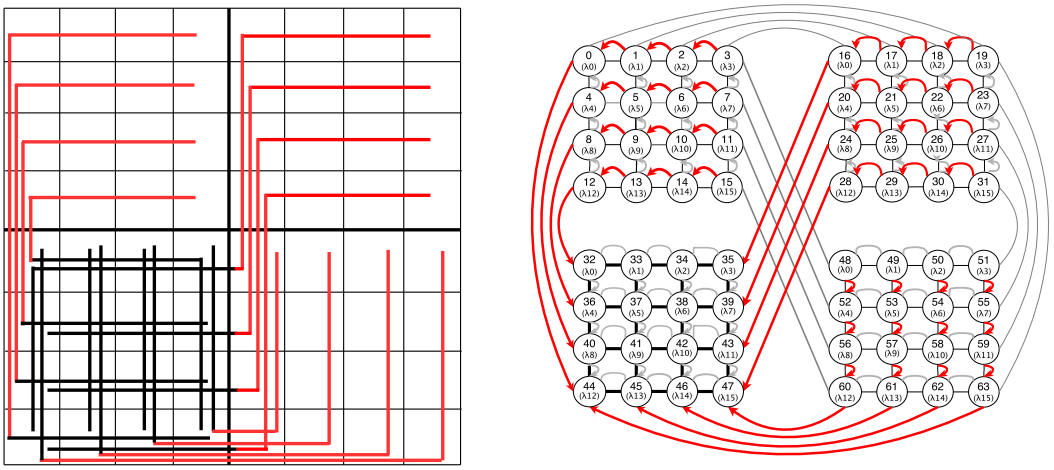


Figure 4.4: Waveguides in the physical layout for Submesh South-West in Amon

take at this point is to the south. This is, for instance, not the case at all other nodes on that path, i.e. 33, 34, and 35: at these points, the signal could either go straight through the switches or being switched to the south.

4.2.2 Laser Power Distribution Network

Previous studies neglect to provide a sample layout, or even mention, the LPDN, i.e. the network on which the light is distributed to all nodes; however, most recently published work has shown that the LPDN plays a significant role to the overall power consumption in WRONoCs and deserves detailed attention [OOTR+17]. As mentioned above, a LPDN is theoretically not necessary if each node can be supplied with one dedicated laser source; however, in NoC sizes of moderate to large scale (e.g. 64 nodes), this would require a large number of lasers to be coupled into the chip, which is impractical, complicates layout [HJH14], and causes high cost overheads in the packaging process. In fact, NoCs should always aim to work with as few laser sources as possible to minimise cost. Having fewer laser sources – ideally 1 – is the more realistic design point and requires to distribute the light from the entry points into the chip to the nodes' injection channels over the LPDN.

Just like in electronics, it was shown that design synthesis tools can reduce path lengths and layout significantly, leading to optimised designs with considerably decreased IL_{max} and in turn laser power (up to 94% reductions for some topologies [BRBS16]).

Both Amon's data network and LPDN should, therefore, be considered a feasible design point rather than an optimised layout; however, a layout proposal is generally necessary to estimate the insertion losses in ONoCs.
Figure 4.5 illustrates a potential LPDN to provide each node with light. The wavelengths can either be provided by a multi-wavelength comb laser or by an array of single-wavelength lasers, depending on how much output power per wavelength is required by the NoC and can be supplied by the laser. To minimise the path from the coupling point of the off-chip laser to each node (and their injection channels), light is coupled into the middle of the chip into what is referred to as a 'power waveguide' and distributed in an H-tree fashion like in clock distribution networks. Optical splitters distribute light by splitting it over multiple waveguides, thereby incurring splitting losses of 0.1-0.2 dB [HJH14] [GMS+14]). In addition to that, splitting light from one waveguide to two output waveguides halves the optical power going down each path for 50:50 splitters, which translates to a 3 dB (50%) signal degradation.
Ideally, optical splitters would split the power ratio going down each of the links based on the actual power requirements of each link; however, Ortín-Obón et al. [OOTR+17] note that using different splitting ratios based on the power requirements of each link does not necessarily lead to the most power-efficient design. In addition, the authors note that highly unbalanced splitting ratios may result in issues during the manufacturing process caused by process variations, and suggest to opt for the most common and reliable splitting ratio for light distribution, which is 50% down each path (i.e. 3 dB signal degradation). Note, however, that efficient place&route tools for WRONoCs and LPDNs are an active research area [OOTR+17] [BRBS16] which may provide optimised designs in the future.

4.2.3 Control Network

As mentioned earlier, Amon's data network must be supplemented with a control network for destination-reservation since each node has only one ejection channel, meaning that it can only receive data from one sender at a time. While this architectural approach allows for large resource and power savings in the data network, it requires to prevent multiple senders from transmitting data to the same destination at the same time. For that reason, nodes must first check the availability of a node on the control network which should ideally impose as little power overheads as possible and minimise the latency overheads incurred by destination reservation.
Early proposals envisioned to perform destination-reservation on an electrical control

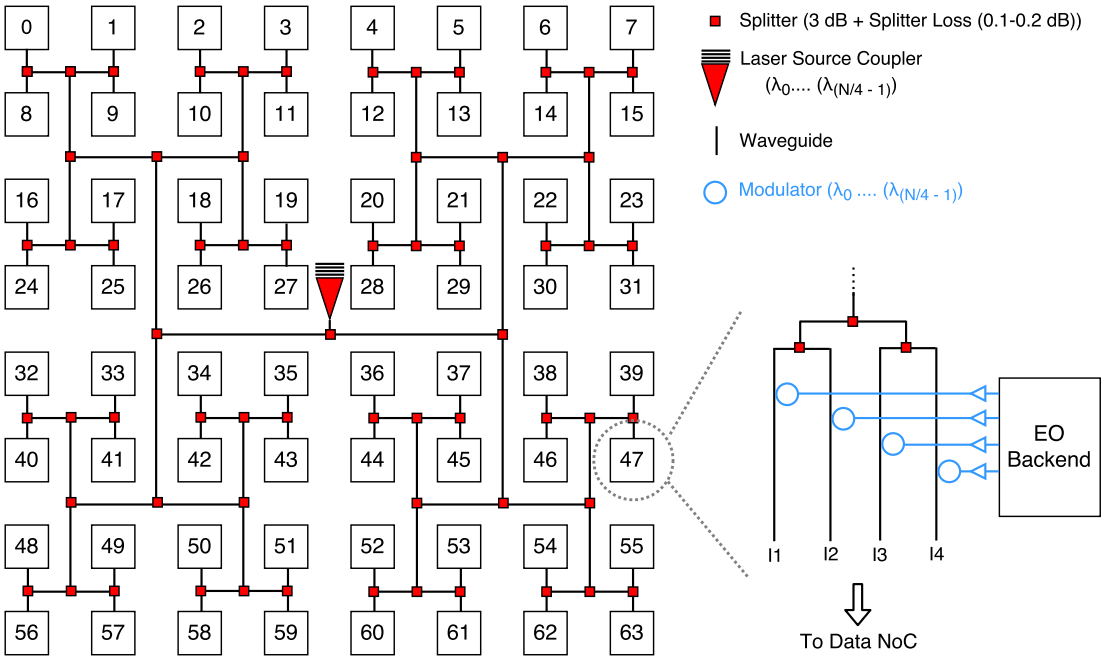


Figure 4.5: Light Distribution in Amon for the Data Network

network [KH12]; however, destination-reservation occurs globally, i.e. each node must be able to check the availability of each destination in the NoC, which incurs both latency and energy overheads when executed electrically, effectively cancelling out any performance gains on the data network. Control packets are significantly smaller than data packets, which allows to implement the control network with much less bandwidth. This combined with the weak distance dependency of optical data transmission in terms of latency and energy makes optical links very suitable to implement a global control network.

Nevertheless, the control network has a critical impact on the overall performance of the total NoC and must provide global contention-free all-to-all communication to enable each sender to check the availability of a destination at any given time. The excessive laser power of global crossbars, however, puts hard limits on the bandwidth that each node can use to modulate control packets. Performance improvements must thus originate from smart control mechanisms to overcome this lack of bandwidth scaling on the control network.

The next section first introduces QuT's control network and destination-reservation mechanism, followed by the presentation of novel modifications and extensions to this mechanism in order to improve performance and lower power without resource overheads.

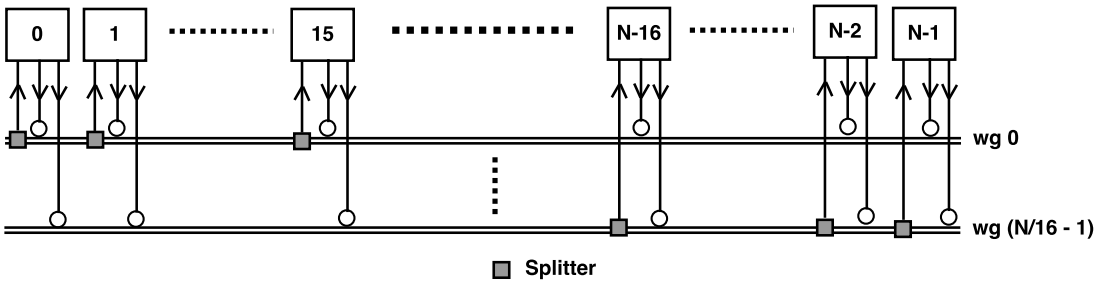


Figure 4.6: Control Network Design of QuT [HJH14]

QuT Control Network

The authors of QuT proposed an all-optical control network based on MWSR buses, as depicted in Figure 4.6. Their design implements N/16 waveguides, and 16 nodes receive data from each waveguide which is guided to them using optical splitters. Each node can send data to every destination by modulating its control packet on the waveguide to which the destination is connected. To do that, each node has modulators placed on each waveguide in the control network.
The control packets exchanged in QuT consist of requests (REQ), acknowledgements (ACK), and negative ACKs (NACK). If a node wants to transmit data to a destination, it modulates a REQ on its assigned λ on the MWSR bus connected to the destination. Upon reception of a REQ, a destination will either reply with an ACK in case it is free or with a NACK otherwise. If the sender receives an ACK, it will start with the data transmission. If a NACK is received, QuT implements a retransmission scheme in which a subsequent REQ will be sent after the sender has waited for a back-off period determined by the average number of cycles required to transmit a data packet.
Since 16 nodes receive from each waveguide, the sender must encode the destination ID into the control packet, which consists of 4 bits ($\log_2 16$). In addition, the control packet type must be encoded, requiring another 2 bits for the three packet types (REQ, ACK, NACK), leading to a total of 6 bits per control packet. Therefore, each node must provide buffer space of $(N-1) \times 6$ bits to be able to receive a REQ from each node in the NoC simultaneously. Besides, QuT assumes one wavelength for modulating control packets at each node to keep laser power at acceptable levels.

Amon Control Network

Designing an optical, low-overhead control network for contention-free communication is challenging, and QuT's control network based on MWSR buses and splitters is

an efficient trade-off between number of waveguides and power consumption. Amon thus adopts the physical implementation of the control network as shown in Figure 4.6. Although low-power, however, this structure may cause performance bottlenecks: for instance, in the 64-node case, each node has one waveguide to send REQs, ACKs, NACKs, and REQ retransmissions to 16 nodes. In traffic patterns (or multicast traffic) where control messages have to be sent to nodes that are all addressed on the same waveguide, this would require to serialise all messages on little optical bandwidth (1λ) since the control network should be low-overhead. Since this architecture is very susceptible to adversary traffic, the control packet mechanisms should be improved to reduce latency overheads on the control network.

Removing NACKs and REQ Retransmissions A sender would not request a destination for data transmission if it has already sent out a REQ to this destination for another packet and has not received an ACK back yet. Therefore, buffer space is accounted for to hold *one* alive REQ for each potential sender in the NoC, which does not require much buffering since REQs are small (6 bits). In addition, as described above, each destination must provide buffer space for just (N-1) REQs because it will never receive more than one REQ per sender. Based on these conditions, there does not seem to be a real need to reply to a sender with a NACK. Therefore, we modify QuT's mechanism by removing NACKs entirely. Destinations will simply keep alive REQs in their buffers, and send out ACKs once it is free again. While this does not cause any overheads to QuT's scheme, it has two benefits: first, requesters will receive an ACK at the earliest point possible rather than waiting for a back-off time and retransmitting the REQ, which may cause unnecessary latency. Second, discarding NACKs and REQ retransmissions reduces both energy and network contention on the control network (and in turn latency) as fewer packets are exchanged. Besides, 1 bit can be saved in the control packets as there are only two packet types now (REQ/ACK).

Parallelising Control Packet Transmission In addition to this described (potential) improvement, saving 1 bit in the control packets is not significant, and that bit can instead be used to further improve latency on the control network: since the bandwidth on the data network is fixed, each node knows how long data transmission will take for a given packet size. This knowledge could be used at the destination to start sending out ACKs in parallel to receiving data: once a destination receives the first flit or

starting sequence of a packet, it knows how much longer packet reception will take if it knows the packet size. In addition, it can also estimate how long it takes to transmit an ACK since bandwidth on the control network is fixed at design time and the ACK size known. A destination could, therefore, start transmitting an ACK in parallel to receiving data since the data and control network are two separate networks that do not interfere with each other. This allows to effectively hide the entire delay that the ACK packet would incur on the network.
Information about the packet size must be added to the REQ to calculate the transmission latency of the data packet. NoCs must typically support two packet sizes in CMPs (coherence traffic and cache line transfers), which would just require one additional bit in the REQ packets. The control packet size would, therefore, not increase compared to QuT's mechanism. With this information, ACKs can be sent earlier so that the requester receives it in the same cycle data transmission as the currently received packet at the destination finishes. This approach allows to hide the ACK delay completely and ensures that the current and following data transmissions do not interfere with each other. This improvement could lower congestion in the NoC, particularly in traffic patterns in which some destinations receive significantly more traffic than others.
The same improvement can be attained by parallelising REQ messages: for instance, imagine the case in which a node has two packets for the same destination in its output buffers and just received an ACK for its first packet. Waiting to send the REQ for the next packet until the data transfer of the first packet is finished causes unnecessary latencies in this case since the next REQ could be sent out immediately on the control network while data transmission occurs on the data network. This mechanism is added to parallelising the ACK messages at the receiver in order to lower latency and congestion at the senders. Note that the authors of QuT do not mention whether they parallelise REQs or not, so it may be possible that this mechanism was already considered in QuT. The effect of parallelising REQs and ACKs on the total performance will, therefore, be studied separately in the following section.

4.2.4 Evaluation

Methodology

This study compares Amon to QuT as the latter constitutes the most efficient control network based WRONoC design in literature and was shown to outperform a number

of previously proposed ONoCs. The goal is to evaluate to which extent Amon can decrease 1) MR heating power by requiring fewer MRs and 2) laser power by reducing path losses. In addition, we are interested in how the improvements of the proposed destination-reservation mechanism translate to latency reductions and throughput gains. To study scalability, we consider both NoCs for 64 and 128 nodes. If electrical interconnects shall be replaced entirely by all-optical NoCs, it must be shown that the latter is capable of significantly outperforming an aggressive electrical baseline to justify a complete technology shift. Therefore, this study is complemented with a 2D electrical mesh topology (2D Mesh) – as it is deployed in many commercial products (i.a. [VHR$^+$08][BEA$^+$08]), – with 64 bits path width and aggressive latency values of 1 cycle link and 2 cycle router traversal at 5 GHz.

In this study, NoCs were simulated using HNOCs [BIZCK12]. To gain insight into the NoC's performance bottlenecks, the NoCs were stressed with various different synthetic traffic patterns to evaluate latency and throughput behaviour under different network loads, as well as application traffic traces to study latency on more realistic network utilisation scenarios.

Synthetic traffic patterns include uniform random, bit complement, hotspot, and neighbour traffic to stress the different corner cases of the topologies. In bit complement traffic, destination coordinates are the bit-wise inversion of the source coordinates, i.e. each sender sends all its traffic to one destination. In hotspot traffic, one node receives 30% of all the traffic, while the remaining traffic is uniformly distributed amongst all remaining nodes. In neighbour traffic, a node randomly sends to one of the adjacent nodes in a tile-based chip layout. All traffic patterns were simulated for varying injection rates with an exponentially distributed inter-packet gap. Packet sizes in synthetic traffic are 256 bit for each data packet.

For realistic traffic, the NoCs were evaluated under applications from the PARSEC benchmark suite [BKSL08], which represents a collection of heterogeneous multithreaded applications for CMPs spanning various application domains (e.g. financial analysis, media processing, computer vision, etc.) and thus represent diverse workloads. Traces were collected with Netrace [HGK10], a tool that gathers traces with dependency tracking in a full-system simulation environment of a CMP with 64 in-order cores with 32 kB private L1I/L1D caches, a shared 16 MB low-level cache, MESI coherence protocol, and for an electrical 2D Mesh NoC with a hop latency of 3 cycles. Traces of the PARSEC applications were captured during PARSEC defined region of interest, i.e. the parallel portion of the applications after caches have been warmed up.

A λ-set consists of *eight* wavelengths for addressing/data transmission on the data network, and each source has *one* wavelength for modulation on the control network, which has been identified as an efficient design point in QuT. We assume a 5 GHz core clock frequency, 10 Gb/s modulators/detectors, and a tile based layout with 1 mm tile dimensions (8×8 for 64 nodes, 8×16 for 128 nodes).

Performance

We compare the latency of QuT and 2D Mesh to an Amon implementation with three different control mechanisms: *Amon_seq* denotes an Amon implementation without starting ACK transmission in parallel to receiving data packets. This allows to evaluate the benefits of discarding the retransmission scheme in QuT and just keep REQs and transmit ACKs once the destination is free again. *Amon_par_ack* utilises the parallel ACK transmission scheme described in the previous section, and *Amon_par_req_ack* parallelises both REQs and ACKs.
Packet latency is measured from the time a packet is injected into the source node until it is fully received by the destination. Optical transmission delay includes the serialisation delay in the EO backends, waveguide propagation delay (10.45 ps/mm [HCC+06]), and delay through the OE backend (1 core clock cycle [KMP+10]).

Synthetic Traffic Figures 4.7 and 4.8 show the average packet latency for the synthetic traffic patterns for 64 and 128 nodes, respectively. Packet latency is mainly determined by the time a packet waits in the output buffers of the sender and by data modulation, where the latter is the major contributor at low injection rates. With 8λ on the data network, 5 GHz core clock, and 10 Gb/s modulation speed, a data packet has a modulation delay of $256/16 = 16$ clock cycles. As network traffic increases, so does contention at the destination nodes, leading to increasingly long waiting times in the output buffers of the senders.
Amon improves latency and throughput compared to QuT for all traffic patterns and both network sizes significantly. For low injection rates, the latency benefits are small since modulation of the data packet requires the most latency, and hardly any retransmissions occur in QuT due to low contention on the network; however, for increasing injection rates, latency gains become more eminent, and the latency reductions on the control network shifts the saturation point significantly, allowing for higher throughput on the NoC. Indeed, we observe that parallelising the transmission of the ACK messages has the most significant impact on the traffic the NoC can sustain. For all

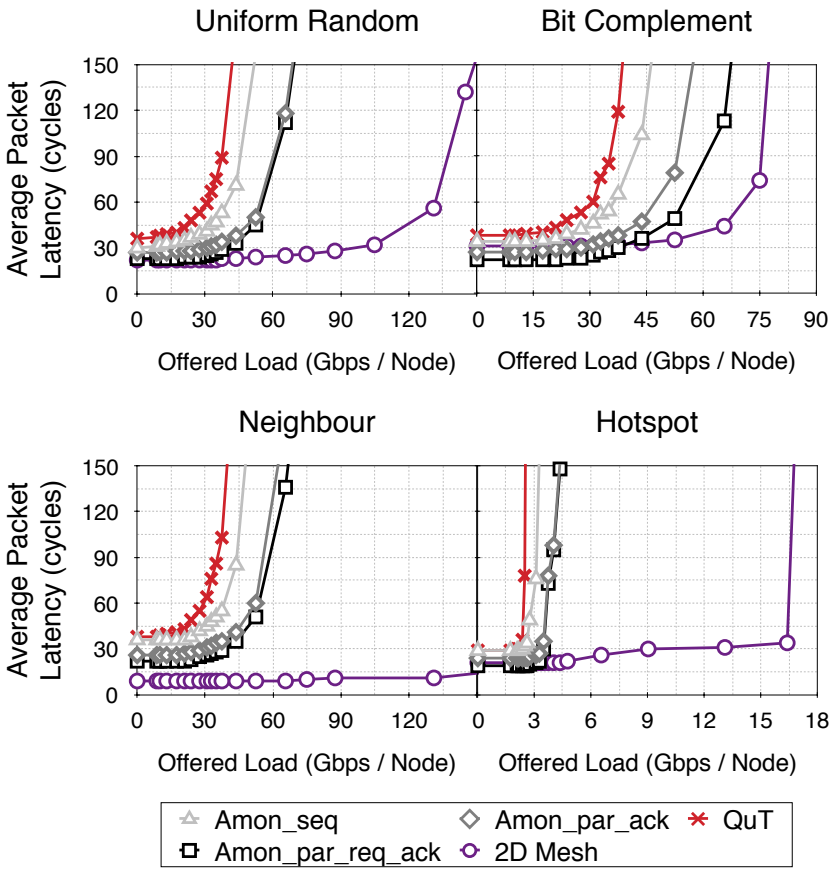


Figure 4.7: Average packet latency for synthetic traffic for 64 nodes

patterns apart from bit complement, parallelising REQs has little impact as the likelihood of a sender to send two subsequent packets to the same destination is low. Parallelising the REQ messages has a decisive impact in bit complement traffic, where each node always sends to the same destination, and is mainly responsible for the attained throughput gains. Parallelising ACKs in this traffic pattern only has little impact since there is no contention at the destination nodes. For all traffic patterns, Amon improves throughput by ∼40% compared to QuT for both network sizes, while offering lower latency for all injection rates (at least 20%).

Hotspot is the most adversarial traffic type for Amon/QuT as a high amount of traffic is sent to just one destination, leading to high contention [HJH14]. Consequently, network saturation occurs significantly earlier than in all of the other traffic patterns (for instance, Amon saturates at 3.5 Gbps/node in hotspot, and at 53 Gbps/node in random traffic for 64 nodes). The suitability of these NoCs for application traffic in which a lot of many-to-one traffic occurs is therefore questionable, and the most benign case is when traffic is evenly distributed (as this leads to minimal contention at

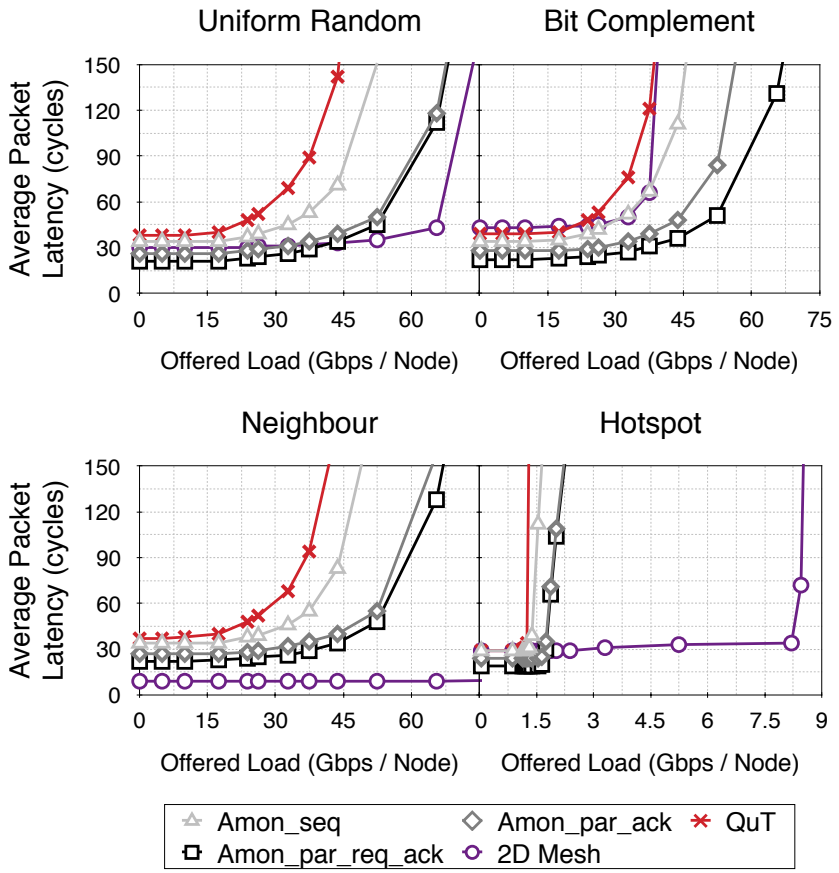


Figure 4.8: Average packet latency for synthetic traffic for 128 nodes

the destination nodes). Although Amon's novel mechanism on the control network can noticeably improve throughput for hotspot traffic, when compared with QuT's, the achieved throughput is still rather low. As we will discuss later on, this has a great impact on realistic application traffic and requires careful consideration at the architectural level.

Compared to the 2D Mesh, Amon can provide competitive latency for low injection rates, but saturates earlier in all traffic patterns. The 2D Mesh is particularly superior for neighbour traffic as it does not require destination reservation or EO/OE conversions (note that the 2D Mesh saturates at ∼350 Gbps/node).

Both latency and throughput scales well in control network based WRONoCs since data communication takes place optically: large distances can be traversed within 1-2 clock cycles for the assumed tile widths of 1 mm. In addition, contention occurs at the destination nodes, and not within routers (as typical in electrical NoCs). We observe similar latency and throughput levels for both network sizes in all traffic patterns apart from hotspot traffic in which a higher number of nodes send to the hotspot node as

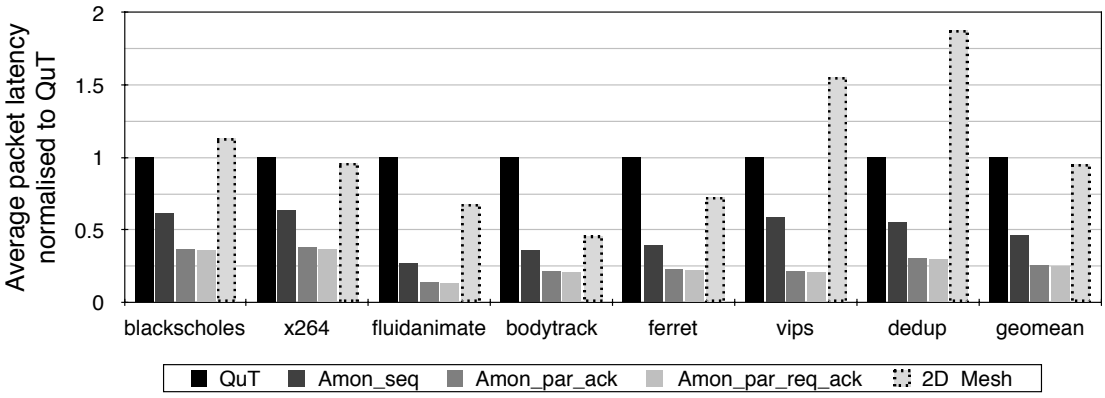


Figure 4.9: Average Packet Latency for PARSEC Workloads

the network size increases, which leads to earlier saturation. The 2D Mesh becomes increasingly inferior for larger network sizes as it requires more hops on average to reach a destination. Consequently, Amon becomes more competitive in all patterns for 128 nodes; however, apart from bit complement, the 2D Mesh can still provide more throughput.

Realistic Traffic Figure 4.9 shows the average packet latency results for a number of PARSEC application traces of the considered NoCs, which are in line with our observations made for synthetic traffic: Amon improves performance compared to QuT significantly. When ACKs and REQs are parallelised in the destination-reservation phase, Amon reduces the average packet latency by $\sim$75% on average. This underlines our assumption regarding the application demands of realistic workloads on the on-chip network: although the average injection rate over the whole course of execution is low in these applications [LNP+13], we observe both structural and transient hotspots when running the simulations, which is the reason why Amon's control network provides significant improvements to QuT.

Although the 2D Mesh offers lower latency and can sustain higher network loads for synthetic traffic, Amon outperforms the 2D Mesh in terms of latency on realistic traffic. The main purpose of synthetic traffic is to stress the topology and expose weak spots, and all packets are unicast, which does not resemble the traffic of realistic workloads in which cache coherence protocols require nodes to send out multicast packets (e.g. for invalidation requests). For multicasts (and/or broadcasts), Amon is superior to electrical NoCs because it can send to each destination in the NoC independently (and thus simultaneously in case each destination is free) since each node has injection channels on which data can be modulated to each destination. In electrical NoCs like

the 2D Mesh, each node has one injection link into the local router, and must therefore serialise multicast packets one after another.

When analysing the packet latencies during our simulation, we observed that packets received at the vast majority of destinations exhibit a low average packet latency; however, for a few destinations, the average packet latency was significantly higher (5-10×). Analysing the number of retransmissions in QuT shows that these destinations send out an order of magnitude more NACK messages than the other destinations, showing that very high contention occurs at these destinations. In addition, the total number of packets received at these destinations is not significantly higher than for the destinations with lower latency, hinting that these contentions are likely caused by transient hotspots, i.e. phases in which a large number of nodes send to the same destination. This can be observed across all PARSEC applications (to a varying degree). For such hotspots, the underlying architecture of NoCs like QuT/Amon in which each destination can only receive data from one node at a time may not be ideal and will require further optimisations, as the one we propose later on in Section 4.3.

Increasing link bandwidth is the first obvious approach to improve performance; however, increasing bandwidth by increasing the number of wavelengths per link would lead to unacceptable overheads in terms of MR count (and thus MR heating) and laser power for networks of 64 nodes and more (the following section will discuss power consumption in more detail). Higher modulation rates (e.g. 20 Gb/s or 40 Gb/s) could improve the link data rate without additional MRs and wavelengths. Modulators of these speeds have already been demonstrated [LZY+11, LZT+12]; however, these devices currently impose large footprints and high energy consumption. The vast majority of previous studies has therefore not considered higher data rates than 10 Gb/s. However, devices mature at a fast pace and it would be interesting to study whether improvements in link throughput could make control network based WRONoCs more competitive. For that reason, all NoCs were also simulated for 16λ link bandwidth on the data network, and results are shown in Figure 4.10.

We make several interesting observations: first, the more efficient the control network mechanism, the higher the benefits of increasing link data rate from 8λ to 16λ. For instance, while Amon_par's average packet latency is more than halved, only ∼20% latency reductions are obtained in QuT (on average), suggesting that contention at the destination nodes are a decisive factor for determining latency. Second, when comparing QuT_16λ with Amon_par_8λ, we observe that Amon outperforms QuT significantly although only having half the link bandwidth. Based on these results, we

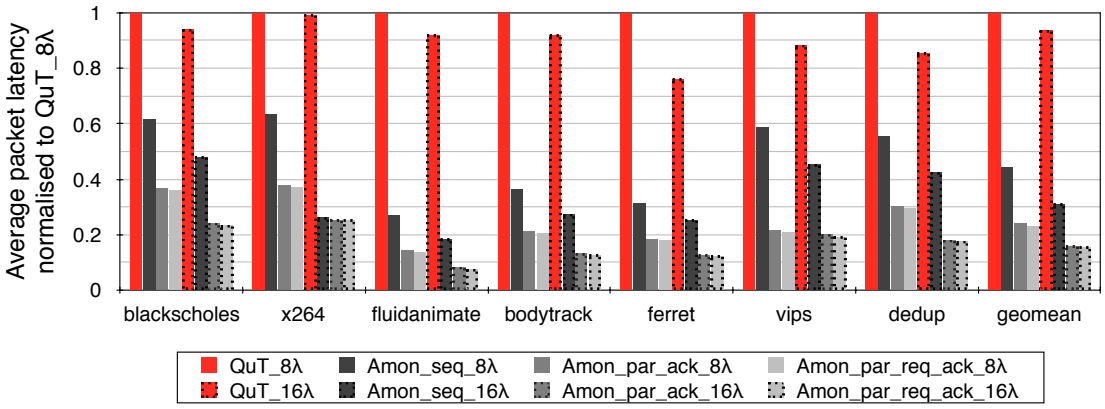


Figure 4.10: Average packet latency for PARSEC workloads for 8λ and 16λ link bandwidth

conclude that, while good latency improvements could be achieved with higher modulation rates, it will not be enough to cancel out the impact of an inefficient control network. An efficient control mechanism thus seems to be the basis for low latency in control network based WRONoCs.

Power Consumption

The total power consumption consists of laser, MR heating, dynamic and leakage power. Laser power dissipated at the off-chip laser source is static. To have a fair comparison, we use the same model as in QuT, which calculates laser power per wavelength based on the IL_{max} from the laser source to the receiver, laser efficiency, and the receiver sensitivity (see Equation 4.1). QuT's study does not mention a LPDN, the number of lasers coupled into the chip, or splitter loss explicitly when calculating IL_{max}. The reported laser power values, however, suggest that the authors assumed one dedicated laser source for each node. We thus assume that IL_{max} was computed starting from the injection points into the data network. Other studies have taken the same approach as it allows to study the impact of the switching topology on IL_{max} in isolation [OOTR+17]. In addition, we assume 20 μW/MR for MR heating. Dynamic power is dissipated in the EO and OE backends for data modulation/demodulation, which requires 100 fJ/bit and 50 fJ/bit, respectively [BJO+09]. For the 2D Mesh, dynamic power was extracted for a low-voltage 22 nm technology with DSENT [SCK+12].

$$P_{laser} = N_{wv} \times Le \times P_{sense} \times 10^{IL_{max}/10} \tag{4.1}$$

Table 4.1: Static Optical Power Requirements

		QuT	Amon	Control Network
64 Nodes	ILmax (dB)	16.36	15.31	17
	Num. of λ	128	128	64
	Laser Power per λ (mW)	3.44	2.7	4
	Laser Power (mW)	440.32	346.13	256
	Num. of MRs	46336	39014	4288
	MR Heating Power (mW)	927	780	85.7
	Total Power (W)	**1.367**	**1.126**	**0.341**
128 Nodes	ILmax (dB)	24.11	27.5	21
	Num. of λ	256	256	128
	Laser Power per λ (mW)	20.5	44.7	10
	Laser Power (W)	5.25	11.44	1.28
	Num. of MRs	182784	147552	17300
	MR Heating Power (W)	3.655	2.951	0.346
	Total Power (W)	**8.9**	**14.39**	**1.626**

Laser Power QuT does not provide a layout of their topology and their utilised modelling simulator PhoenixSim [CHB+10] not publicly available. It is thus not possible to reproduce their reported IL_{max} accurately since layout can have a tremendous effect on path lengths and waveguide crossings, and it is fair to assume that their simulator has done some sort of optimisation. To allow for a fair comparison, we utilise the technology parameters used in QuT to compute IL_{max}. Specifically, MR-drop loss (0.5 dB), MR-through loss (0.01 dB), waveguide crossing (0.12 dB), waveguide bending (0.005 dB), and waveguide propagation loss (0.1 dB/mm) is included in QuT's IL_{max} calculation [HJH14]. Coupling loss (1 dB) is not included in the reported IL_{max} and added separately together with laser efficiency (5 dB / ∼30%) to obtain the required laser power. Finally, the receiver sensitivity P_{sense} (-17 dBm / 20 μW) is multiplied as shown in Equation 4.1. IL_{max} for the control network is shown as reported in QuT. Table 4.1 lists our laser and MR heating power results.

The path with the maximum loss in Amon is between the two nodes in the opposite corners of the chip. For 64 nodes, IL_{max} is 15.31 dB. With 1 dB coupling loss, this translates to 2.7 mW per wavelength. Multiplied by the number of wavelengths, this leads to 346.13 mW, which is 21% less laser power than in QuT.

Amon induces lower IL_{max} as it has an improved switching topology. QuT is based on a ring topology that implements 'cross links' spanning across the topology originating

at each even node ID and 'bypass links' that cross the switch designs originating at each odd node ID (as depicted in Figure 3.4 in Section 3.2.2). While ring topologies are typically benign to layout in the place&route process due to their simplicity, QuT has numerous topological properties that make it inferior to Amon in terms of IL_{max}. One of them is the large number of cross and bypass links that cause significant numbers of waveguide crossings, both within the switch design (up to 3 per switch), and across switches by crossing each other. In addition, the injection MRs of each node are placed on one of the two ring waveguides (one per direction), resulting in a large amount of MR-through loss. Amon provides more spatial-division multiplexing by implementing one waveguide per row/column in each Submesh, which reduces MR-through loss, and has a topology tailored to a mesh-layout, which lowers waveguide crossings. On top of that, between each destination pair in Amon, a wavelength only has to be dropped from one waveguide to another at most once in the switching topology (not including injection/ejection which is required in both QuT and Amon). In QuT, the optical signal may be dropped multiple times, which adds additional 0.5 dB to IL_{max} for each drop.

Interestingly, 128-node Amon actually causes higher IL_{max}, which translates to twice the laser power. Theoretically, the excessive number of waveguide crossings and MR-through losses in QuT should make Amon superior for higher number of nodes. Amon's overhead can be explained by two reasons: scaling Amon to 128 nodes requires an 8×16 layout with 4×8 Submeshes. This layout is highly imbalanced, which leads to a network topology that has low IL_{max} in the vertical connections (only 8 rows), and high IL_{max} in the horizontal direction (16 columns). Our analysis has shown that this leads to very large overheads in MR-through losses and waveguide crossings. Amon should, therefore, be scaled in a balanced fashion (ideally equal number of rows and columns) to have even paths through the network and low IL_{max}. A second reason why QuT has lower loss than Amon could be the simulator that the authors used (Phoenixsim), which may perform layout optimisations. Layout tools were shown to be able to significantly reduce IL_{max} in topologies like QuT in which large numbers of waveguide crossings occur [BRBS16]. Unfortunately, it is not possible to estimate the extent of layout optimisation that Phoenixsim performs since it is not open-source. From an analytical point of view, Amon should scale more efficiently than QuT since QuT inherently has more waveguide crossings, MR-drop, and MR-through losses with worse scalability.

Analysing and comparing the data networks of different switching topologies is useful to identify the most efficient design; however, as discussed before, estimating laser power based on IL_{max} through the data network and assuming one laser source at each injection point is highly optimistic. More realistic designs assume a very low number of lasers coupled into the chip (ideally one), which requires to factor in a LPDN for distributing light across the chip into the power model. This has not been considered in previous control network based WRONoC proposals. In addition, no information about the injection backend design was provided in recent studies, and the evaluation results appear to have missed to include the heating requirements of the modulators. The modulator count, however, is a non-negligible portion of the total MR count (QuT/Amon needs $(N/4 \times \lambda)$-modulators at each injection channel in each node, i.e. $N \times (4 \times (N/4 \times \lambda) - \lambda))$ ($-\lambda$ since a node does not need modulators to address itself). We, therefore, include modulators into the power model, too.

Each node must be provided with (N/4) wavelengths in order to address every other node in the NoC. In fact, for a link bandwidth of 8λ, each node requires $(N/4) \times 8\lambda$ at each injection channel. Figure 4.5 illustrates the loss that optical signals experience on a LPDN for a 64-node WRONoC prior to actually entering the data network, assuming one laser source coupled into the chip. Each split introduces losses in the splitter (0.1-0.2 dB), and, in addition, the signal loses 50% (or 3 dB) of its strength when its split across two waveguides. As shown in Figure 4.5, for 64 nodes, light needs to be split six times on the path from the coupler to each node, and then another two times within each node to distribute the wavelengths to each injection channel. This leads to $8 \times (3\,dB + 0.1\,dB) + 0.8\,dB = 25.6\,dB$, for a splitter loss of 0.1 dB, 0.8 dB waveguide propagation loss (for 0.1 dB/mm and 8 mm path length), and 50% power splitters, which is more optical loss than in the data networks themselves (see Table 4.1). With 15.31 dB IL_{max} in Amon, 1 dB coupler loss, and 5 dB laser efficiency, this would lead to a total path loss of 46.925 dB, which translates to 0.981 W per wavelength laser power, and a total of 125.56 W for 128λ ($N/4 \times 8\lambda$). For QuT (16.36 dB IL_{max}), this would translate to 47.96 dB, 1.25 W per wavelength laser power, and a total of 160 W. These power overheads are impractically high and make a scaling above 64 nodes infeasible since the power requirements are unsuitable for on-chip interconnects and currently no laser technologies exist that could supply them. Therefore, the studies in the rest of this chapter will only consider NoCs with 64 nodes.

Fortunately, the loss parameters assumed in QuT do not represent the latest device technologies – more advanced devices have already been fabricated and verified since

Table 4.2: Conservative and Aggressive SiP Technology Parameters

Loss Parameter	Conservative	Aggressive
Waveguide propagation	0.1 dB/mm	0.0271 dB/mm
Waveguide crossing	0.12 dB	0.04 dB
Waveguide bending	0.005 dB/ 90	0.027 dB/ 90°
MR-through	0.01 dB	0.001 dB*
MR-drop	0.5 dB	0.5 dB
Splitter	0.1 dB	0.1 dB
Coupler	1 dB	1 dB
Laser efficiency	5 dB	5 dB
Receiver Sensitivity	-17 dBm / 20 μW	-21 dBm / 7.94 μW

QuT's publication in 2014 – and devices are widely expected to mature over the next few years. Table 4.2 lists the technology parameters utilised in QuT ('conservative') and more aggressive technology parameters of more advanced SiP devices. Silicon hybrid rib/strip waveguides of 460 nm width have been demonstrated with propagation losses of 0.0271 dB/mm and bend losses of 0.027 dB/90° [BS11], and so have CMOS compatible waveguide crossing arrays that decrease crossing loss down to 0.04 dB [LSZP14]. Photodetectors with better sensitivity than -17 dBm have also been demonstrated, and receivers can exhibit sensitivities of -21 dBm at 10 Gb/s [BRNB16, MNM+12]. In addition, numerous studies expect MR-through loss of 0.01 dB to decrease down to 0.001 dB or even 0.0001 dB [ZAU+15, BSK+10, JBK+09, LBGP14]. Table 4.2 lists these improvements under 'aggressive device parameters'. The asterisk (*) indicates device projections/speculations rather than demonstrated devices (only MR-through loss).

In order to assess the impact that these improvements have on designs like Amon, it is necessary to analyse what contributes the most to IL_{max} in these networks. For Amon, waveguide crossings and MR-through losses are the major contributors for the utilised conservative parameters in the previous section. Figure 4.11 presents a breakdown of the losses contributing to IL_{max} for the conservative (cons) and aggressive (aggr) technology parameters of Table 4.2 for Amon's data network and LPDN (note that, for the sake of clarity, waveguide bending losses are omitted in these charts due to their negligible contribution to IL_{max}).

Our first observation is that Amon's data network benefits significantly from more advanced device technologies – all improvements translate to noticeable reductions of IL_{max} in Amon's topology. In particular, we observe that the technological improve-

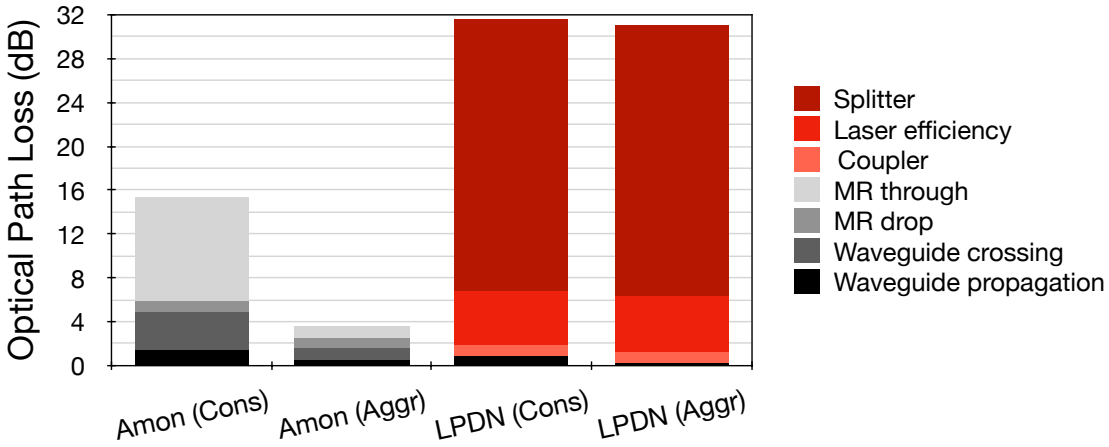


Figure 4.11: IL_{max} breakdown of Amon and the LPDN for 64 Nodes

ments have a large impact on MR-through and waveguide crossing loss. The overall IL_{max} is reduced by $\sim 4\times$. The second observation is regarding the LPDN: the most losses to the optical signal actually occur in the LPDN. Losses caused due to low laser efficiencies of current technologies and loss for coupling light from a fibre into the chip are unavoidable and would occur even if light did not have to be distributed; however, the vast majority of the optical loss stems from splitting, i.e. distributing, light across the chip to each node. Splitting loss includes loss within the splitter (0.1 dB) and 3 dB for a splitting ratio of 50% down each path of a 1:2 split, where the latter is the decisive loss factor. Advanced technologies have no impact on the fact that optical power needs to be split when being distributed. Technological advances only provide slight improvements regarding the waveguide propagation loss in the LPDN. Note that QuT's laser power requirements would decrease as well, however, not to an extent to which it would be superior to Amon as its switching topology is inherently less efficient (as discussed earlier).

Table 4.3 lists the total IL_{max} of the entire optical NoC (Amon + LPDN) from the point light enters the chip to all receivers for both conservative and aggressive technology parameters for 64 nodes. While Amon requires impractical amounts of laser power for conservative parameters, advanced technology parameters have a large impact and can actually reduce the total laser power to an acceptable level. Improvements in laser efficiency, advanced coupling devices, and higher receiver sensitivities could lower this value even further. In summary, with advanced technologies, Amon could be a suitable alternative for CMPs with high network utilisation demands. In particular, apart from the devices already demonstrated, MR-through loss projections are crucial in the future to enable power-efficient WRONoCs.

Table 4.3: Total power results for Amon with LPDN for 64 Nodes (dynamic power is listed for uniform random traffic prior to network saturation)

	Amon + LPDN (cons)	Amon + LPDN (aggr)
ILmax (dB)	46.925	34.580
Laser power per λ (W)	0.985	0.0574
Total Laser Power (W)	126.08	2.918
Num. of MRs (Modulators + Filters)	71270	71270
MR heating power (W)	1.43	1.43
Total Static Optical Power (W)	127.51	4.35
Dynamic Power (W)	0.587	0.587
Leakage Power (W)	0.091	0.091
Total Power (W)	**128.06**	**5.02**

The exponential relationship between laser power and IL_{max} shows how important it is to reduce path losses through smart network architectures and advanced technologies; however, ultimately, losses to distribute light across the chip will remain to dominate the IL_{max} due to the physical properties of splitting light, which will effectively limit the power efficiency of WRONoCs for larger scale NoCs.

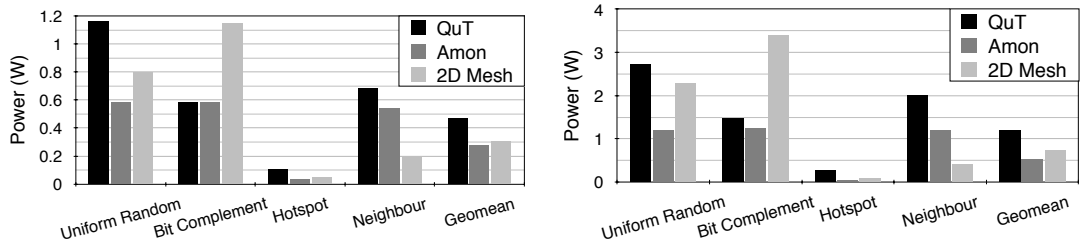
From an architectural point-of-view, core clustering can reduce the number of injection points in the NoC, leading to less light splitting and thus large laser power reductions; however, each node would only have half the bandwidth, and contention at the destination nodes, which already poses a performance challenge for realistic traffic (as revealed in Section 4.2.4), would increase even further in this case, possibly leading to unacceptable latency overheads. However, when considering the injection points of Amon, and in fact most control network based WRONoCs in literature, we observe that each node has injection channels which allows to send a packet to every node in the network simultaneously at any given time. The question arises whether this is actually necessary to provide sufficient network performance, or whether fewer injection channels can satisfy the performance demands, too. For instance, reducing the number of injection channels to one would allow to discard two splits on the optical path without requiring core clustering. This would reduce IL_{max} by 6.2 dB, which would reduce the total laser power to 30.25 W and 0.881 W in the conservative and aggressive case, respectively. Section 4.4 will present an approach to achieve these reductions and evaluate its efficiency.

MR Heating Power Amon's topology allows for fewer MRs than QuT, particularly as network size increases, and in turn saves MR heating power, showing that improved topologies can save both power and resources. Table 4.1 lists the MR requirements of Amon and QuT. Amon's topology requires fewer MRs than QuT with better scalability: Amon requires 16% and 19% fewer MRs than QuT for 64 and 128 nodes, respectively, thus decreasing the total MR heating power consumption by the same percentage.

Dynamic Power Figures 4.12a and 4.12b report the dynamic power consumption for packet buffering, modulation, detection, and in the EO/OE backend circuitry for the different considered synthetic traffic patterns (prior to network saturation of QuT) for 64 and 128 nodes, respectively. Note that we only consider Amon with the control mechanism that parallelises both REQs and ACKs as the differences between Amon's control mechanisms in terms of dynamic power were negligible.

Our control network mechanism reduces dynamic power compared to QuT because neither NACKS nor REQ retransmissions are required. As dynamic power is measured slightly before network saturation of QuT, contention at the destination nodes is high, leading to more NACKs and REQ retransmissions in the control network. In addition, each control packet is received by 16 nodes in the control network as the optical signal is distributed using splitters. Therefore, each control packet consumes dynamic power for detection and in the OE backend circuitry in 16 nodes, which explains the large extent of power savings by discarding NACKs. The amount of QuT's power overheads thus stem from the number of additional NACKs and REQ retransmissions, which depend on the contention at the destination nodes, which in turn depends on the traffic pattern and injection rates. For instance, in bit complement traffic, each node sends to a different unique destination, leading to no contention and in turn no dynamic power overheads. The biggest difference is observed in the pattern with the highest contention – hotspot traffic – which, however, also exhibits the lowest absolute dynamic power since these NoCs saturate very early. On average, the absence of NACKs and REQ retransmissions saves $\sim$45% and $\sim$55% for the 64 and 128 node case, respectively.

Dynamic power savings compared to the 2D Mesh vary significantly based on the traffic pattern due to its sensitivity to distances on chip (transmitting to destinations far away from the source causes high numbers of hops and, in turn, more energy consumption for each router and link traversal). In optics, data communication itself is very low power once the power for the laser and MR heating is paid for. This explains



(a) 64 Nodes (b) 128 Nodes

Figure 4.12: Dynamic Power Consumption for Synthetic Traffic

the dynamic power savings of Amon/QuT compared to the 2D Mesh, which are on average 8% and 23% for 64 and 128 nodes, respectively.

Leakage Power We assume buffer space for 24 $\times$ 64 bits at the output buffers at each node which corresponds to the buffers space in the 2D Mesh at each input port of a router. In addition, we apportion buffer space for one cache line (576 bit) at each ejection channel since each destination can only receive data from one sender at a time, which can be at most the size of a cache line. Our evaluation has shown that leakage power in WRONoCs plays an insignificant role, attributed to the absence of buffering at intermediate routers. In fact, many WRONoC studies do not even explicitly include leakage power in their power model [KH12] [HJH14]. In addition, Amon does not incur any additional overheads compared to QuT since both NoCs have the same buffer requirements; however, for a fair comparison to electrical NoCs, at least the input and output buffer power values should be included. In total, leakage power in WRONoCs contains leakage power in the input and output buffers, the EO and OE backends, and driver circuits in our model. Compared to the buffer requirements of electrical NoCs in the routers, however, these are low (see Table 4.4).

Total Power Table 4.4 compares the power consumption of Amon and QuT to the 2D Mesh. Power values include the entire NoCs, i.e. the data network, the control network, and the LPDN. Dynamic power represents the dynamic power dissipated prior to network saturation of QuT for uniform random traffic. Power for the control network in the aggressive case was computed based on the IL_{max} reported in QuT with the conservative parameters, but with an aggressive receiver sensitivity of -21 dBm. The power of the control network is therefore slightly overestimated in the aggressive case; however, note that the power consumed for that receiver sensitivity is 100 mW and is therefore only a very small contribution to the overall laser power ($< 1\%$) –

Table 4.4: Total Power Consumption of 64-node NoCs

Power	QuT (Cons)	Amon (Cons)	Amon (Aggr)	2D Mesh
Laser	160.256	126.336	3.018	-
MR heating	1.66	1.51	1.51	-
Leakage	0.091	0.091	0.091	0.41
Dynamic	1.158	0.587	0.587	0.8
Tota Power (W)	**163.16**	**128.5**	**5.2**	**1.21**

most power is consumed on the data network. For the conservative parameters of QuT, Amon reduces power compared to QuT by 21%. However, both WRONoCs are not competitive compared to an aggressive 2D Mesh at 22 nm. Although most of the power is consumed at the laser source which is off-chip, the overall system efficiency would be seriously decreased. For aggressive parameters, results look more promising, although 5.2 W is still well above the 2D Mesh (1.21 W).
Dynamic power, however, increases significantly faster in the electrical NoCs, and the dynamic power in Table 4.4 was extracted at QuT's saturation point, which is fairly low. Amon, however, can sustain higher network loads and optical NoCs become increasingly efficient as injection rates increase. The highest dynamic power savings of Amon compared to the 2D Mesh are therefore at the saturation point of Amon, which is at 65 Gbps/node. At this point, Amon consumes just 0.88 W whereas the Mesh consumes 1.2 W dynamic power.

Resource Requirements and Area

QuT was shown to be the design with the lowest area requirements compared to a number of alternative WRONoCs, thus a comparison of Amon with QuT allows to assess how Amon can compete amongst these proposals. Area requirements of WRONoCs, however, are typically difficult to estimate in a meaningful way for various reasons. Based on whether integration is envisioned to be monolithic or on a separate die and 3D integrated, placement constraints vary. Since integration is widely envisioned to be performed on a separate die, placement of the components depends on the locations of the TSVs, and placement of the on-chip components (cores, caches, etc.), and thus varies between designs. Also, the layout of the SiP components requires consideration of spacing between them to avoid crosstalk. On top of that, place&route tools can

Table 4.5: SiP Resource Requirements

		QuT	Amon	Control Network
64 Nodes	Num. of MRs	46336	39014	4288
	Num. of Waveguides within NoC	130	64	4
	Num. of Injection Waveguides	256	256	-
	Num. of Ejection Waveguides	128	240	64
	Total Num. of Waveguides	514	560	80
128 Nodes	Num. of MRs	182784	147552	17300
	Num. of Waveguides within NoC	258	96	8
	Num. of Injection Waveguides	512	512	-
	Num. of Ejection Waveguides	256	480	128
	Total Num. of Waveguides	1026	1088	144

change the total required area through optimisation. Nevertheless, the MR and waveguide requirements of a topology are typically a good first indicator to estimate the area overheads of a WRONoC topology.

The resources requirements of Amon and QuT are listed in Table 4.5 for 64 and 128 nodes, and are categorised into waveguides necessary to perform the switching within the data network, and waveguides required to inject and eject optical signals into the switching topology. Both Amon and QuT require four injection waveguides to inject data into the network. Amon needs more ejection waveguides because optical signals must be ejected from each cardinal direction, while QuT is a bidirectional ring and only needs to eject signals from two directions. Since optical signal injection/ejection occurs within the switch design of the local node, these waveguides are much shorter than the waveguides connecting the switches. The impact on the overall area is therefore most likely low. Therefore, although Amon needs more waveguides than QuT overall, the waveguides required in the switching topology are actually much fewer (less than half), most likely leading to less waveguide area requirements overall. In addition, the number of MRs, as reported in the previous section, is lower in Amon than in QuT, which further reduces area.

Note that the control network consists of MWSR buses on which data is directly modulated onto the bus and is not injected into the network like in WRONoCs. Therefore, no injection waveguides are necessary. Ejection waveguides on control network are normally not necessary on MWSR buses neither since photodetectors are placed right behind the ejection MRs; however, QuT's control network utilises splitters, which requires one additional waveguide per split to guide the optical signals to the local node, which results in N ejection waveguides.

In both designs, the number of waveguides is ∼500 and ∼1000 for 64 and 128 nodes, respectively. To put this into perspective to VLSI layout: waveguides of 470 nm width have been successfully fabricated. In order to avoid high crosstalk between two adjacent waveguides, they should be spaced 0.5-3 μm from each other [PD10]. Therefore, often a total (pessimistic) waveguide pitch of 4-5 μm is assumed in the scientific literature (e.g. in [LBGP14] or in the technology library of DSENT [SCK+12]), which enables a compact implementation. In addition, MRs can be as small as 3 μm [NFA11]. From a layout perspective, such dimensions are thus within the feasible design range.

4.2.5 Discussion

WRONoCs based on control networks are likely the only viable solution to implement all-optical NoCs without excessive static power consumption. Although QuT outperforms numerous state-of-the-art proposals in terms of power consumption, advanced designs such as Amon can improve power efficiency even further. In fact, Amon reduces power consumption by 21% for 64 nodes compared to the topology QuT.
In terms of performance, our results confirm our hypothesis that contention in the control network is the decisive performance delimiter for NoCs like Amon/QuT, and even doubling the link bandwidth on the data network cannot hide inefficiencies on the control network. Parallelising ACKs and REQs to data transmission and removing NACKs from the control mechanism reduces contention on the control network, hides arbitration latency, and decreases dynamic power significantly. This alleviates most of QuT's shortcomings and improves throughput on synthetic traffic by up to 45% and reduces packet latency on PARSEC traces by 75% (on average). In addition, Amon outperforms aggressive electrical baseline 2D Mesh by similar margins on realistic traffic (70% on average).
Despite the laser power reductions of Amon compared to QuT, the total laser power is still too high even for the most advanced SiP devices. This significantly decreases its power efficiency compared to electrical baselines and must be further addressed. Besides, neither QuT nor Amon includes non-linear effects in their power model. Non-linear loss of ∼0.35 dB for ∼100 mW optical power in a 10 mm waveguide has been reported [LBGP14], which is fairly low. However, the 64-node Amon for aggressive SiP devices requires ∼3 W at the laser source, and although this power is halved at each split in the LPDN, this could lead to considerable non-linear losses. One possible solution to this problem is to simply utilise more laser sources and couple them into the lower branches of the LPDN, which would reduce the power per laser source

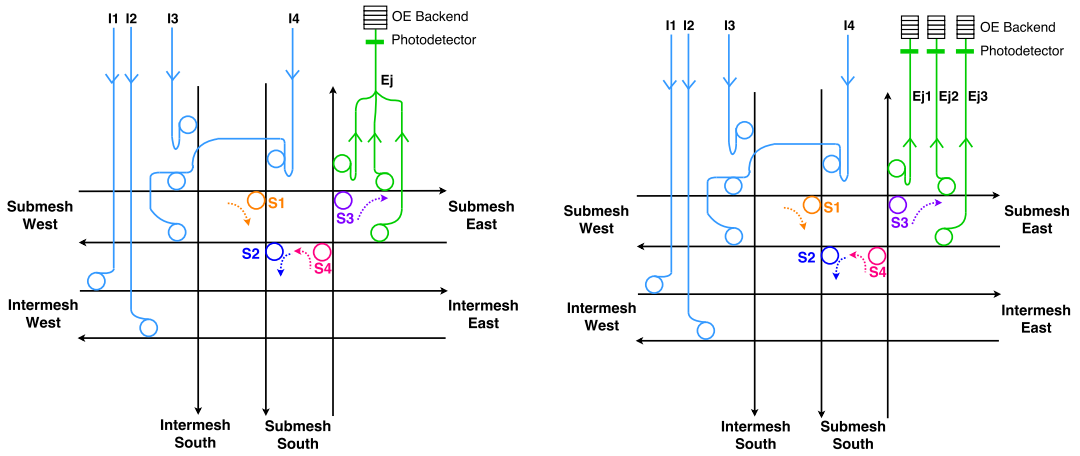
and in the waveguides and in turn non-linear losses. The trade-off is that more laser sources are required in the system. Reducing laser power is therefore not only important to design efficiency, but also to keep non-linear losses at levels at which they only have little impact on the total laser power consumption. Later on in this chapter (see Section 4.4), we will discuss an architectural improvement to the injection back-end which substantially reduces the number of laser splits required to drive the signals into the injectors.

We will move on, first, into tackling the performance degradations arising from communication hotspots, which are common in emerging workloads and, as we just discussed, one of the weakest points of Amon/QuT. In order to avoid latency overheads and make these NoCs more robust against these kinds of traffic patterns we present an architectural modification to the ejection backend that reduces the contention at the receiving nodes by increasing the number of ejection channels.

4.3 Deploying Multiple Ejection Channels

The availability of a single ejection channel at the nodes can generate unnecessary contention scenarios, which can be especially pathological under transient hotspots as those arising from realistic workloads. The incidence of such hotspots depends on the CMP system architecture and applications but is typically low (one or two hotspots in the PARSEC traces used in our study), meaning that only a very small fraction of the nodes in the NoC must be provided with higher ingress bandwidth or different access patterns. Unfortunately, light (i.e. the wavelengths for modulation) is distributed to the injection channels of the nodes in NoCs like Amon/QuT, which makes a destination-based bandwidth scaling impossible to manage in real-time.

Improved control mechanisms, however, can have a large impact. When analysing Amon's switch design, one can observe that the design of the ejection channels can be modified to accept data from more than one sender at the same time, which could take off some of the contention on the control network and improve performance. While one benefit of control network based WRONoCs compared to contention-free WRONoCs is the reduced number of receivers at a destination (1 vs. $N-1$), increasing the number of receivers at a node from 1 to *1 for each cardinal direction* will likely not eliminate these benefits. In fact, switches in Amon have incoming links from either 3 or 4 directions, which is still much less than $N-1$ in contention-free WRONoCs.



(a) Original Amon Switch 33 with one ejection channel

(b) Amon Switch 33 with OE backends at each ejection channel

Figure 4.13: Switch Design with OE Backends at Each Ejection Channel

4.3.1 Deploying Multiple Ejection Channels in Amon

A possible improvement that does not require additional MRs or bandwidth on the links is to allow nodes to receive data from more than one sender at a time. Figure 4.13a shows the design of Switch 33 in the original 64-node Amon design, and Figure 4.13b our proposed modification. Each switch in Amon receives optical signals from each cardinal direction and ejects these signals with its MR filters, which will then be combined to the same waveguide and absorbed by the photodetector. Instead of combining these waveguides, each ejection waveguide – which is already provisioned for – could guide its optical signals to its own, separate photodetector. This would enable a node to receive from three senders at the same time (rather than one), just by adding two additional photodetectors and backend circuitry. These overheads are very small, and will not even lead to higher MR heating power since no MRs are added.

The destination-reservation mechanism must be adjusted because the access rights to such destinations change: instead of allowing only one node to send to a destination at a time, a destination can now grant access to *one sender per cardinal direction* simultaneously provided their optical signals enter from different cardinal directions so that they are ejected to different photodetectors. Since switching in WRONoC topologies is static/deterministic, each destination knows the cardinal direction from which optical signals will enter the switch based on the sender ID. In Amon, for instance, this

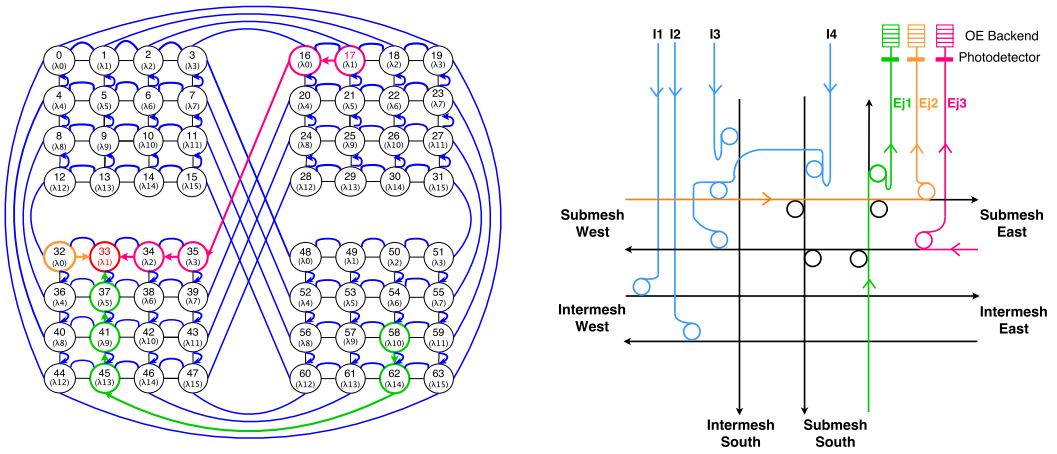


Figure 4.14: Example of Simultaneous Data Reception from Three Different Senders

depends on the relative location of the sender to the destination. Figure 4.14 exemplifies this for Switch 34. Node 17, 32, and 58, can all send to Node 34 at the same time because their optical signals are entering Switch 34 from different cardinal directions and are thus absorbed by different photodetectors.

Structural or transient hotspots may be known at design time. For instance, in embedded systems or systems-on-chip, the application domain and communication patterns are known a priori and will not change over the lifetime of a system [JP09]. In these cases, it is possible to just equip the hotspot nodes with multiple ejection channels which would allow for very low overhead modifications. However, in many other domains, such as server computing or data centres, applications and workloads change and the operating system may execute the same program on different resources. In these cases, it may be desired to have one ejection channel for each cardinal direction at *each node* in the NoC, provided that the total overheads are acceptable. Therefore, we evaluate the overheads and drawbacks of both designs in the following. We expect the overheads to be small in either case because, even if multiple ejection channels exist, each ejection channel can only receive data from one sender at a time, therefore requiring only very limited buffering (e.g. for one cache line).

4.3.2 Evaluation

PARSEC workloads were simulated in order to identify the hotspot nodes in each application, which were then equipped with the backend modifications described above. In all applications, it was mainly one destination that constituted a significant hotspot. It should be noted, however, that the hotspots are also determined by the cache hierarchy

and coherence protocols, and not necessarily by the applications only, so for different microarchitectures the hotspot nodes may differ. One hotspot destination leads to a design in which the modified switch design is added to only *one node* in Amon, denoted as *Amon_multiple_ej_1*. Similarly, we evaluate Amon with multiple ejection channels at *each destination*, denoted as *Amon_multiple_ej_each*. We utilise the most efficient control mechanism of the previous section, which is the design in which both REQs and ACKs are parallelised to data transmission. We compare our design to the standard Amon with one ejection channel per node and the same control mechanism. For our evaluation with synthetic traffic, we only compare Amon to Amon_multiple_ej_each since it is not possible to identify hotspots in these traffic patterns (or hotspots may not even exist). Both studies only consider NoCs with 64 nodes given the lack of scalability in terms of laser power exposed in the previous section. We include QuT and the 2D Mesh to the evaluation to show the improvements both in relation to the standard Amon and the alternative NoCs.

Performance Analysis

Figure 4.15 depicts the latency results for synthetic workloads. Latency and throughput improvements are very promising, with throughput doubled for most traffic patterns. Latency and throughput remain unchanged in bit complement traffic in which each destination receives traffic from only one sender, therefore more ejection channels have no impact on performance. Throughput is improved to a point at which Amon can now compete with a 2D Mesh. In summary, using multiple ejection channels improves throughput for all traffic patterns, and is competitive to the 2D Mesh (apart from neighbour traffic).

Figure 4.16 shows the latency results for PARSEC workloads. Adding three photodetectors and OE backend circuitry to each ejection channel in one node has tremendous performance benefits across all applications. On average, Amon's packet latency is halved just by adding our modification to one node. In fact, adding multiple ejection channels to each destination does not lead to significant further latency reductions in these workloads ($\sim$4%). Another interesting observation is that the previous section showed that doubling the link bandwidth on the data network to 16λ also halves the latency on average, which means that adding more ejection channels has the same effect as doubling link bandwidth. The difference, however, is that doubling link bandwidth would lead to unsustainable laser power overheads, whereas adding ejection channels causes only very little overheads (as we will see in the following). Compared to QuT

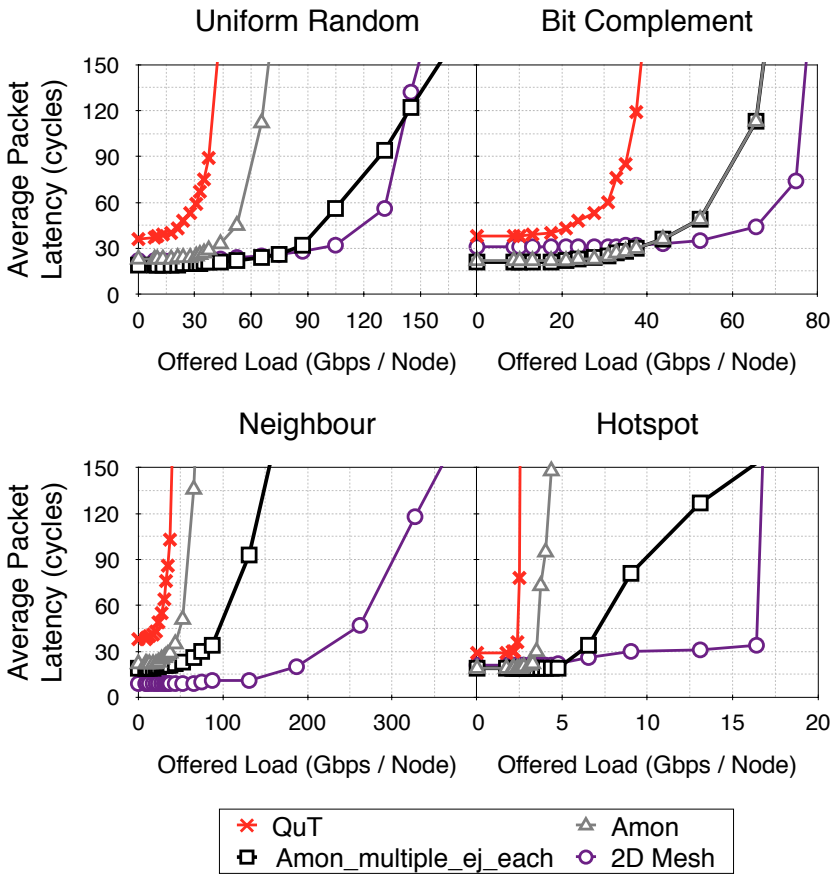


Figure 4.15: Average packet latency for synthetic workloads: Multiple Ejection Channels

and 2D Mesh, the latency improvements are now even higher than before, making our design the far superior choice in terms of performance. Note that QuT could technically also be extended to function with multiple ejection channels, but its topology would only allow this approach to a limited extent since it consists of a bidirectional ring and thus only receives data from two directions (rather than 3 or 4 in Amon). Nevertheless, performance gains of this approach could likely be attained in QuT, too, just to a lower extent.

Power Overheads

Figure 4.17 shows the power breakdown of the standard Amon, Amon with only one destination with multiple ejection channels, and Amon with all destinations having multiple ejection channels. The power overheads include leakage power of the additional input buffers, photodetectors and drivers, and OE backend circuitry for each

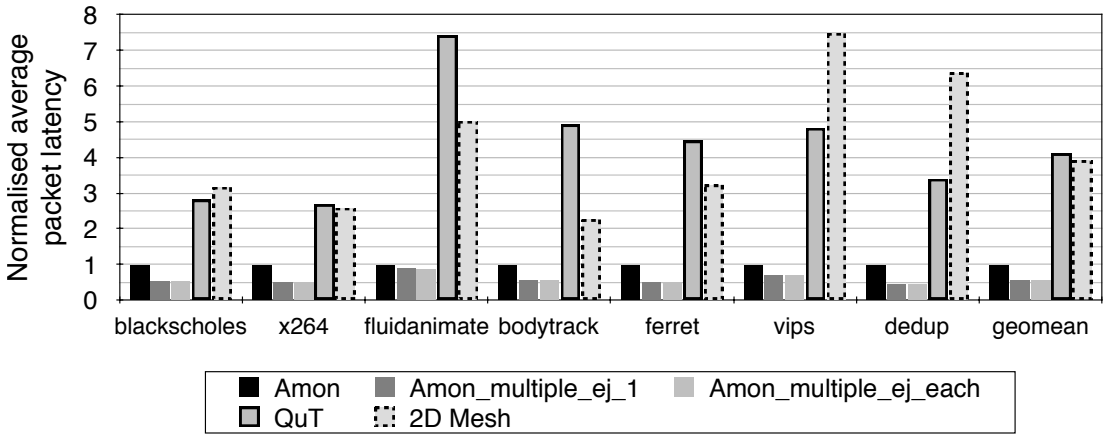


Figure 4.16: Average Packet Latency for PARSEC Workloads: Multiple Ejection Channels

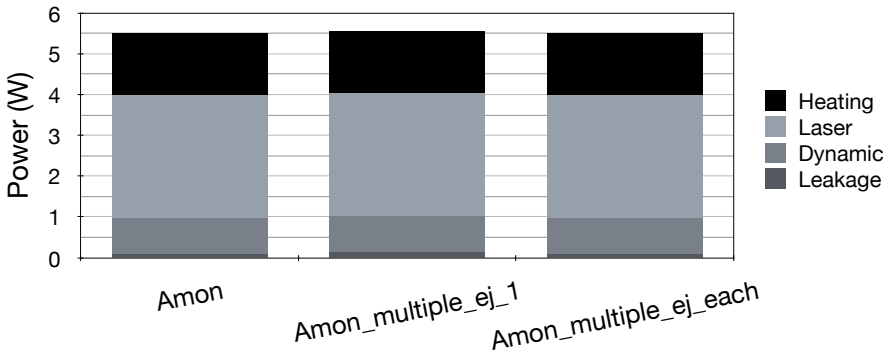


Figure 4.17: Power Overheads of Implementing Multiple Ejection Channels

ejection channel. We apportion each receiver with 576 bits input buffer space, which corresponds to one cache line. Leakage power in optical NoCs is widely known to have a very small impact on the overall power, which is also reflected in our power results. Therefore, the total leakage power overheads, which are $\sim$40% when each destination has multiple ejection channels, translate to negligible power overheads overall ($<$ 0.5%). The latency and throughput improvements of our approach, therefore, come with very little overheads.

4.3.3 Discussion

Implementing multiple ejection channels turns hotspot sensitive control network based WRONoCs into high-speed NoCs that outperform state-of-the-art all-optical and aggressive electrical baselines for realistic traffic patterns. In addition, contention at the destination nodes is alleviated, which makes Amon robust towards various different

traffic patterns at fairly uniform latency and throughput levels, which is a desirable property in many-core architectures [JBK+09]. Power overheads of additional ejection channels are negligible, even if implemented at each node, making this a well-rounded architectural optimisation which could be incorporated to future or other existing [KAH11][KH12][HJH14] all-optical NoC designs.

4.4 Leveraging MR Tuning to Reduce Splitting Losses

As shown in the experiments and discussed before, the power requirements – particularly at the laser source – are still too high to make WRONoCs a viable candidate to replace electrical NoCs. As identified in Section 4.2.4, the LPDN is a main contributor to the total laser power since light distribution degrades the signal by 3 dB for each 50:50 split (+ 0.1 dB splitter loss). Any reduction in the amount of splitting would, therefore, allow for large laser power savings. The first obvious approach to reduce splitting is to decrease the number of switches by deploying clustering, i.e. two nodes are connected to one switch through which they enter the network; however, this would also halve the bandwidth per core. Ideally, we would reduce splitting without causing significant bandwidth losses.

When investigating QuT's and Amon's switch design, the question arises whether the number of injection channels, in particular, the number of destinations that data can be transmitted to simultaneously, is actually necessary to support the NoC's traffic demands. With the current architecture and assuming that each destination is free, a node can send to every other node in the NoC simultaneously (i.e. broadcast) as they are addressed by different λ-sets on different injection channels. We argue, that this is rarely the case as most cache coherence protocols in large-scale CMPs deploy directory-based protocols, which exhibit (typically) low fan-out multicast traffic, but not broadcast [JP09]. Indeed, most operations for data communication are of unicast nature. For this reason, decreasing the number of injection channels may not have a diminishing impact on performance in realistic systems but could save two splits (1 *vs.* 4 injection channels) and thus huge amounts of laser power. This section will present the injection backend proposal with a single injection channel which allows for substantial power savings with little performance overheads that are more than compensated by the other proposed optimisations discussed above.

4.4.1 MR Tuning to Reduce the Number of Injection Channels

If the four injection channels are reduced to one injection channel, it must still be ensured that each node can inject data on all the waveguides to reach each destination in the data network. Figure 4.18a shows a backend design in which *one* injection channel is connected to all required waveguides (for example Switch 33 of a 64-node Amon), providing all necessary injection points into the network. Injection filters are basically chained one after another on the injection waveguide.
Since one injection channel is now connected to each injection point, some sort of arbitration must occur prior to data transmission since four MR filter banks respond to the same λ-set (four destinations share the same λ-set for addressing). For instance, in Figure 4.18a, if data shall be injected into the Intermesh link to the west, it would be filtered by a MR filter responding to the same λ-set to the Submesh south before that optical signal could ever reach the Intermesh west injection point. Luckily, MRs can be dynamically tune/detuned by integrated heaters to respond to a certain wavelength channel or not (as discussed in Section 2.3.2). Therefore, we propose to perform MR detuning of all injection filters responding to the λ-set that shall not inject into the network prior to data transmission. This is shown in Figure 4.18b: data shall be injected into the Intermesh west waveguide. Therefore, all injection MRs responding to the injection λ-set are detuned prior to data transmission (grey) and only the injection MR on the Intermesh west waveguide is tuned in, ensuring correct data injection into the network. Unfortunately, this approach does not only have a decreasing effect on power consumption, but also a negative effect on latency and thus requires a detailed analysis of its benefits and drawbacks.

Impact on Performance

Two factors may degrade performance: i) delay for MR tuning prior to data transmission, and ii) bandwidth reductions due to fewer injection channels.
i) Previous studies assumed tuning delays of one core clock cycle [Van10], and devices of at most 500 ps have been demonstrated [PTDS16]. Besides, MR tuning could be performed in parallel to the delay through the EO backends during data transmission. We, therefore, assume one core clock cycle of additional delay caused by MR tuning in our model.
ii) The previous design with four injection channels allowed to send data to each of the destinations in the NoC simultaneously since there was one injection channel per

Submesh, and one λ-set to address each destination within a Submesh (N/4 λ-sets in total per injection link). In our design with one injection channel, a sender can only transmit to N/4 different destinations simultaneously. For instance, if a node wants to send data to two destinations that are addressed with the same λ-set, it would have to transmit data one after another since only one injection channel is available.

Impact on Power

From a power perspective, as described earlier in this chapter, reducing the number of injection channels would reduce the splitting requirements by 6.2 dB. Connecting one injection waveguide to all injection points, however, will lead to higher path losses in the NoC due to more MR-through losses (the worst-case λ now has to pass $(3 \times (N/4) \times \lambda)$ additional MRs. For 8λ link bandwidth and 64 nodes, this would result in additional 384 MR passings, which would result in additional 3.84 dB (for conservative 0.01 dB) and 0.384 dB (for aggressive 0.001 dB) MR-through loss. In either case, the savings in splitter loss outweigh these overheads, and, in the aggressive case, the MR-through loss overheads are insignificant compared to the splitter loss savings. Apart from savings in IL_{max}, MR heating power is also saved: as we only require one injection channel now, the number of required modulators is sliced by four (see Figure 4.19). Since these reductions take place in each node, the savings in MR heating could be significant. In addition, design complexity is reduced as much less EO backend circuitry is required. The question that remains is: can these power savings actually provide a more power-efficient solution, or will the performance degradation outweigh the power savings? We aim to provide an answer to this question in the following.

4.4.2 Evaluation

Methodology

This approach is applied both to the standard Amon (one ejection channel per node) and to the Amon with multiple ejection channels at each node proposed in the previous section. '4Inj' indicates four injection channels per node (standard Amon), and '1Inj' the proposed one-injection-channel approach. All Amon NoCs utilise the most efficient destination-reservation mechanism in which both REQs and ACKs are parallelised. We evaluate these NoCs both under synthetic and realistic workloads and provide a power comparison.

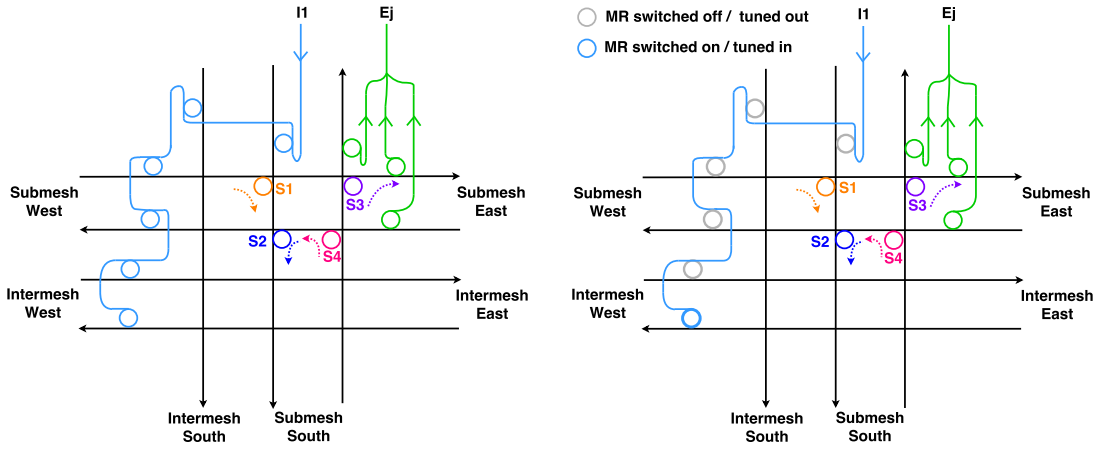


Figure 4.18: Backend modification example Switch 33: one injection channel is connected to all waveguides. MR filters are tuned/detuned prior to data transmission to allow the optical signal to enter the network into the correct waveguide

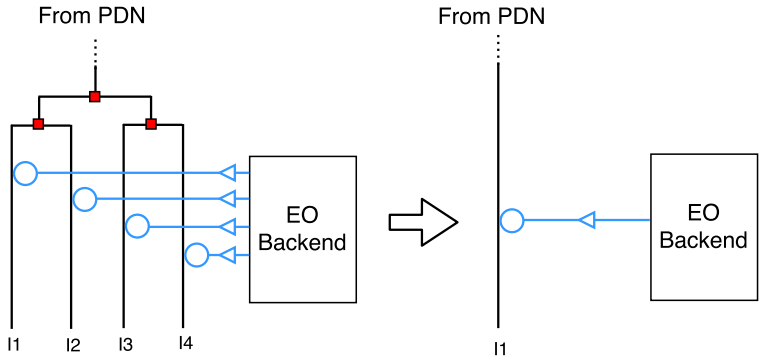


Figure 4.19: Amon Backend: One vs. Four Injection Channels

Performance Analysis

Figure 4.20 shows the latency results for synthetic traffic. Reducing the number of injection channels from four to one only increases packet latency by a very small percentage, and the network saturates slightly earlier when only one injection channel is used. In bit complement traffic, each source has only one destination, so a source will never have to send packets to two different destinations that share the same λ-set. The same applies to neighbour traffic since two destinations with the same λ-set would be located in different Submeshes far away from each other on the chip. In these two cases, the performance differences are caused by the latency for MR tuning only. In hotspot traffic, contention at the injection channel can occur; however, performance is mainly limited by the contention at the hotspot destination node, leading to small

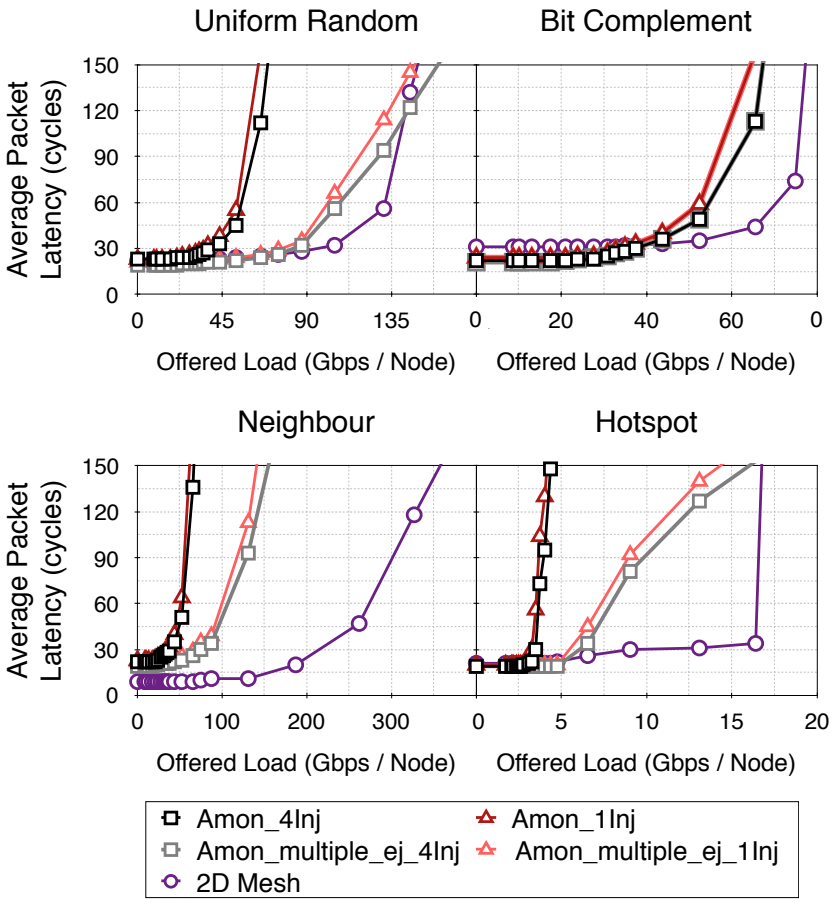


Figure 4.20: Average Packet Latency for Synthetic Traffic: One vs. Four Injection Channels

latency and throughput differences. In uniform random traffic, latency is very similar for low-to-moderate injection rates, but Amon saturates 13% earlier with one injection channel. One possible explanation for this is that the likelihood of two subsequent packets in a source node to be sent to a destination with the same λ-set is actually quite low since only four (out of 64) nodes share the same λ-set in Amon.

Figure 4.21 shows the latency results on PARSEC workloads. Average packet latency is increased on all of the workloads, and on average by 20%. For realistic workloads, these modifications may thus lead to some latency drawbacks; however, in none of the workloads to a diminishing extent. Compared to the 2D Mesh and QuT, the latency savings are still large. The question is whether these latency overheads are justified by the power savings, which is analysed in the following.

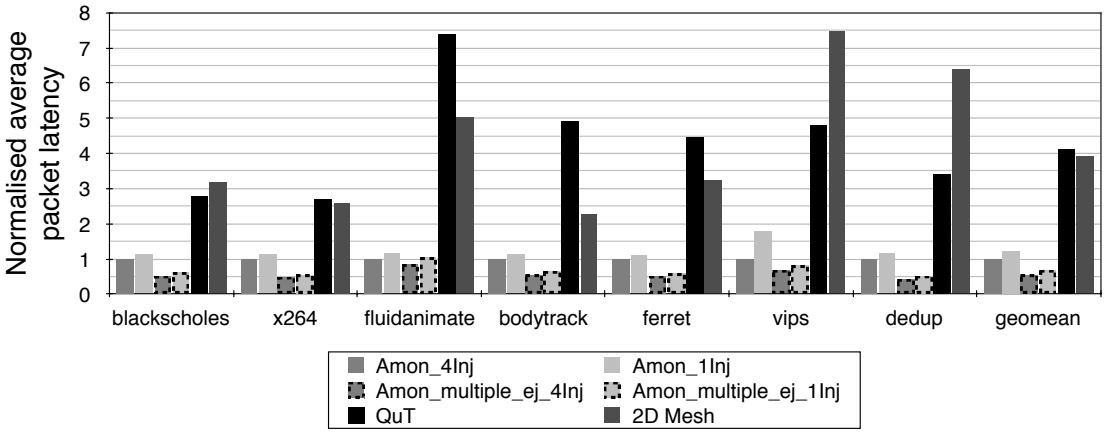


Figure 4.21: Average packet latency normalised to the standard Amon design with four injection channels and one ejection channel for PARSEC workloads.

Power Analysis

Table 4.6 lists laser and MR heating power for conservative and aggressive technology parameters for the data network and LPDN. Reducing the number of injection channels to one per node results in 34% fewer MRs and in turn less MR heating power. The reductions in IL_{max} lead to 42% laser power reduction for conservative parameters, and to $\sim$87% for aggressive parameters. In total, utilising four injection channels per node leads to 1.71$\times$ and 4.11$\times$ higher power consumption for the conservative and aggressive case, respectively; however, even with one injection channel, conservative parameters will not be able to deliver sufficiently low power levels to be competitive to the 2D Mesh. The results for the aggressive parameters, on the other hand, are promising, particularly as most of them are demonstrated devices and not speculations.

Table 4.7 compares the power values to the 2D Mesh when factoring in the control network and dynamic and leakage power. In addition, in order to compare power efficiency, we compute and compare the throughput per Watt (TPW), which is the maximum sustained throughput (in Gbps/node) divided by the consumed power (in our study for uniform random traffic). We only compare Amon to the 2D Mesh for aggressive technology parameters, since the conservative ones lead to an order of magnitude higher power consumption due to overheads at the laser and is thus unlikely to make designers consider it for replacing electrical NoCs.

The power savings of utilising only one injection channel combined with its low performance degradations lead to significant improvements in terms of power efficiency. Compared to the four-injection-channel case, the one-injection-channel design improves power efficiency by 1.95$\times$ and 1.66$\times$ for one and multiple ejection channels,

Table 4.6: Static optical power consumption comparison of Amon (data network + LPDN) with one and four injection channels for conservative and aggressive technology parameters

	Conservative		Aggressive	
	Amon: 1 Inj.	Amon: 4 Inj.	Amon: 1Inj.	Amon: 4 Inj.
ILmax	44.565	46.925	28.764	34.580
Laser Power per λ (W)	0.572	0.985	0.00597	0.0228
Total Laser Power (W)	73.23	126.08	0.765	2.918
Num. of MRs	47206	71270	47206	71270
MR Heating Power (W)	0.944	1.43	0.944	1.43
Total Power (W)	**73.17**	**127.51**	**1.71**	**4.35**

respectively. Although power efficiency can be significantly improved with our approach, only the design with multiple ejection channels and one injection channel – the most efficient Amon design of all – can seriously compete with the 2D Mesh in terms of TPW (though the 2D Mesh is still 17% more power-efficient).

Note that this does not necessarily mean that the 2D Mesh is always the preferred choice over Amon. In fact, there are numerous reasons why Amon is actually the most efficient design. Firstly, the technology library of DSENT for the 2D Mesh is likely very aggressive, particularly clocked at 5 GHz. Other studies in literature or future projections often assume significantly higher power values (i.a. [LBGP14][HJH14]). Secondly, although the 2D Mesh can sustain higher network loads than Amon for synthetic traffic, it is less efficient for realistic traffic with multicast traffic, in which Amon reduces the average packet latency significantly. Besides, latency in the 2D Mesh is highly dependent on the traffic pattern due to latency at intermediate hops, whereas Amon shows constant performance levels across all traffic patterns (apart from hotspot). Thirdly, although laser power should be included in the total power consumption to assess energy efficiency in a meaningful way, the power dissipated at the laser source is off-chip and has, therefore, no impact on the thermal design power. For CMPs in which the on-chip fabric is stressed, Amon would thus impact the power budget less as it has much lower dynamic power consumption than the 2D Mesh.

Table 4.7: Total Power Breakdown and Throughput per Watt of All Amon Designs vs. The 2D Mesh

Power (W)	Amon_4Inj	Amon_1Inj	Amon_mult_ej_4Inj	Amon_mult_ej_1Inj	2D Mesh
Laser	3.018	0.865	3.018	0.865	-
MR heating	1.51	1.024	1.51	1.024	-
Leakage	0.091	0.0766	0.14	0.126	0.41
Dynamic	0.88	0.84	1.9	1.8	2.6
Total Power	**5.5**	**2.63**	**6.57**	**3.64**	**3.01**
Throughput	70	65	163	150	145
TPW	**12.73**	**24.76**	**24.82**	**41.27**	**48.17**

4.4.3 Discussion

Our results show that exploiting MR tuning to reduce the number of injection channels per node improves power efficiency significantly compared to the four-injection-channel case by allowing for large power savings with low performance degradations. The results for aggressive parameters are particularly encouraging: at a total power consumption of 2.63 W, WRONoCs could actually be more power efficient than electrical NoCs, especially as the vast majority of the devices used in the aggressive case have already been demonstrated to be viable. Our injection backend solution is a step forward to making WRONoCs like Amon more power-efficient. In addition, our architectural findings, i.e. the power saving potential of utilising MR tuning to reduce the number of injection channels, can also be applied to other WRONoC architectures that also over-provision the injection bandwidth per node (e.g. [KAH11][KH12][HJH14]).

4.5 Summary

This chapter introduced Amon, a novel WRONoC topology that reduces the static optical power consumption compared to the state-of-the-art topology QuT by decreasing the number of MRs for wavelength-routing and optical path losses. In addition, we avoided the use of NACK packets and REQ retransmissions which enables reductions in both contention on the control network and dynamic power. REQs encoding the packet length can be leveraged to forecast the duration of data transmission which allows destinations to send out ACKs to requesters while receiving data packets at the same time. This novel control scheme improves both performance and dynamic power significantly. In total, Amon saves 21% of power, improves throughput by up to 45%

for synthetic workloads, and reduces latency by 75% on PARSEC traces (on average.) Our study revealed that the standard practice of having a single ejection channel per destination makes WRONoCs particularly susceptible to traffic hotspots, which arise frequently in realistic multi-threaded workloads and might thus lead to large latency overheads which, in turn, increases execution times. Equipping Amon with photodetectors and backend circuitry to eject optical signals from each cardinal direction in the topology can largely resolve this issue: while this modification introduces negligible power overheads ($< 0.5\%$), they double throughput on most synthetic traffic patterns and decreases latency by $\sim$50% on PARSEC traces (on average) compared to the one-ejection-channel design.

Although our architectural proposals improve the efficiency of WRONoC considerably, our study showed that, without further modifications, Amon cannot compete with an aggressive 2D electrical mesh baseline in terms of power consumption, even for most recent SiP devices. The main reason for this is the high static power consumption at the laser source and for MR heating. In addition, scaling WRONoCs above 64 nodes requires either core clustering or significant improvements in SiP devices such as lower device losses, higher laser efficiencies, and improved receiver sensitivities. Although not solving the scalability problem entirely, we take a step forward towards low-power WRONoCs for 64 nodes by proposing to reduce the number of injection channels to one and control the injection waveguide by tuning/detuning MR filters responsible to inject optical signals into the NoC. Although that leads to latency overheads ($\sim$20%), it roughly halves power consumption, thus improving the overall power efficiency significantly.

We close this chapter by highlighting that the most power-efficient Amon design, i.e. with one injection channel and ejection channels for each cardinal direction, has a 21% higher power consumption than the 2D Mesh baseline prior to network saturation, but also decreases latency by $\sim$5.5$\times$ on PARSEC traces on average. For CMPs that stress the on-chip network, Amon may, therefore, become the preferred choice when a high-performance NoC is required, especially if the dynamic power of the 2D Mesh exceeds the thermal design power limit when dealing with high network loads.

Chapter 5

Combining Electrical and Optical Links

5.1 Introduction

Chapter 4 showed how optical data transmission can be utilised to overcome the energy inefficiencies of electrical interconnects for long distances, and how sophisticated network architectures like Amon can largely reduce power consumption in all-optical NoCs; however, the static power overheads to enable all-to-all optical communication are still considerable, and the overheads of signal conversion for short distances detrimental compared to electrical links. As discussed in Section 3.3, recent literature is replete with proposals that combine electrical and optical links in the NoC's topology which demonstrate that discarding electrical interconnects in NoCs altogether leads to unnecessary inefficiencies. In the current state of electronics and SiPs, electrical links are actually superior in terms of performance and power consumption if distances are short enough.

Our analysis and comparison of electrical and optical interconnects in Section 2.4 showed that the properties of these technologies are somewhat complementary: while electrical links are efficient for short-distance communication, optical links suffer from latency overheads for EO/OE; however, as distances grow, electrical interconnects become increasingly power-hungry as they require repeaters. Optical links, on the other hand, do not, and can leverage propagation delay of light with only marginal energy overheads even for long distances. Depending on the technology used, the cross-over point (with regard to on-chip distance) at which optical links become more efficient than electrical links varies. Previous proposals used approximate distances based on

intuition rather than detailed analysis [BP14] [KMP+10][PKK+09], which may not result in the most efficient design and possibly limit the potential of this approach. A detailed analysis would explore the cross-over distance P_{dist}, and a NoC would ideally route packets on electrical links for distances $\leq P_{dist}$, and optical links otherwise. At the same time, optical links could be leveraged to decrease the average hop count in a NoC – two measures that would result in very low dynamic power.

As discussed in Chapter 2, static optical power scales proportionally to link bandwidth; however, this relationship is exponential and could, depending on the bus architecture, grow sharply. This chapter discusses this issue in more detail for the R-SWMR bus, on which laser power is particularly susceptible to link bandwidth due to high path losses. Therefore, although R-SWMR buses are a very efficient design, bandwidth scaling is critical.

As many application domains underutilise the on-chip network, the question arises whether high-bandwidth optical links are necessary to satisfy the communication demands of a system, or whether link bandwidth can be lowered to avoid static power being wasted in low-utilisation phases. Although utilising low-bandwidth links inevitably results in performance degradation, they could still be superior compared to an electrical NoC if the distances for which optical links are used are large enough. The serialisation delay imposed by low-bandwidth optical links needs to be studied in relation to the latency of packets if routed through an electrical NoC (i.e. hop latency, serialisation and contention delay) to identify the distances at which the optical links are superior.

This chapter analyses electrical and optical interconnects with respect to latency and power consumption based on the distance on chip, and proposes a novel topology 'Lego' that adopts a distance-based combination of these two interconnect technologies and aims to lower power consumption while maintaining performance levels. In particular, it makes the following novel contributions:

- A detailed study of optical interconnects for current SiP devices and a comparison to electrical interconnects for a 22 nm technology which reveals that laser power can be reduced by implementing higher numbers of low-bandwidth optical links rather than few high-bandwidth links. Inserting higher quantities of low-bandwidth optical links into a topology can thus provide similar bisection bandwidth at lower laser power.

- The NoC design 'Lego' that implements higher quantities of low-bandwidth optical links in its topology to reduce laser power, and utilises a distance-based

routing approach where a simple electrical NoC is used for distances lower than a parameter P_{dist} and optical links otherwise. This approach hides the serialisation delay imposed by low-bandwidth optical links, deploys both interconnect technologies at distances where they are most energy-efficient, and offers low average hop counts.

- Lego exhibits 3.25× and 2× higher TPW compared to an electrical 2D mesh and the hybrid NoC Meteor [BP14]. Compared to an all-optical mesh topology like LumiNOC [LBGP14], Lego decreases packet latency by 70% on realistic workloads while consuming slightly less power. For current technologies, Lego is particularly suited for CMPs that require high network utilisation as its low dynamic power is maintained even for very high injection rates.

5.2 Optical vs. Electrical Links

Deciding on when to utilise optical and electrical links depends on a number of design trade-offs affecting latency and power consumption. This section discusses these implications and identifies their benefits and drawbacks.

Power Consumption

Optical Links As discussed in detail in Section 2.4.1, static power consumed at the laser source and for MR heating dominates the power consumption of optical interconnects, which necessitates to keep the numbers of MRs and IL_{max} low. The number of wavelengths on an optical link not only impacts the laser source itself but also IL_{max}: the more wavelengths are used to transmit data on a waveguide, the more modulators and MR filters are required, which in turn increases MR-through loss. Depending on the number of wavelengths and the number of receivers on a bus, this can lead to large laser power overheads.
Figures 5.1 and 5.2 plot the laser power consumption for a 10 mm SWSR and a SWMR bus, respectively (values modelled with DSENT [SCK+12] with SiP technology parameters of Table 5.1). For the SWSR bus, laser power grows only slightly greater than linearly with the number of wavelengths, suggesting that the additional losses on the waveguide are not considerable with only one receiver. Crosstalk within a waveguide, which also increases along with the number of wavelengths, also does not seem significant in this configuration.

Table 5.1: SiP Technology Parameters (loss values based on measurement values of a recently demonstrated SiP link in 45 nm silicon-on-insulator technology [OMS+12, GMS+14, LSZP14]

Loss Parameter	Value	Loss Parameter	Value
Splitter	0.2 dB	MR-through	0.01 dB
Coupler	1 dB	Waveguide propagation	0.3 dB/mm
MR-drop	0.5 dB	Photodetector	0.1 dB
Waveguide crossing	0.04 dB		
Laser efficiency	25%	MR heating	20μW/MR

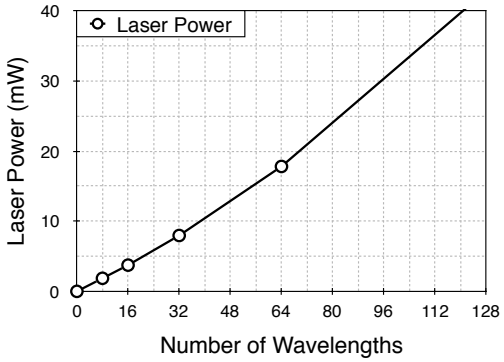


Figure 5.1: Laser Power vs. Num. of Wavelengths on a SWSR Bus

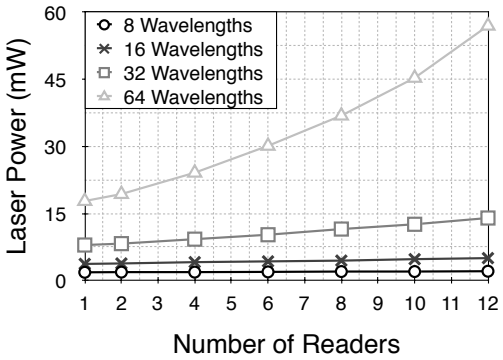


Figure 5.2: Laser Power vs. Num. of Receivers on a SWMR Unicast Bus

For a SWMR bus, however, the impact of increasing number of wavelengths are much more significant as the number of receivers increases (note that Figure 5.2 plots a SWMR *unicast* bus, i.e. only one receiver will draw power from the laser source at any time, all other receivers are switched off (like in the R-SWMR bus in Section 2.3.2)). The plots illustrate that both the number of wavelengths and the number of readers must be apportioned carefully to avoid excessive laser power, particularly as SWMR buses are much more suitable than SWSR buses to design optical NoCs (as discussed in Section 2.3).

MR-through loss is the main reason for this exponential relationship. Its impact on laser power on a SWMR bus depends on the absolute through-loss value per MR and the losses of other devices on the link. For instance, the laser power results in Figures 5.1 and 5.2 assume 0.3 dB/mm waveguide propagation loss; however, loss values of 0.1 dB/mm or even 0.0271 dB/mm have already been demonstrated (as discussed in the previous chapter). The lowest demonstrated MR-through loss is 0.01 dB/mm (to the best of our knowledge), and although lower values have been assumed in recent studies, these values are speculations rather than demonstrated devices.

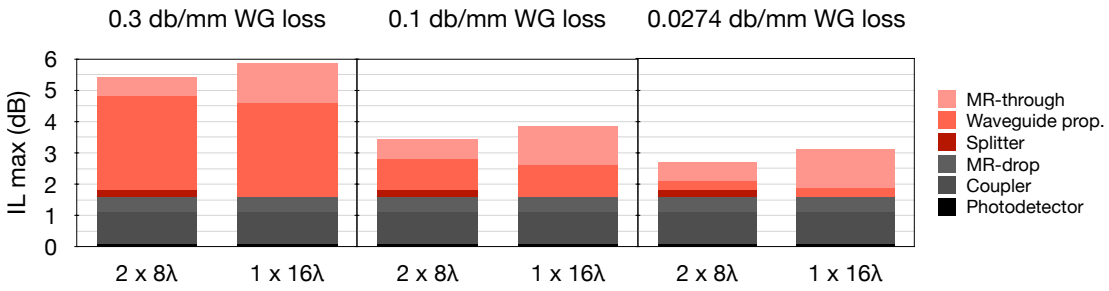


Figure 5.3: IL_{max} on a SWMR bus consisting of 1) two 8λ buses and 2) one 16λ bus for different waveguide (WG) propagation losses found in the literature.

Lower waveguide propagation values increase the impact of MR-through loss on IL_{max} on a SWMR bus. Figure 5.3 illustrates this by plotting the IL_{max} breakdown for 16λ SWMR buses consisting of 1) two 8λ buses and 2) one 16λ bus. Note that the total device loss values in black do not differ between the different designs. Implementing a 16λ SWMR bus using two 8λ buses requires an optical splitter that adds 0.2 dB [1] loss to the critical path (assuming both buses are supplied with *one* laser source (rather than one for each 8λ bus) to keep the impact on chip packaging the same). The main observation is that savings in MR-through losses on the 8λ buses outweigh the splitting loss overheads compared to the 16λ bus (which does not require splitting). The loss savings of the $2 \times 8\lambda$ bus increase as waveguide loss decreases since MR-through loss becomes more important to the overall IL_{max}, and ranges from 8% in the 0.3 dB/mm case to 14% in the 0.0274 dB/mm case. In either case, although incurring area overheads, implementing a higher quantity of low-bandwidth links is more laser power efficient. As the laser power plots in Figure 5.2 suggest, these savings would be even larger for link bandwidth higher than 16λ. The previous chapter discussed that waveguides and MRs are compact structures, therefore area overheads would probably not lead to infeasible layouts. A more detailed analysis of this will be provided in the following section.

Electrical Links The analysis in Section 2.4.1 showed that electrical links are more energy-efficient for short distances as they do not require EO and OE conversion; however, for link lengths > 0.5 mm, the relatively-distance-independent energy consumption of optical data transmission outperforms electrical links (for the used 22 nm technology). From an energy perspective, it is therefore only beneficial to utilise electrical links for destinations < 0.5 mm, which is a short distance for typical tile dimensions

[1] 0.2 dB splitter loss is a conservative assumption and allows us to assess our approach pessimistically. Optical splitters of less than 0.1 dB have been demonstrated [WGWS15] and would favour our approach even further.

Table 5.2: Core Clock cycles (EO + t_{prop} + OE) for transmitting 64, 256, and 576 bit packets at different optical bandwidth. For simplicity we assume t_{prop} = 1 and OE = 1.

Number of Wavelengths	**Packet Size**		
	64 Bits	**256 Bits**	**576 Bits**
4λ	10 (8+1+1)	34 (32 +1+1)	74 (72 +1+1)
8λ	6 (4+1+1)	18 (16 +1+1)	40 (36+1+1)
16λ	4 (2+1+1)	10 (8 +1+1)	20 (18+1+1)
32λ	3 (1+1+1)	6 (4+1+1)	11 (9+1+1)
48λ	3 (1+1+1)	5 (3+1+1)	8 (6+1+1)
64λ	3 (1+1+1)	4 (2+1+1)	7 (5+1+1)

of 1-2 mm (e.g. [VTL+16] [BSP+16]). This cross-over point will increase for smaller technology nodes (e.g. 14 nm, 7 nm) since, although electrical interconnects do not scale nearly as well as transistors, energy is still decreased compared to larger nodes. Besides, the energy required to traverse a router adds up for every hop, mainly for buffering and crossbar traversal. These energy overheads must be considered when comparing electrical to optical connections.

Latency

As shown in Section 2.4.2, large distances on chip can be traversed optically within one 5 GHz clock cycle, which is much lower than electrical signal propagation on a repeated wire (10.45 ps/mm vs. 131 ps/mm), and data transmission is less distance-dependent in terms of energy. Although all optical components add to the optical delay, the major contributor is data modulation, i.e. the time it takes to serialise a packet based on the available bandwidth. Table 5.2 further outlines this by listing the impact on the delay of common packet sizes in NoCs for varying number of wavelengths, assuming link propagation delay of 1 cycle for simplicity, 10 Gb/s modulators and 5 GHz clock frequency. These values are an important guideline to trade-off power and latency. For instance, increasing link bandwidth from 16λ to 32λ decreases latency only by one clock cycle, but more than doubles laser power (see Figure 5.1). Bandwidths lower than 8λ introduce too much latency and are unsuitable to satisfy on-chip bandwidth demands.

As discussed before, electrical NoCs can satisfy current communication demands in CMPs for small to moderate core counts (< 32), but become increasingly energy-inefficient for larger sizes due to transferring data packets over large distances. Optical

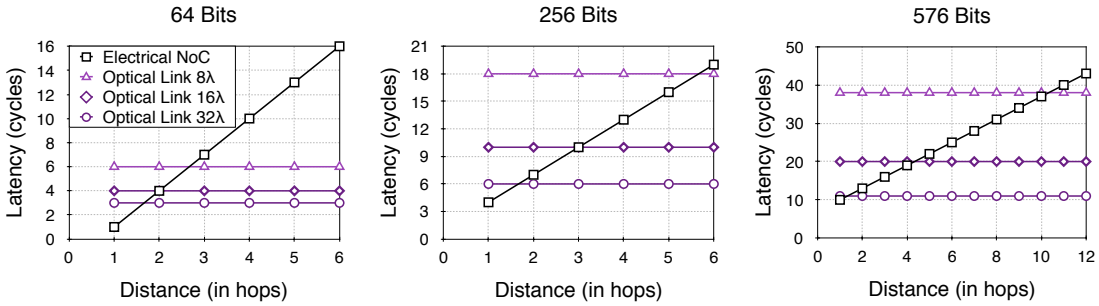


Figure 5.4: Latency for transmitting data in electrical vs. optical NoCs (5 GHz core clock, 10 Gb/s modulators, and 3 cycles per hop in an electrical NoC with 64-bit flits)

links, however, are ideal for these distances both in terms of energy and latency. Although lowering the bandwidth on optical links increases serialisation delay, latency and throughput can be maintained compared to electrical links – with higher energy efficiency – if distances are large enough. This reduces static power on optical links due to lowered optical bandwidth, and at the same time maintains (and for sufficiently large distances even improves) the latency and throughput properties of electrical NoCs.

In order to combine electrical and optical links with this approach efficiently, optical delays must be studied in detail and compared to the delay in electrical NoCs. Although electrical links do not need EO and OE conversions, the only energy-efficient way of reaching distant cores is through several hops in a topology, which includes router delay. If we assume aggressively pipelined routers that can be traversed in two clock cycles, one hop would take 3 cycles (assuming no contention). While this delay adds up for each additional hop to reach a destination, hardly any delay is added on optical links when the distance increases (assuming direct connections).

Figure 5.4 illustrates this comparison for a range of different optical bandwidths on an optical link with an electrical NoC for common packet sizes (64 bits for coherence traffic and 576 bits for cache line transfers, with 256 bit being a rough average). The charts assume zero load in the network, and latency is measured from the time a packet leaves the output port of the source router and arrives at the input port of the destination router, e.g. sending to a direct neighbour would require one link traversal, to a node in two hop distance two link and one router traversal, etc. Note that latency (in cycles) in the electrical NoC and optical links also depends on the clock frequency; however, the trends would remain the same as both the electrical NoC (lower frequency allows for fewer pipeline stages in routers/links) and optical links (lower frequency allows to modulate more bits in one clock cycle) would speed up similarly.

For small-sized coherence traffic (64 bits), which comprise ∼70% of the total messages in a large number of multi-threaded applications [LNP+13], the cross-over distance (P_{dist}) at which optical links are superior to electrical links is quite small (< 2 hops for 8λ). As cache line transfers (576 bits) require larger throughput, P_{dist} can be up to 10 hops compared to 8λ. Although this seems quite large, these distances are not necessarily uncommon, e.g. the average hop count in a 2D mesh for 64 or 256 nodes is 5.3 and 10.6, respectively. In addition, sending packets through an electrical NoC imposes further latency due to buffer and link contention.
Therefore, a topology could improve performance and power consumption by combining low-bandwidth optical links for large distances with an electrical NoC for short distances. The cross-over distance P_{dist} could be determined based on the bandwidth demands of the application and latency and energy values of the used technology. The optical links would allow to maintain the zero load latency properties of the electrical NoC if distances are large enough and would 1) lower energy/dynamic power significantly (no hops, no repeaters) and 2) take load off the electrical NoC, thereby decreasing contention delay. The following section presents the novel topology 'Lego that adopts this approach.

5.3 LEGO: A Locally-Electrical Globally-Optical NoC

Lego is a NoC comprising topology, layout, and routing algorithm that exploits the design principles discussed above to lower laser and dynamic power, and packet latency.

5.3.1 Topology

Lego implements higher numbers of low-bandwidth optical links to reduce laser power and utilises them for larger distances to hide serialisation delay and to deploy optical data transmission where it is most energy-efficient. Optical links are paired with an electrical NoC that is only utilised for communication with nodes at small distances to reduce energy and latency. Figure 5.5 illustrates Lego's topology, which combines an electrical 2D mesh (electrical links in green) with optical links in its rows/columns. Each node can transmit to every other node in the same row/column using its optical link if the destination is further away than the cross-over distance parameter P_{dist}, which determines whether packets are routed over the electrical or optical links.

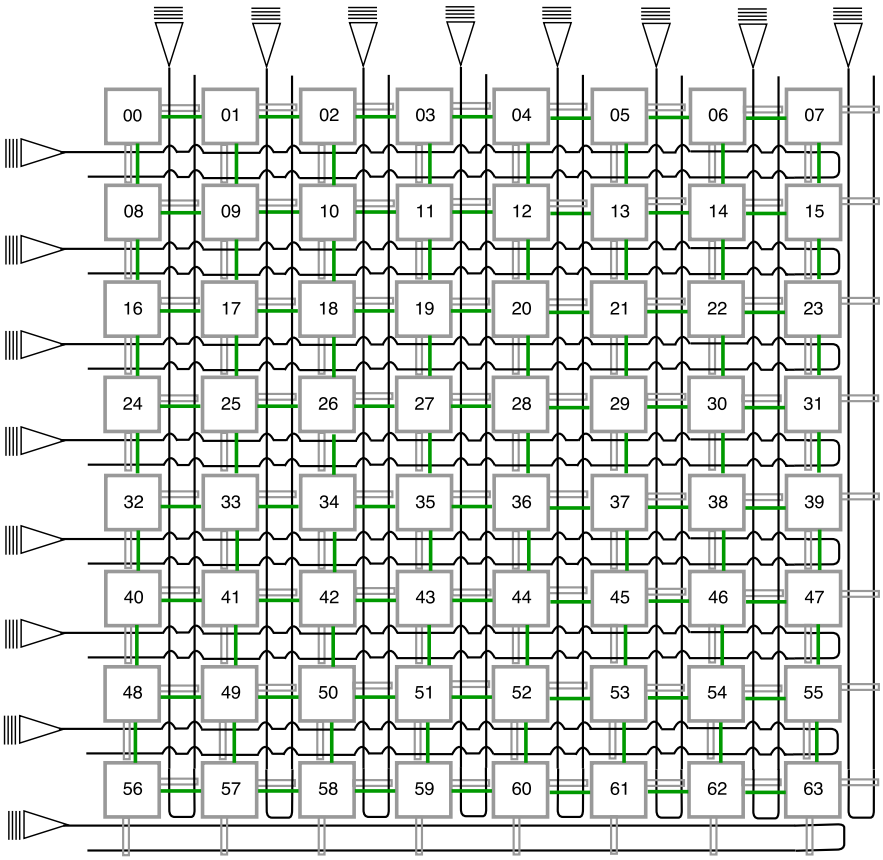


Figure 5.5: Lego Topology for 64 Nodes

5.3.2 Routing Algorithm

Our routing algorithm aims to minimise link traversals and always chooses the link/interconnect technology that offers the lowest energy and latency to keep dynamic power and latency as low as possible. Therefore, based on the relative position of a sender and its destination, the former either sends on an optical link, electrical link, or a combination thereof. If the number of hops required to reach the destination node is greater than P_{dist}, a sender would choose to transmit over the optical links in Lego, and on the electrical links otherwise. Routing is classified into four cases based on the location of a sender S and its destination D:

1. **Case 1 (Figure 5.6a):** *D is in $\leq P_{dist}$ distance to S*: S routes its packets over the electrical mesh to D.

2. **Case 2 (Figure 5.6b):** *D is in $\geq P_{dist}$ distance to S, but in either the same row or column group of S*: S routes its packets over its optical link to D.

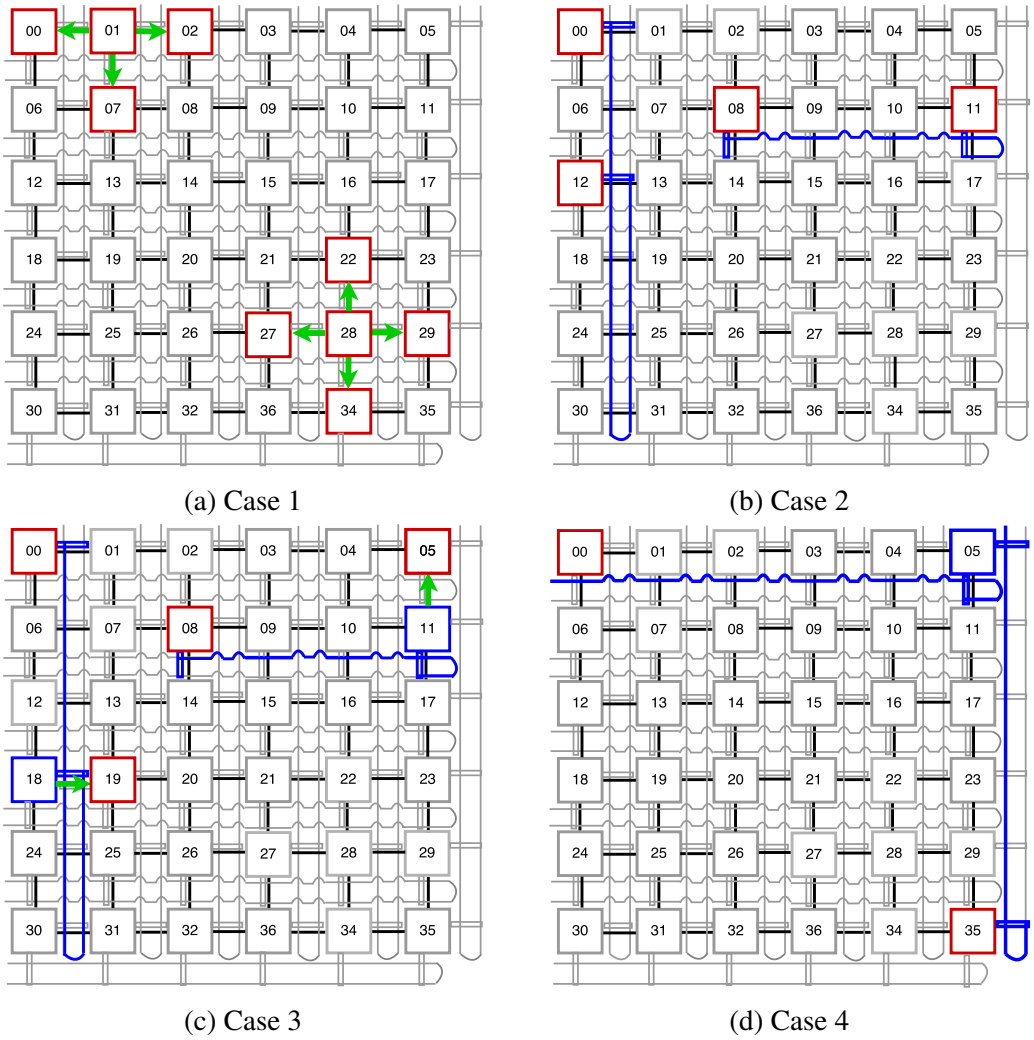
3. **Case 3 (Figure 5.6c):** *D is* $\geq P_{dist}$ *away from S, but* $\leq P_{dist}$ *to a node N that is in the same column as S*: *S* uses its optical (column) link to route the packet to *N*, which will then proceed to route the packet to *D* using the electrical mesh (since *D* is $\leq P_{dist}$ to N). Similarly, if *D is* $\geq P_{dist}$ *away from S, but* $\leq P_{dist}$ *to a node N that is in the same row as S*: *S* uses its optical (row) link to route the packet to *N*, which will then proceed to route the packet to *D* using the electrical mesh (since *D* is $\leq P_{dist}$ to N).

4. **Case 4 (Figure 5.6d):** *D is* $\geq P_{dist}$ *away from S and* $\geq P_{dist}$ *to every node that is in the same row/column as S*: *S* uses its optical (row) link to route the packet to the node that is in the same column as *D*, which will then proceed to route the packet to *D* using its optical (column) link (since *D* is $\geq P_{dist}$ to N).

Effectively, with this routing algorithm, the maximum number of hops that a packet has to travel is $1 + P_{dist}$ (Case 3), which is highly efficient in terms of dynamic power, particularly as this routing algorithm – in combination with Lego's topology – always chooses to route a packet over the interconnect technology that requires the lower energy for data transmission (i.e. electrical links for short, optical links for long distances).

Optical Router Groups

All nodes that belong to the same optical router group, i.e. nodes within the same row or column, are connected to each other in a crossbar fashion, with one R-SWMR bus for each node. Figure 5.7 gives a close-up to an optical router group with 8 nodes. For simplicity, the figure only shows the buses of a few nodes, and only the data buses (as described in Section 2.3.2, every R-SWMR bus requires an additional control bus to notify destinations to tune/detune their MRs). Every node has its own bus for sending. Router groups adopt a U-shaped 'double-back' waveguide layout proposed by Li et al. [LBGP14] which allows nodes to reach every other node by modulating data on the transmit side of the link (red). All receiving nodes filter out the optical data on the receive side (green). In total, 16 waveguides are required in each optical router group (8 data buses, 8 control buses).
Note that the optical router group in Figure 5.7 shows the layout without the underlying electrical NoC. As described, the proposed routing algorithm chooses the optical link only if the destination is $\geq P_{dist}$ away. This, in turn, means that nodes do not need to place MR filters on the R-SWMR buses of other nodes that are within P_{dist} to them



(a) Case 1

(b) Case 2

(c) Case 3

(d) Case 4

Figure 5.6: Routing cases in 6×6 a Lego with $P_{dist} = 1$ (green links indicate hops over electrical links, blue links over optical links)

since these nodes would use the electrical links to transmit data to them. Fewer MRs placed on the buses reduce laser power (fewer MR-through losses), and MR heating. Therefore, routing based on P_{dist} is not only more efficient from a latency/energy point of view, it also allows to reduce the resources on the optical links.

Each node in the router group owns one SWMR control bus for its SWMR data bus. Both buses have the same layout, merely the number of wavelengths, and thus modulators and MR filters, differ. In accordance with the functionality of an R-SWMR bus, transmitting data over optical links thus obeys the following process:

1. Initially, all nodes are detuned and do not filter the wavelengths.

2. When a node wants to send data, it first sends out a control packet containing the

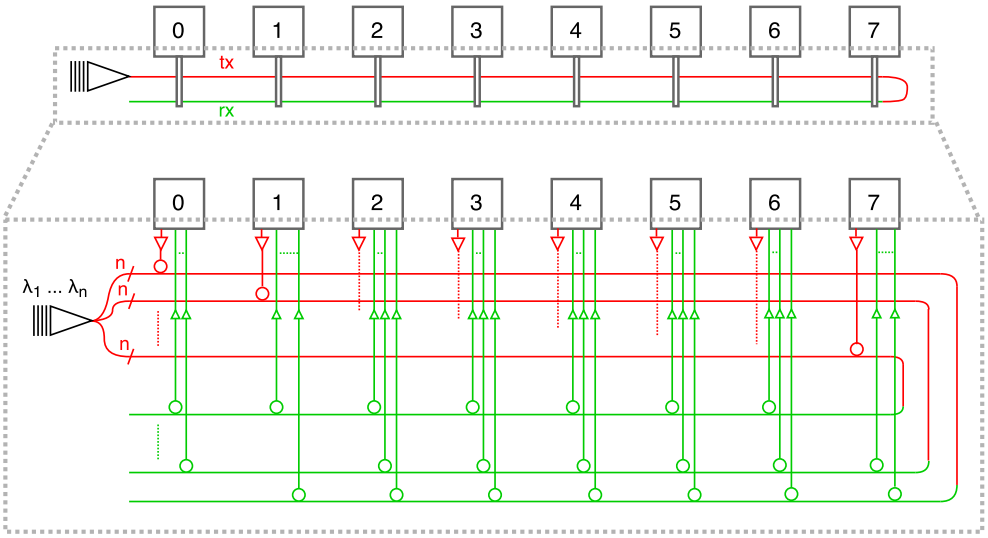


Figure 5.7: Optical router group layout. The control network has the same layout but is omitted for simplicity. Nodes modulate data on the tx path (red) and receive on the rx path (green).

destination node and packet size on the control network to all receivers.

3. The destination node tunes in and all other nodes keep their MR filters on the data bus detuned. The packet size indicates the duration of which MR filters have to stay tuned/detuned.

4. The sender transmits its data.

Control packets are very small since only destination ID and packet length must be encoded to notify nodes about the destination and duration of data transmission, and therefore require low bandwidth. NoCs having to support two packet sizes (64 bit and 576 bit for coherence traffic and cache line transfers, respectively) require just 1 bit. $log_2(N)$ bits are necessary to encode the destination for N receivers on the bus. N depends on the router group size and P_{dist} since nodes in $\leq P_{dist}$ receive data on the electrical NoC. In a 64-node (8×8) Lego, a sender has at most 6 receivers (corner node for $P_{dist} = 1$). Therefore, a control packet in a 64-node Lego can be at most 4 bits, which requires just 2 wavelengths to be modulated in one core clock cycle.

Layout

Implementing ONoCs with SiPs requires a careful consideration of layout and device technologies. Targeting low laser power by providing a larger number of low-bandwidth links rather than few high-bandwidth links only decreases laser power if the

higher number of waveguides does not lead to a higher number of waveguide crossings, which could increase IL_{max} and possibly diminish some of the power savings. In Lego's layout, waveguide crossings occur when optical links located in the columns cross the ones in the rows. For that reason, we assume 3D-integration with the optical circuitry of row and column router groups placed separate photonic layers to eliminate in-plane waveguide crossings [BPH+11].
An optical router group requires 16 waveguides for the data and control buses. Although current technologies allow waveguide dimensions of 520 nm width [BCB+14], sufficient clearance is needed to avoid optical signal interference and crosstalk, which can conveniently be ensured with a waveguide pitch of 5 μm [PD10]. MRs with a 5 μm diameter are often considered in the literature [LBGP14] and with 5 μm clearance needed between the buses, this adds up to 15 μm width per bus [LBGP14]. For 16 waveguides per router group, this layout requires < 0.25 mm for all optical links in the rows and column router groups. Conventional die sizes of 225 mm^2 would allow common tile sizes of 1 mm^2 for 64 nodes while providing sufficient area for interfacing and placement of the SiP devices in the topology's rows and columns. Lego is therefore technologically feasible and allows to be conveniently integrated into conventional CMPs.

5.4 Evaluation

5.4.1 Methodology

This study aims to investigate how the introduced approach of combining electrical and optical links in a NoC topology compares to other designs proposed in the scientific literature in terms of performance, power consumption, and area, and compare Lego to a both an electrical, hybrid, and all-optical NoC. In addition, it identifies which choice of the distance parameter results in the most efficient design.

Experimental Set-up

This study was conducted with Graphite [MKK+10], a distributed parallel simulator infrastructure that allows for fast simulation and modelling of NoCs within a large-scale multi-core environment. For estimating power consumption, Graphite integrates DSENT [SCK+12], the state-of-the-art NoC modelling tool capable of modelling both electronic and SiP components, as well as their interaction. The tiled CMP modelled

in this study is the standard configuration of Graphite in which each tile has private L1I/L1D (16/32 kB) and L2 (512 kB) caches, and a memory controller with 5 Gb/s bandwidth. Caches implement the MSI full-map directory-based cache coherence protocol. Graphite was configured with a 22 nm low-voltage technology library of DSENT, 5 GHz core, router, and link clock, 10 Gb/s modulators/detectors, and 1 mm tile dimensions (square tiles). Table 5.1 lists the technology parameters used to estimate laser power and MR heating. All NoC topologies are simulated with 64 nodes with an 8×8 layout.

Alternative NoCs and Configuration

Several ONoCs utilising optical links for long-distance and electrical links for short-distance communication have been proposed in the recent literature. We compare Lego to different design approaches of combining these two technologies to evaluate its effectiveness. In particular, we compare Lego to the topologies Meteor and LumiNOC, as well as an aggressive baseline 2D electrical mesh (2D Mesh).
We study three different distances in *Lego* that denote the cross-over point at which optical links will be used for communication. Lego_dist1 denotes a Lego implementation in which electrical links are utilised for nodes within 1-hop distance (i.e. direct neighbours), and optical links otherwise. In the same token, Lego_dist2 and Lego_dist3 denote Lego implementation in which packets are routed on the electrical mesh if a destination is at most 2 and 3 hops away, respectively. This allows to study the effects of putting more emphasis on the optical or electrical network, as well as distances for which low-bandwidth optical links are the most effective. Lego implements R-SWMR buses with eight wavelengths (8λ) as our study showed that 8λ could be an efficient design point in terms of laser power consumption and serialisation delay. In addition, R-SWMR buses have control buses with 2λ on the control bus to allow 1-cycle modulation, and each row/column group is supplied with one laser source.
Meteor [BP14] is a topology that implements a 2D electrical mesh and overlays it with an optical network that can be accessed through hub routers. Photonic Regions of Influence (PRI) determine the grouping of nodes to the hubs. With an 8×8 layout, their study shows that grouping 16 nodes to each PRI is the most efficient design variant. We divide the 8×8 layout into four square 4×4 submeshes and place the hub router in the middle of each submesh for the highest efficiency. For inter-hub communication, each hub is equipped with 32λ SWSR buses – one separate bus to each hub. Meteor

thus constitutes a topology in which multiple nodes share a small number of high-bandwidth optical links for long-distance communication.
LumiNOC [LBGP14] utilises the same topology as Lego, but does not include electrical links in its topology, i.e. data communication is optical only. This topology is often referred to as 'optical mesh' in the literature and performs XY-routing to forward packets to their destinations, leading to a maximum number of two hops through the network. LumiNOC utilises a shared optical bus to connect rows and columns; in this study, we are interested in the effect of combining electrical and optical links and are only interested the topologies rather than shared bus arbitration mechanisms. Therefore, to have a fair comparison, LumiNOC's row/column connections are the same as Lego's, i.e. 8λ R-SWMR buses (with 2λ on the control bus). Chapter 6 will discuss shared optical buses, as proposed in LumiNOC, in much greater detail.
2D Mesh is an aggressive baseline electrical mesh with two cycle router and one cycle link traversal delay. Packets are routed based on XY-routing. We include this NoC to reveal the benefits and trade-offs of adding optical links to electrical topologies, and whether this approach can outperform existing designs in the industry [Ram11].
All NoC were simulated with 64-bit flits and one-flit path width through the routers and electrical links. Graphite simulates NoCs with virtual-cut through switching and models output link contention. All NoCs are modelled with a router traversal latency of two cycles, and electrical link traversal latency of one cycle.

Workloads

We evaluate all NoCs under both synthetic and realistic workloads.
For synthetic workloads, we apply uniform random, bit complement, and tornado traffic to stress different corner cases of the topologies. Neighbour traffic is not included in this study since Lego, Meteor, and 2D Mesh all utilise the same electrical links to its neighbours, and thus exhibit similar performance. LumiNOC does not include electrical links, and the simulation results in the previous chapter has shown that ONoCs are inferior to electrical NoCs for neighbour traffic due to EO/OE overheads. Synthetic traffic packets are 256 bits.
For realistic workloads, we simulated a range of different applications from the SPLASH-2 [WOT+95] and PARSEC [BKSL08] benchmark suites, which are the most widely used applications in the scientific community, and represent diverse workloads of shared-memory, multi-threaded applications.

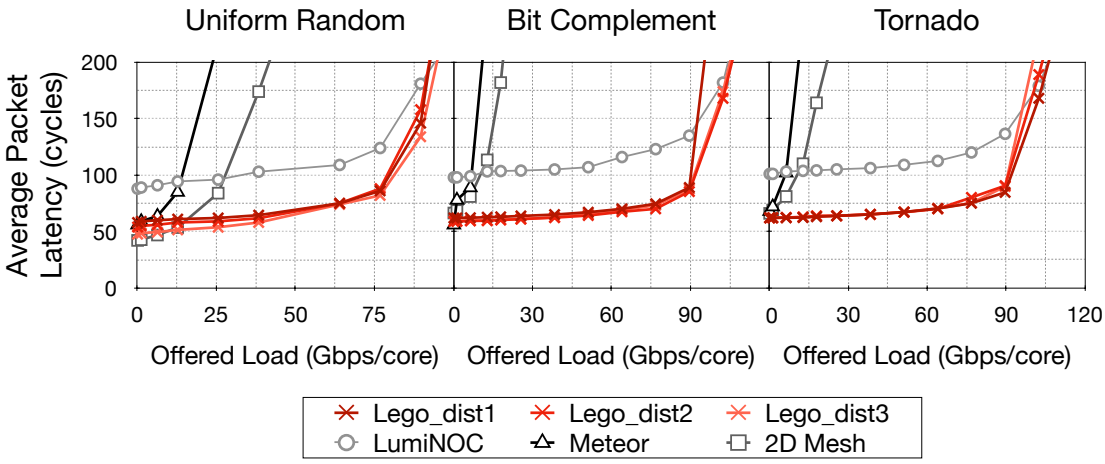


Figure 5.8: Average Packet Latency for Synthetic Traffic

5.4.2 Performance

Synthetic Workloads

Figure 5.8 depicts the average packet latency for different injection rates. Changing the distance parameter in the Lego NoCs does not have a large impact on latency or throughput; however, increasing the parameter can have an impact on packet latency for low network loads, as the results for uniform random traffic show: packet latency is the lowest for the distance parameter set to 3, indicating that using optical links for destinations < 3 is not the most efficient design point.

All Lego implementations saturate significantly later than Meteor and the 2D Mesh. The higher abundance of optical links combined with the very low average hop count in Lego allows it to sustain higher network loads. LumiNOC saturates at similar injection rates; however, LumiNOC is inefficient in terms of latency as it almost doubles packet latency compared to all other NoCs (for low/moderate injection rates). This is likely because LumiNOC relies on low-bandwidth optical links only, which is inefficient for short distances. To combat this, bandwidth on LumiNOC's links could be increased; however, this would also increase power consumption significantly as the previous study has shown.

Realistic Workloads

Figure 5.9 illustrates the average packet latency for a range of SPLASH-2/PARSEC applications normalised to the 2D Mesh baseline. The throughput improvements of

Lego in synthetic traffic do not translate to average latency reductions for realistic traffic. In fact, only Lego_dist3 can provide similar packet latency as the 2D Mesh (on average). This may be due to the packet sizes sent in realistic traffic and the bandwidth on the optical links: cache lines can be much larger than the 256-bit packets in synthetic traffic and are thus more bandwidth critical. While Lego is efficient for smaller packet sizes that do not require much bandwidth (i.e. coherence traffic), lowering the bandwidth to reduce laser power on optical links makes Lego inefficient for cache line transfers. Both 2D Mesh and Meteor only have high-bandwidth links in their NoCs (i.e. 64-bit wide) and can thus transfer large cache lines faster. The highest impact on this can be observed in the latency results of LumiNOC, which exhibits the highest overheads as it relies on low-bandwidth links only. Lowering the bandwidth of optical links should thus be considered carefully for realistic traffic; however, if distances are large enough, the serialisation overheads can be hidden (e.g. Lego_dist3). In this case, lower bandwidth links can provide power savings without increasing latency noticeably.
Figure 5.10 shows the impact that the average packet latency has on the overall execution time of the applications (note that the y-axis begins at 0.85). Compared to the 2D Mesh baseline, the packet latency overheads of Lego translate to at most 1% overheads in execution time (for Lego_dist1). Meteor shows the lowest execution time (1% less than 2D Mesh). The overheads imposed by LumiNOC illustrate that relying just on 8λ is not sufficient to compete with electrical NoCs in terms of performance.
In summary, combining electrical links with low-bandwidth optical links does not translate to overheads in the execution time (on average) if the distance parameter is large enough (i.e. three), showing that 8λ link bandwidth can satisfy the demands of the sample applications (to a varying degree) as long as they are combined with electrical links for short distances. The amount of (dynamic) power that can be saved by this approach will be discussed in the following.

5.4.3 Power Consumption

Synthetic Workloads

Figure 5.11 plots the dynamic power consumption versus the injection rate. The results confirm the hypothesis that Lego – thanks to its very low average hop count and utilising electrical and optical links in cases where they are more energy-efficient – requires significantly less dynamic power than the alternative NoCs. Only LumiNOC consumes

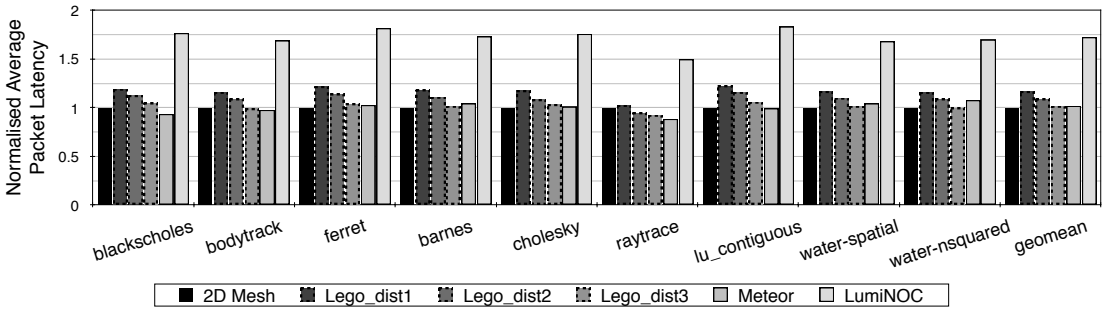


Figure 5.9: Average packet latency for SPLASH-2/PARSEC applications normalised to 2D Mesh

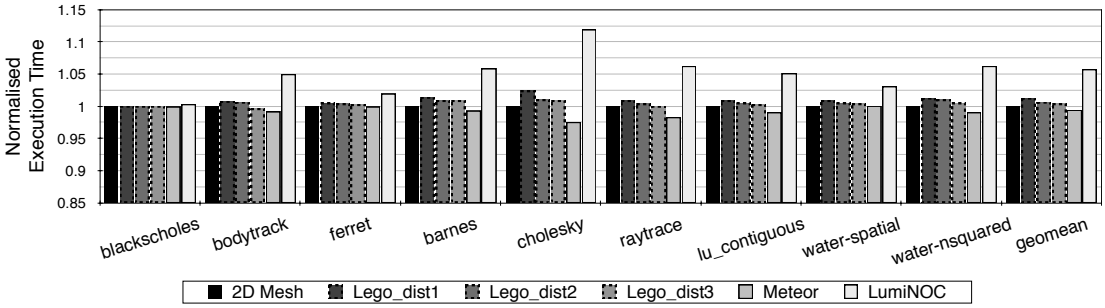


Figure 5.10: Execution time of SPLASH-2/PARSEC applications normalised to 2D Mesh

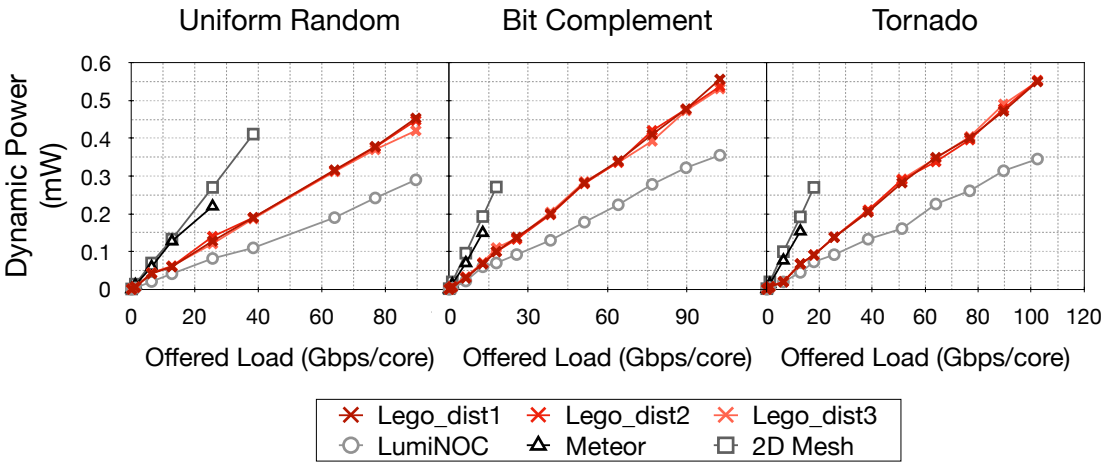


Figure 5.11: Dynamic Power vs. Offered Load for Synthetic Traffic

less dynamic power as it utilises low-energy optical data transmission only; however, as previously discussed, LumiNOC cannot compete in terms of latency. Figure 5.12 shows the power breakdown of the different NoCs. To have a fair comparison, we assume that the same amount of buffering for each NoC is equally distributed across the input ports to estimate leakage power. In particular, we model a typical amount of

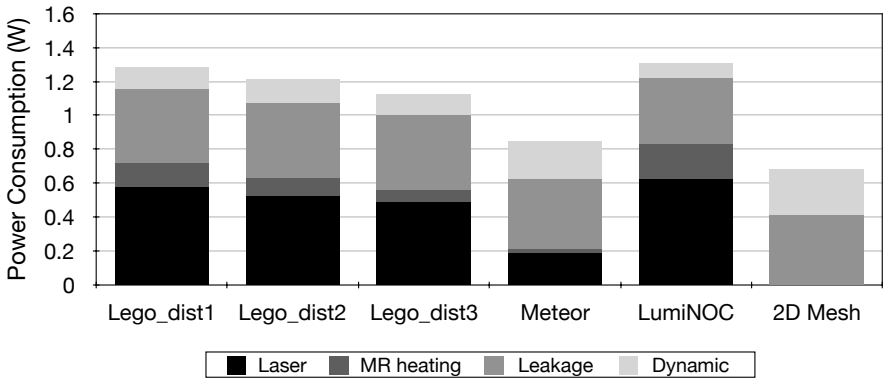


Figure 5.12: Power Breakdown (dynamic power represents the power consumed for uniform random traffic at the saturation point of Meteor)

buffering for a 2D Mesh NoC, which is buffered for 6 virtual channels per input port, each 4-flit deep with 64-bit flits [SCK+12].

Laser power and MR heating consume a significant amount of Lego's power. Meteor relies more on electrical links than Lego and implements fewer optical links, therefore its laser and MR heating power is lower; however, this also results in much earlier network saturation for all workloads in synthetic traffic. As shown in Section 5.2, conservative waveguide loss parameters like the ones assumed in this study (0.3 dB/mm) de-emphasise MR-through loss and therefore the effectiveness of using higher quantities of low-bandwidth optical links. The same applies to advanced splitter technologies. If higher loss values for waveguide and splitter loss were used, Lego would likely perform much better compared to Meteor in terms of laser power.

Note that dynamic power in Figure 5.12 represents the power consumed at the saturation point of Meteor, which is at a fairly low injection rate. Leakage power consumes a considerable amount of the total power for this (low) network load. Therefore, the absolute dynamic power in all NoCs is low; however, dynamic power will get more important for higher injection rates, particularly for the 2D Mesh (see Figure 5.11). For instance, the 2D Mesh consumes 0.4 W dynamic close to saturation, whereas Lego just consumes 0.19 W. The 2D Mesh could provide more competitive throughput compared to Lego by extending link widths and paths through the routers to e.g. 128 bits, but the rate at which dynamic power increases in the 2D Mesh would make it consume large amounts of power at higher injection rates. Moreover, more leakage power would be consumed by the additional circuitry. Therefore, Lego would significantly outperform the 2D Mesh for applications that have high utilisation rates and require large network bandwidth. The same holds true when comparing Lego to Meteor.

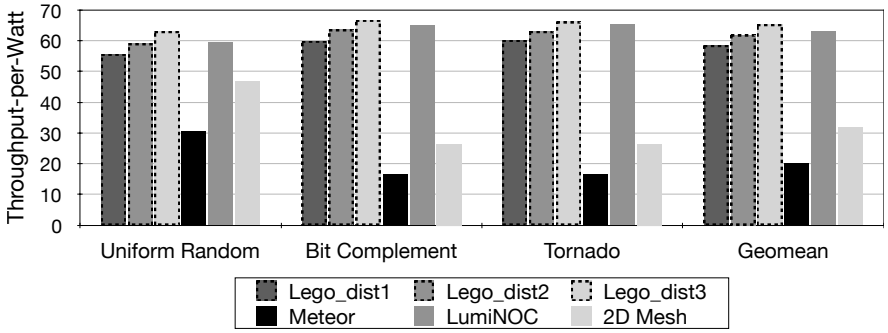


Figure 5.13: Throughput per Watt

In addition, the off-chip laser is not part of the on-chip power budget, so for applications that stress the on-chip network, Lego is actually the preferred design as it has less of an impact on the thermal design power. Besides, as revealed in the previous chapter, the technological assumptions with regard to device losses are pessimistic and advanced devices could further reduce laser power.

Figure 5.13 illustrates the TPW (highest injection rate before network saturation divided by the consumed power at this point). All Lego implementations more than double TPW compared to Meteor and 2D Mesh because Lego saturates significantly later and becomes increasingly power-efficient with growing injection rates. If the networks are utilised at high rates, Lego is the preferred choice. LumiNOC offers TPW values similar to Lego; however, it also doubles packet latency which makes it a less attractive candidate. In summary, laser and MR heating power overheads of Lego make it less efficient for very low injection rates. For high network utilisation, Lego improves power efficiency significantly (up to 3.25$\times$ and 2$\times$ higher TPW (on average) compared to the 2D Mesh and Meteor, respectively) because it can sustain much higher network loads and consumes only very little dynamic power due to its efficient combination of electrical and optical links. With advanced SiP technologies and improved MR heating techniques (ideally athermal MRs), static optical power will become a smaller part of the power budget, and topologies like Lego that feature very low dynamic power will become an even more attractive option, particularly for applications that stress the on-chip network.

Realistic Workloads

Figure 5.14 presents the dynamic power results for the SPLASH-2/PARSEC workloads. We make two main observations: firstly, the savings in dynamic power of Lego

are also present in realistic workloads: Lego reduces dynamic power compared to Meteor and 2D Mesh by more than 50%. Similarly, LumiNOC offers the lowest dynamic power as it is all optical; however, as discussed before, its latency overheads make LumiNOC less competitive. Secondly, the total dynamic power is in the range of 1 mW for all NoCs, which is at least two orders of magnitude lower than the static power consumption in these NoCs.

When analysing the injection patterns of the cores into the NoC across these workloads we noticed that the average injection rate over the entire course of program execution is significantly lower than the saturation points of the NoCs. In particular, the average packet injection rate is between 1-2 packets per 1000 cycles per core, which has also been observed by other studies [LNP+13]. Both benchmark suites provide applications from the high-performance domain which are computation-intensive rather than communication-intensive. Although these applications do exhibit phases of high network utilisation, the average network utilisation is low. Applications from other domains (such as from servers or clouds) might exhibit much higher average injection rates, and evaluating these NoCs with such workloads may shift the impact of static optical power and dynamic power to make Lego the more favourable design.

Next to the communication profile of these applications, some studies suggest that the network utilisation rate can never be very high (on average) in CMPs because of the *self-throttling* effect [MM09]: processors will stall sooner or later (and thus stop injecting packets into the NoC) if their internal buffers that keep track of outstanding cache line requests (e.g. MSHRs [MM09]) are full. The internal buffer size thus limits the average injection rate, even for memory-intensive workloads in which processors may require to issue further memory accesses [LNP+13].

From a power perspective, ONoCs, at the current state of technology, are less suitable for these applications as dynamic power plays an insignificant role compared to static power, which is much higher for optical links. Other applications and CMP architectures may make NoCs like Lego more favourable due to its low dynamic power consumption and competitive latency.

5.4.4 Area

Figure 5.15 illustrates the area breakdowns of the considered NoCs normalised to the electrical mesh baseline. For the area estimations, we utilised the 22 nm technology library in DSENT for the electronic components, and 5 μm waveguide pitch and 10 μm^2 MR area. All NoCs (apart from LumiNOC) are based on a 2D electrical Mesh

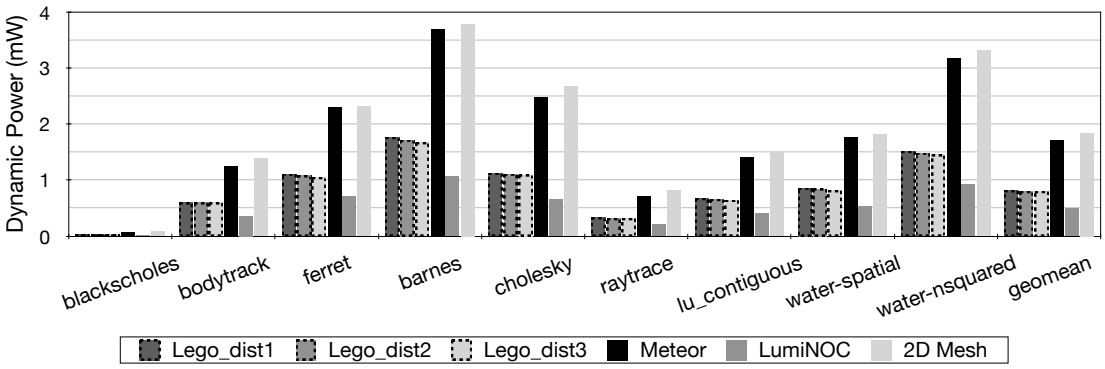


Figure 5.14: Dynamic Power Consumption for SPLASH-2/PARSEC Workloads

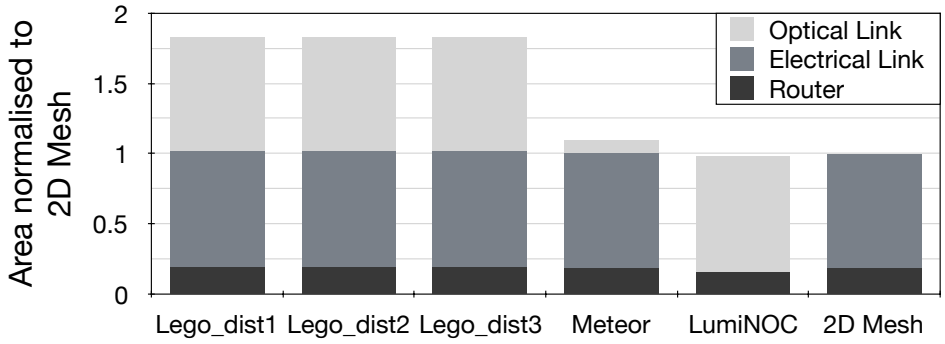


Figure 5.15: Area Breakdowns

and therefore require the same electrical link area. Router area is slightly higher in Meteor and Lego since the number of input ports is higher which causes higher area of the crossbar within the router (in Meteor this is only the case in the hub routers); however, since we apportion the same buffer space to each router in each NoC and the buffers dominate the router area, the overheads are small. The area requirements for all Lego implementations are almost the same although Lego implementations with a larger distance parameter take more load off the optical links (as discussed earlier). This is because the U-shaped waveguides in Lego require much more area than the MRs, which renders the MR savings insignificant in terms of area. In summary, Lego does impose area overheads, however, not to an excessive extent (1.8$\times$ compared to the 2D Mesh baseline). It has been suggested that area overheads are acceptable if it enables higher power efficiency [JBK$^+$09], which Lego provides.

5.5 Summary

This chapter proposed a novel approach of combining electrical and optical link in a NoC topology to utilise both interconnect technologies where they are most efficient to strike both low latency and low power. Analysing the energy and latency properties of both interconnect technologies in detail allows to identify the on-chip distances at which it is best to use either electrical or optical links. In addition, laser power consumption on optical links grows significantly as link bandwidth is increased, and we advocate utilising optical links with lower bandwidth to avoid these overheads. To mitigate the serialisation overheads of low-bandwidth links, they are used for large distances only and supplemented with electrical links short distances. The novel NoC design 'Lego' adopts this approach in a topology and shows that performance levels of aggressive electrical baselines can be maintained at significantly reduced dynamic power. Although Lego implements a much higher quantity of optical links than other hybrid topologies (which rely on low numbers of high-bandwidth optical links), its laser power overheads do not grow to the same extent. In summary, Lego showcases that a fine-grained, distance-based topology can offer large dynamic power reductions at low latency and high throughput.

This study showed that laser and MR heating power dominate the power budget for low network utilisation rates – common in the high-performance domain – which makes NoCs with optical links inferior to their electrical counterparts. As injection rates grow, Lego becomes increasingly efficient and is a superior design for NoCs that must support high network utilisation at low dynamic power. The results suggest that in order for Lego to be the preferred choice for NoCs with low utilisation rates, advances in SiPs must be made; however, this study evaluated Lego with conservative device values, and more advanced devices that have already been demonstrated would make the approach of using higher quantities of low-bandwidth optical links more favourable. This is particularly the case for lower splitting and waveguide propagation losses, which have already been demonstrated. Lower MR-through loss values would reduce the efficiency of the proposed approach; however, aggressive values assumed in recent studies are speculative and have not been demonstrated yet. In either case, the very low dynamic power results of SPLASH-2/PARSEC workloads suggest that further technological improvements are required before optical links will emerge in these application domains.

Chapter 6

Efficient Bandwidth Sharing on Optical Buses

6.1 Introduction

As discussed in the previous chapters, static power consumption in ONoCs is directly related to network bandwidth, and much of it is wasted in many realistic applications due to low average injection rates. Consequently, designs that can efficiently utilise the available bandwidth are key to power efficiency.

Several recent studies investigated bandwidth sharing mechanisms in which multiple sender-receiver pairs share the available optical bandwidth through TDM with promising results (see Section 3.4). While many proposals investigate bandwidth sharing mechanisms on the NoC level, shared optical buses represent a particularly interesting design approach due to their modularity and practicality: just like in traditional electrical buses, multiple nodes are connected to a common bus on which an arbitration mechanism manages bus access. If implemented in a NoC, bandwidth and bus size can be scaled flexibly according to the NoC's performance demands and power constraints. In fact, many designs utilise buses as the backbone of higher-order topologies (e.g. [JBK+09][PKK+09]). Therefore, any improvements regarding bus utilisation are beneficial as they carry over to enhance the efficiency of the NoC as a whole.

Although some critical design points of shared optical buses have been analysed for different loss values [LBGP14], the scientific literature lacks a thorough investigation of latency, bandwidth, and power consumption for different bus sizes, bus bandwidths, and critical SiP device parameters. In addition, bus arbitration and scheduling mechanisms that are more sophisticated than the ones proposed could improve the throughput

on these buses. Finally, the question arises whether bus arbitration should be performed on the same bus as data transmission ('in-band'), or whether performing arbitration in parallel on a dedicated bus is more power-efficient overall.
This chapter explores all of these research questions and makes the novel following contributions:

- A detailed analysis of the shared optical bus studying IL_{max}, laser and MR heating power for different bus sizes and bandwidth. MR-through loss is identified as the critical loss factor and limits bandwidth on a bus to 32λ for current SiP technologies. Device predictions forecasting reductions in MR-through, however, would result in more than 50% laser power reductions.

- Splitting the optical bandwidth into subchannels, enabled by tuning MRs in groups, to allow parallel transmissions over the same waveguide and, in turn, to reduce the latency overheads of sequential scheduling.

- A low-overhead scheduling algorithm that allocates transmissions to resources in both time slots and subchannels. The proposed algorithm is adaptable to any bus bandwidth and size, number of subchannels, and flow control mechanism.

- Both a centralised and distributed arbitration mechanism to illustrate that subchannel scheduling can be implemented efficiently in shared optical buses.

- A study evaluating and comparing parallel arbitration to in-band arbitration, including the design of a parallel arbitration bus to support subchannel scheduling and an efficient mechanism to parallelise arbitration that aims to maximise throughput.

- An evaluation of all proposed shared buses when implemented as the backbone into a realistic NoC topology for 64 and 256 nodes.

Compared to the state-of-the-art sequentially-arbitrated bus LumiNOC [LBGP14], subchannel scheduling improves throughput up to $2\times$ with both arbitration schemes (centralised and distributed) for in-band arbitration. Although exhibiting higher arbitration complexity, these approaches do not incur any power overheads; however, for very low injection rates, latency is increased by 10-20% (depending on bus bandwidth).
Performing bus arbitration in parallel on a separate bus makes subchannel scheduling even more beneficial to LumiNOC since arbitration latency overheads can partially be hidden. Compared to in-band arbitration, throughput per Watt is doubled because

power overheads of the additional arbitration bus are less than 10% and throughput improved by more than 2×. In addition, parallel bus arbitration offers higher design flexibility as bandwidth on the arbitration and data bus can be scaled independently, suggesting that parallel bus arbitration is overall a more power-efficient design point. When implementing buses as the backbone of realistic NoCs all these benefits carry over, and the throughput gains can be leveraged to either provide high-bandwidth NoCs, or low-power NoCs by utilising clustering.

6.2 The Shared Optical Bus

Li et al. [LBGP14] proposed the shared optical bus (see Figure 6.1) in which *N* nodes are connected to *one* waveguide on which they share the available optical bandwidth ($\lambda_0..\lambda_{63}$). As discussed in Section 2.3, this reduces laser power compared to other bus-based crossbars significantly as it decreases the total number of wavelengths which can now be scaled independently from the number of nodes. Furthermore, it just requires one waveguide and allows for more efficient bandwidth utilisation.
Enabled by a U-shaped waveguide, each node modulates on the transmit side (tx path, in red) and receives on the receive side (rx path, in green). In contrast to a contention-free crossbar, this approach requires TDM to avoid data corruption of simultaneously transmitting nodes. Therefore, shared optical buses work in two phases: 1) an arbitration phase in which nodes request the bus and are granted access by an arbitration mechanism and 2) a data transmission phase in which nodes transmit their data. If multiple senders request the bus simultaneously, bus arbitration is required and contending nodes must be scheduled sequentially for data transmission.
Sharing wavelengths through TDM on the same waveguide is enabled by MR tuning, which allows to dynamically switch on/off modulators and MR filters by shifting their resonance wavelengths (as discussed in Section 2.3.2). Figure 6.2 exemplifies a use case in which Node 0 and Node 15 own the bus for communication, thus tuning in their modulators and MR filters, respectively. All other nodes detune their modulators and MR filters to prevent interfering with the data transmission. If another sender-receiver pair is scheduled subsequently on the bus, all nodes tune/detune their MRs accordingly to enable correct TDM utilisation of the bus.

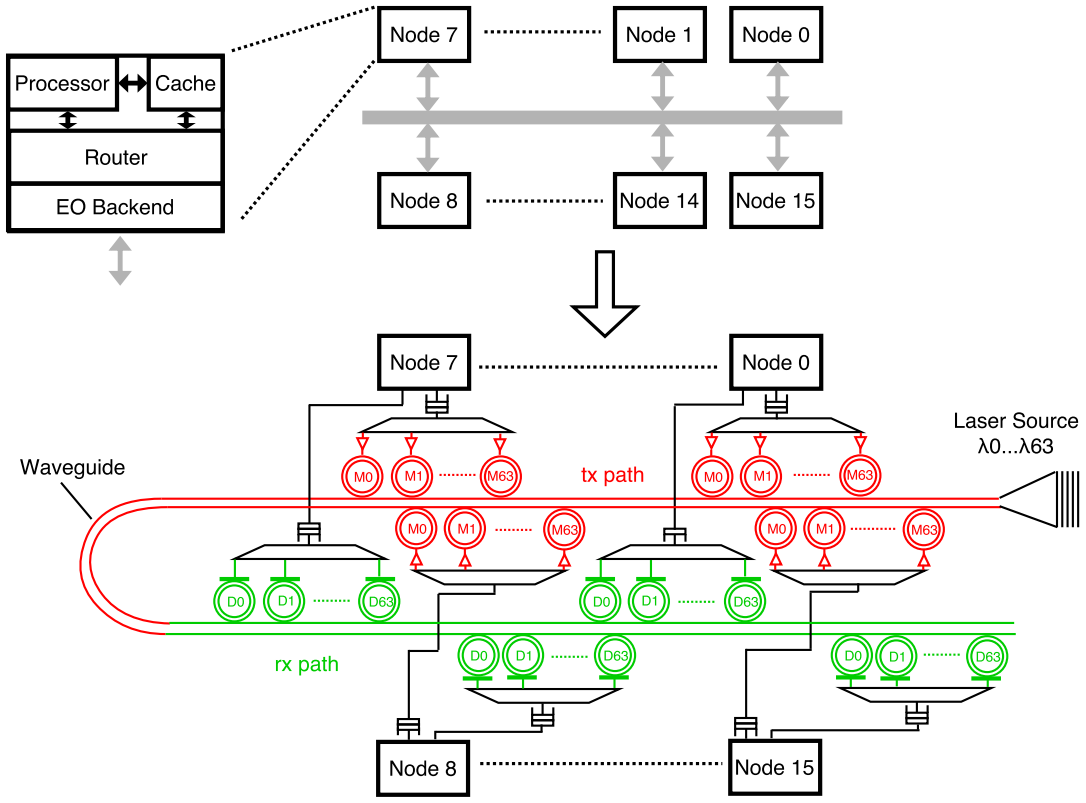


Figure 6.1: Shared Optical On-chip Bus: Layout

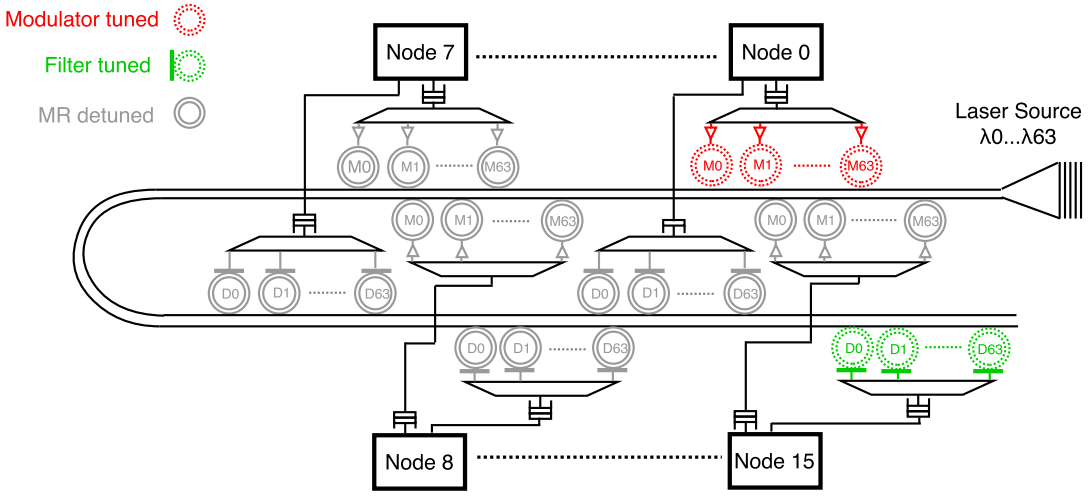


Figure 6.2: Data Transmission Example on a Shared Optical Bus

6.2.1 Challenges

Next to all the benefits of shared optical buses, their implementation requires to address a number of design challenges to assess its suitability and power efficiency for current

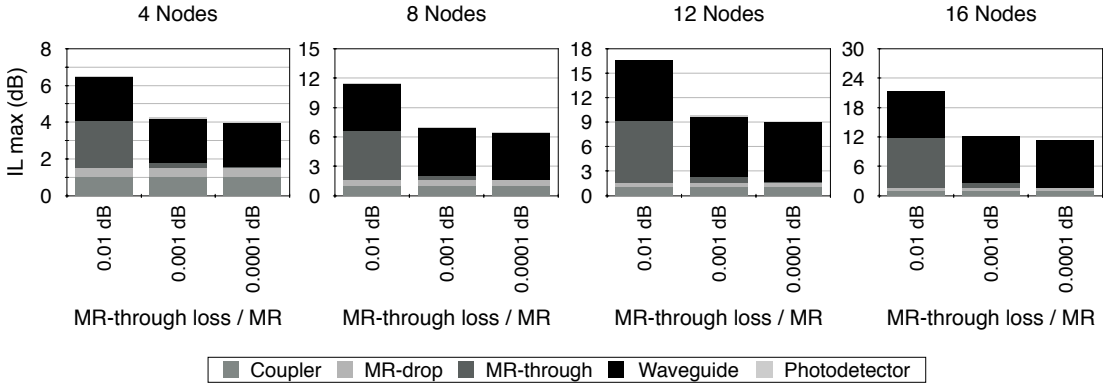
SiP device technologies. Optical bandwidth on the bus must be apportioned sensibly and put into relation with IL_{max} and laser power. Since SiPs are a quickly evolving technology, analysing the critical device losses and their impact on laser power for both conservative and speculative loss values is of high interest to assess the potential benefits and drawbacks of shared buses. In addition, the power consumed for MR heating must be considered carefully as every node on the bus requires one modulator and one filter for each wavelength. This section identifies critical design parameters in shared optical buses and analyses the power requirements for successfully demonstrated technologies as well as future device forecasts to identify the current state and future potentials.

6.2.2 Insertion Loss

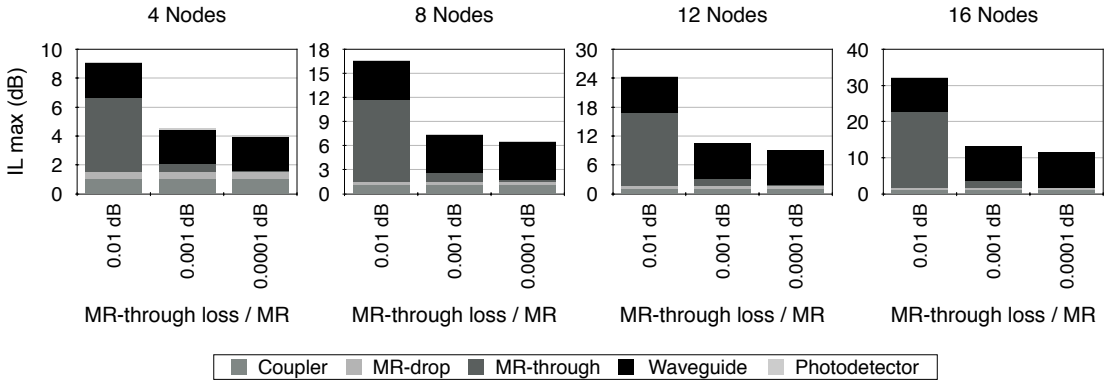
This study analyses the critical device losses and their dependency on the number of wavelengths (i.e. optical bandwidth) on shared optical buses and their impact on laser power. The IL_{max} path on a shared optical bus is the path to the receiver that is the furthest away from the laser source. For instance, in Figure 6.1, that would be the path to Node 15. While all SiP devices contribute to IL_{max}, MR-through loss is the most critical contributor in shared buses. Although low in absolute value relative to other device losses, the number of MRs a signal passes on a shared bus is high as each node connects both modulators and filters on the whole range of wavelengths on the bus. Future device speculations suggest very low MR-through loss values per MR (0.001 dB/0.0001 dB) [JBK$^+$09, LBGP14]; however, the most recent demonstrated prototypes exceed these values by 10$\times$ (0.01dB) [GMS$^+$14]. Although reporting the impact that higher MR-through losses per MR can have, previous studies base their laser power results mainly on optimistic loss values [LBGP14]. Evaluating IL_{max} for both conservative and speculative parameters would, therefore, cast light on both current and potential future power requirements.

Shared buses with optical bandwidth of 32λ and 64λ are commonly considered in literature (e.g. [LBGP14] [BP14] [PKK$^+$09]) and will, therefore, be analysed in the following. Figures 6.3a and 6.3b depict IL_{max} for varying MR-through loss values and bus sizes. Other than for MR-through losses, we assume the technology parameters listed in Table 5.1 (see the previous Chapter).

MR-through loss is a major contributor to IL_{max} for demonstrated MR devices with loss values of 0.01 dB, particularly as the number of wavelengths and nodes increases.



(a) 32λ Bus Bandwidth



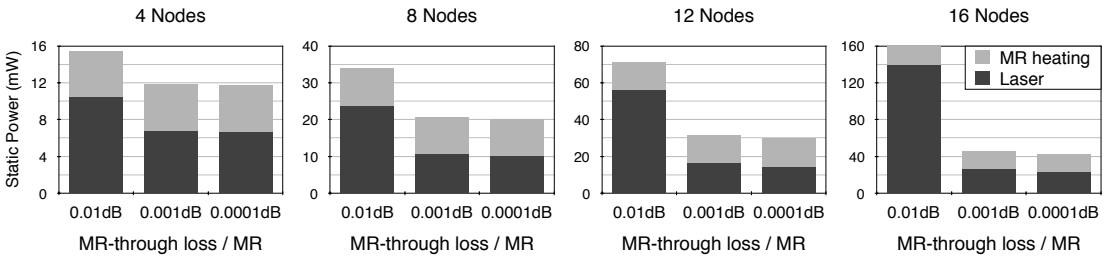
(b) 64λ Bus Bandwidth

Figure 6.3: Insertion Loss Analysis of a Shared Optical Bus

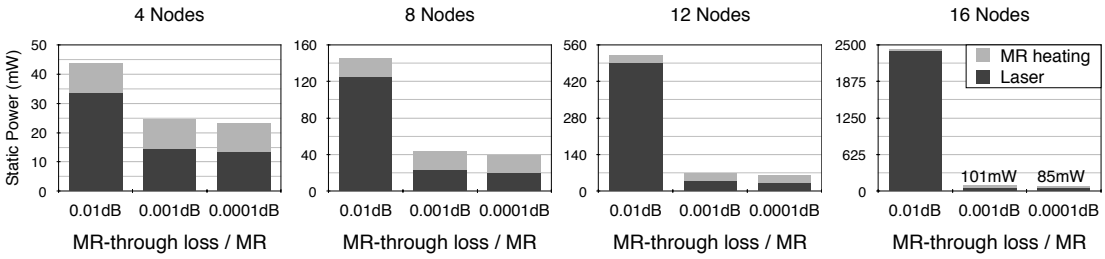
For both 32λ and 64λ, device improvements to 0.001 dB MR-through loss would decrease the overall IL_{max} significantly compared to 0.01 dB (for all bus sizes). These benefits increase along with the number of wavelengths on the bus and the number of nodes since both of these factors determine the number of MRs that have to be passed. Once these device predictions are in place, waveguide loss would dominate IL_{max} for losses of 0.3 dB/mm; however, waveguides with 0.0271 dB/mm loss have already been demonstrated [BS11], which would decrease waveguide loss by more than 10×, and thereby reduce the total impact of waveguide loss in the shared bus, too. Nevertheless, the potentially achieved reductions of advanced MR-through losses are very promising. Improvements from 0.001 dB to 0.0001 dB MR-through loss has only a small impact on shared buses of the considered bandwidth values and sizes.

6.2.3 Power Consumption

To determine the power requirements of a shared optical bus, the key questions are how changes in IL_{max} for different MR-through loss values translate to laser power, and what the ratio between laser and MR heating power on shared buses is. Figures 6.4a and 6.4b show the power breakdown for the considered bus sizes and wavelengths.
One can observe the same trends for laser power as for IL_{max} for different MR-through loss values: laser power savings of more than 50% could be achieved by more advanced technologies (for buses with more than 8 nodes). Similarly, improvements from 0.001 dB to 0.0001 dB MR-through loss are small. MR heating contributes less to the total power than the laser, although the number of MRs on the buses is high (each node needs λ modulators and MR filters); however, it will gain more significance as MR-through losses decrease. The previous chapter revealed that the relationship be-



(a) 32λ Bus Bandwidth



(b) 64λ Bus Bandwidth

Figure 6.4: Static Optical Power Requirements of a Shared Bus

tween the number of wavelengths and laser power is exponential on R-SWMR buses, mainly due to MR-through losses. Shared optical buses further exacerbate this laser power problem because each node places modulators next to the waveguide (vs. only one node on a R-SWMR bus), which leads to almost twice the number of MR passings. For instance, for 8 nodes and 0.01 dB MR-through loss, laser power more than triples as bandwidth is increased from 32λ to 64λ. For the lower MR-through loss cases, doubling the bandwidth has a more linear relationship to laser power, although

not perfectly linear due to MR-through losses and higher crosstalk. This underlines the impact that MR-through loss has on laser power on shared buses, and the importance of advanced SiP devices with lower MR-through losses.
Increasing the bandwidth on shared optical buses, particularly for the current device technologies with 0.01 dB MR-through losses, is achieved with much less laser power by implementing two buses with 32λ rather than one bus with 64λ. Although physically implemented as two separate buses, logically all nodes would transmit data on the bus as if it was one. From a layout perspective, this approach would most likely not pose any major obstacles because shared buses are already very layout-friendly (only one waveguide). Light can be provided to two buses by either coupling one more laser source into the chip or by utilising optical splitters to distribute the optical signals. Although splitters introduce further losses, the previous chapter identified that these overheads are relatively small compared to the MR-through losses incurred by providing the same bandwidth on just one waveguide.

6.2.4 Discussion

This study confirms the significance of MR-through losses to laser power in shared optical buses, particularly as the number of nodes and wavelengths on the bus increases. While the outlook of laser power reductions with more advanced devices and lower MR-through loss is very promising, no clear roadmap for SiPs exists, and it is difficult to estimate when exactly these technological advancements will be in place. The following studies of this chapter will, therefore, stick to 0.01 dB MR-through loss to provide a more realistic assessment of the current state of SiPs.
For 0.01 dB MR-through loss, the power overheads of implementing 64λ buses with *one* waveguide are large, and even become unsustainable for buses with more than 8 nodes. Therefore, the rest of this chapter assumes 32λ on one waveguide, and increasing the bandwidth on the optical buses is achieved by adding 32λ buses. For instance, a (logical) 128λ bus would consist of four physical 32λ buses.

6.3 Subchannel Scheduling

State-of-the-art shared optical bus proposals schedule nodes that simultaneously request the bus sequentially on the entire optical bandwidth. Communicating nodes thereby tune in their MRs for the duration of their assigned time slot, and detune

them otherwise [LBGP14]. Rather than performing strict sequential scheduling, large throughput improvements could be achieved by allowing multiple sender-receiver pairs to utilise the bus both sequentially *and* in parallel by logically dividing the available optical bandwidth into *subchannels*. Leveraging subchannels on the same physical channel is a concept well-known in electrical interconnects (i.a. [DNSD13, VSG+12]) and could also be adopted in shared optical buses since MRs can be tuned/detuned individually by their integrated heaters. For shared optical buses, this has not been studied by the scientific community yet, but could have great potential to improve throughput and power efficiency.
This section will first explain the idea of logically splitting a bus' bandwidth into subchannels, followed by introducing a low-overhead scheduling algorithm that assigns simultaneously requesting nodes to both time slots and subchannels. Subsequently, it presents two different arbitration mechanisms – one centralised and one distributed – that enable subchannel scheduling on a bus, followed by a simulation study comparing the proposed approaches to the state-of-the-art timeslot-only mechanism.

6.3.1 Efficient, Light-weight Subchannel Scheduling

The fact that MRs are typically tuned/detuned individually or in groups by either integrated or co-located heaters [GLM+11] allows to schedule requesting nodes both sequentially and in parallel on different λ-subsets – *subchannels*. Extending sequential with parallel scheduling adds to the complexity of computing a correct and efficient scheduling of requesting nodes, which may result in latency, power, and resource overheads. To tackle this issue, this section presents a subchannel scheduling approach that attains high bus utilisation while allowing for a simple scheduling mechanism that should allow to compute time slots and subchannels fast and with low resource requirements. Besides, we briefly review the components that add to the latency in optical data transmission on shared optical buses, and analytically show the superiority of subchannel to timeslot-only scheduling.

Bus Splitting into Subchannels

Subchannels are formed by splitting the optical bandwidth available on the bus *logically* into non-overlapping subsets. This allows multiple sender-receiver pairs to communicate simultaneously on the same bus by utilising different subchannels. For instance, in Figure 6.5, Node 0 sends to Node 8 on Subchannel 0 and 1, while Node 7

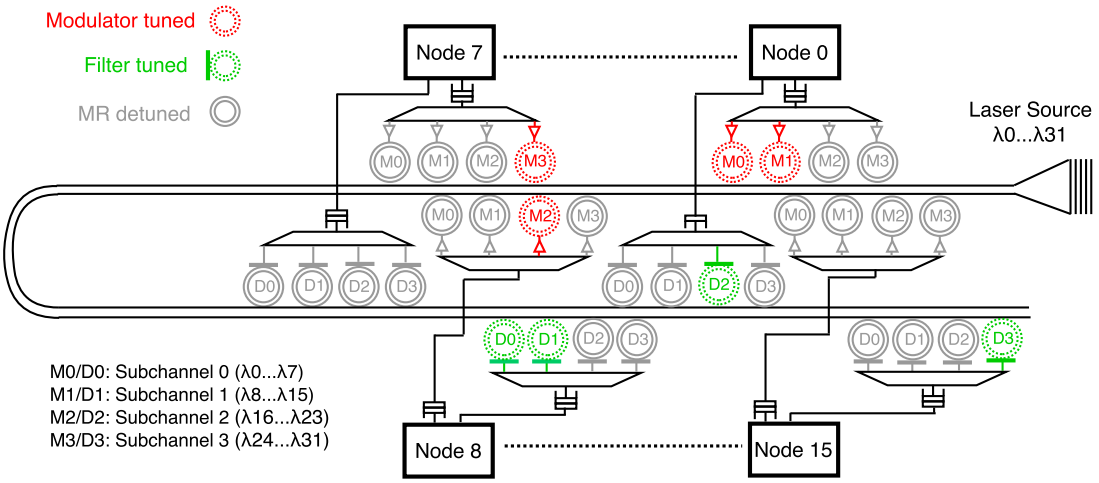


Figure 6.5: A Shared Optical Bus during Data Transmission with Subchannels

sends to Node 15 on Subchannel 3, and Node 8 to Node 0 on Subchannel 2. This is enabled by each node tuning in sets of modulators (red) and filters (green) according to their assigned subchannel(s), while detuning all other MRs (grey). In this example, this optical bandwidth of 32λ would be divided into four subchannels with 8λ each.

Minimising Bus Utilisation Cycles

As discussed in previous chapters, optical data transmission includes 1) EO data conversion through modulation, 2) signal propagation delay on the waveguide, and 3) OE data conversion through detection, where 1) depends on core and link data rate, 2) on link length and propagation delay, and 3) on detector and OE backend speed (typically one core clock cycle). Chapter 4 discussed tuning speeds of MRs – i.e. the time it takes to shift and stabilise their resonance wavelength – which have been subject to extensive research [SBC08, PTDS16, TLY+16]. This study assumes a tuning delay of one core clock cycle at 5 GHz like previous studies [Van10]; however, as we will see in the following, longer tuning delays would further increase the efficiency of subchannel scheduling compared to sequential scheduling as in total fewer consecutive tuning cycles are required compared to sequential scheduling without subchannels.
These latencies are illustrated in Figure 6.6, which depicts how bus time slots are utilised for transmitting several 64-bit packets. The leftmost example demonstrates sequential data transmission over *one* subchannel comprising the entire bandwidth. Each packet (P0, P1, etc.) occupies the bus for 4 cycles (including MR tuning delay between the packets). Note that, for a 64-bit packet, only 32λ are needed for modulating it in one clock cycle, effectively wasting half of the bandwidth in this case – one of the

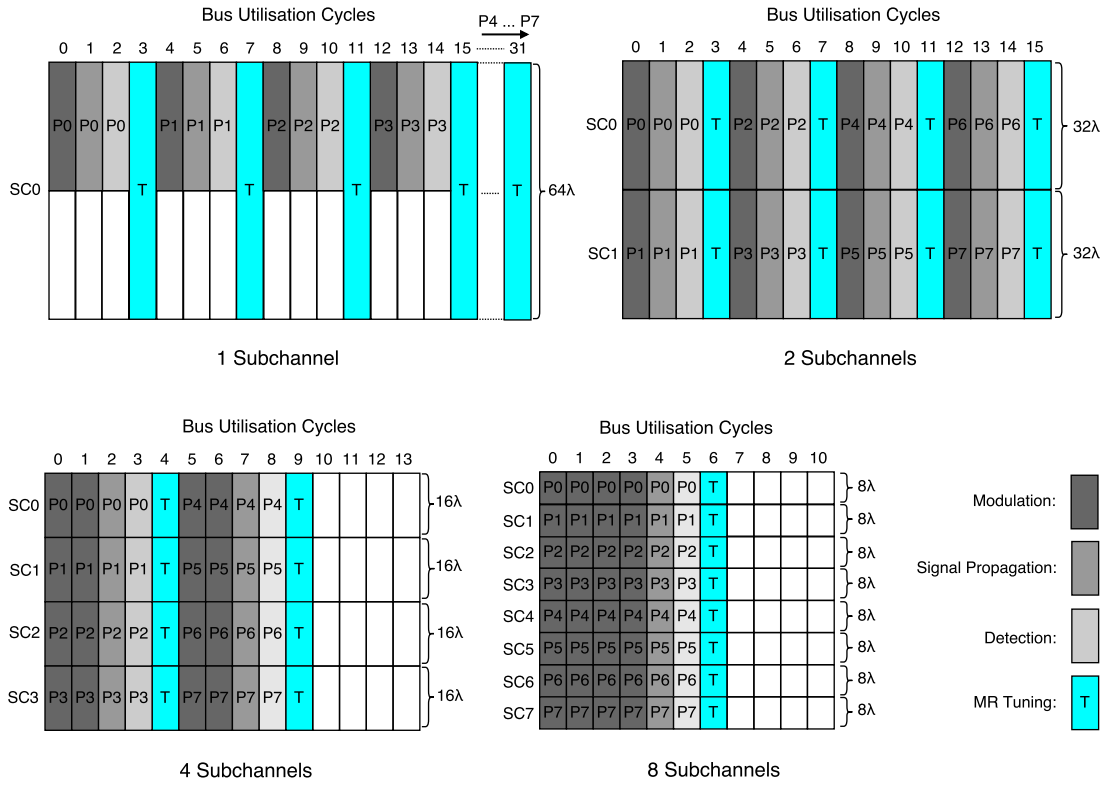


Figure 6.6: Data transmission phase on a 64λ bus for 64-bit data packets with varying numbers of subchannels (SC).

weaknesses of strict sequential scheduling. As we increase the number of subchannels to 2, 4, and 8 we notice a positive effect on the total utilisation cycles. Halving the bandwidth of each subchannel from e.g. 32λ (2 subchannels) to 16λ (4 subchannels) leads to twice the modulation latency, so the overall modulation time in parallel and sequential scheduling is the same; however, propagation, detection, and MR tuning overheads are parallelised, which leads to large latency savings overall. For instance, in the 8 subchannel case, propagation/detection/tuning delay of each packet is only impacting the total latency *once*, while e.g. in the 1 subchannel case, this latency is added up for every single packet scheduled on the bus. The ideal case, from an analytic perspective, is to provide one subchannel for each node connected to the bus (i.e. maximum number of simultaneous requests), with a total bus bandwidth that is divisible by N (to avoid uneven subchannel widths).

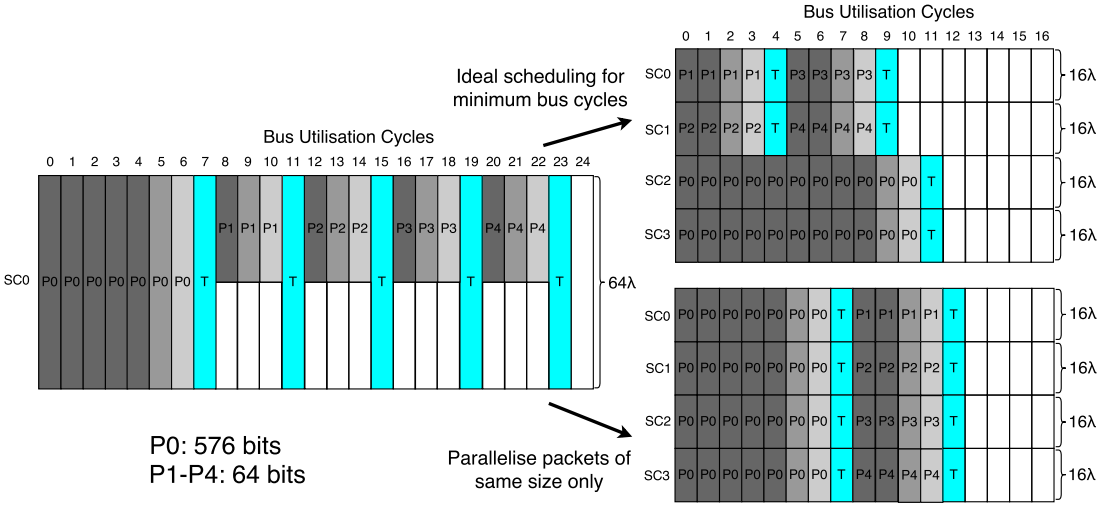


Figure 6.7: Scheduling Multiple Packet Sizes on Subchannels

Subchannel Scheduling Algorithm

Since scheduling computation is performed in the arbitration phase, it should ideally take only one cycle to determine an efficient scheduling of the incoming requests. While computing the minimal latency for nodes requesting the bus for the same packet size is simple, this is not the case if the bus is requested for multiple packet sizes, e.g. 64 bits and 576 bits, which makes the task of finding the ideal scheduling with minimum bus utilisation cycles more complex and in turn possibly causes significant latency overheads.

Figure 6.7 shows a simple example of this: assume the bus is requested for one 576-bit packet (P0) and four 64-bit packets (P1-P4) and that four subchannels are available (on the right). On the left, sequential scheduling takes 24 cycles in total. The top right figure depicts the ideal scheduling of these packets which schedules packets of different sizes in parallel and minimises the latency down to 12 cycles. This requires to determine all possible combinations of these packets on all possible time slots and subchannels, leading to a computation complexity that has factorial growth with the number of requests.

Grouping requests according to the packet sizes they are requesting the bus for and parallelising only packets of the same size tremendously simplifies the scheduling algorithm and only leads to small bus cycle overheads that are likely neutralised by the latency saved during scheduling computation in the arbitration process. The bottom right figure in Figure 6.7 illustrates this scheduling, which, in this example, sends the 576-bit packet on the entire bandwidth and parallelises the 64-bit packets subsequently,

resulting in just one additional cycle. While this is a simple example for illustration purposes, this trend holds true for higher quantities of requests and combinations of packet sizes.

Algorithm 1 shows the algorithmic definition of the allocation circuitry. This extends to multiple packet sizes by having a separate queue for each and applying the allocation separately. At this point, we make no assumption on which packet sizes should go first, and leave this for future analysis as our focus is on maximising bus utilisation. Our greedy algorithm attempts to schedule as many packets in parallel as possible. If there are more requests than subchannels, each subchannel is assigned to a different packet, and the remaining packets will be scheduled in the next time slot. The pointer to the starting point of the time slot in Algorithm 1 is the 'slot_start_cycle' variable, which is determined by the time occupied by the current (and all previous) time slots ('current_slot_duration') based on the packet size and number of wavelengths, subchannels, and requests. For instance, in the '2 Subchannels' example in Figure 6.6, the first slot would start at cycle 0, the second at 4, the third at 8, etc. Prior to time slot and subchannel allocation, the incoming requests are stored in a queue ('sorted_reqs') based on their priorities/credits. This ensures that the flow control mechanism is obeyed by ensuring that priorities are maintained. If there are fewer requests than subchannels, i.e. not all subchannels can be filled with requests, the optical bandwidth must be assigned evenly to minimise bus latency. The number of subchannels assigned to each requester ('#SC') is thus the total number of subchannels divided by the number of requests. For instance, if only one request remains it will use all subchannels, if there are $< \#subchannels/2$ requests each will use two subchannels, $< \#subchannels/4$ requests each will use four subchannels, etc. Each allocation is executed by the function *assign(request, subchannel-range, starting slot)*.

6.3.2 Bus Arbitration Mechanisms

The arbitration phase is the default state of the bus and is entered by each node once data transmission is over and the bus is free again. Arbitration can be centralised or distributed, with both approaches entailing different opportunities and trade-offs. In both cases, the exchange of arbitration packets manages bus access. After the arbitration phase, each node must know the time slot and subchannel(s) on which it 1) is allowed to send its data (in case it contended for sending on the bus), 2) has to tune in its MR filters (in case it is a receiver), 3) has to keep its filters detuned, and 4) when the next arbitration round begins.

ALGORITHM 1: Subchannel and Slot Allocation

```
Queue sorted_reqs = sortReqsBasedOnCredits(requests_in);
while (!sorted_reqs.empty()) do
    if (num_reqs >= num_subchannels) then
        num_current_reqs = num_subchannels;
    else
        num_current_reqs = sorted_reqs.size();
    #SC = num_subchannels / num_current_reqs;
    for (int i = 0; i < num_current_reqs; i++) do
        assign(sorted_reqs.pop(), C(i*#SC, #SC*(i+1)-1), slot_start_cycle);
    end
    slot_start_cycle += current_slot_duration;
end
```

Arbitration packets should be small while carrying all information to enable correct scheduling of time slots and subchannels. The previous section showed that sequential scheduling is inefficient in terms of bus utilisation; however, it simplifies arbitration since allocation is unidimensional (time slots only, no subchannels). This section introduces both a centralised and distributed arbitration scheme for subchannel scheduling that exhibit low overheads compared to the sequential approach.

Centralised Arbitration for Subchannel Scheduling

In centralised arbitration, an arbiter computes the scheduling and notifies all nodes about when and how they can access the bus, implemented by exchanging request (REQ) and acknowledgement (ACK) packets between the nodes and the arbiter. The arbiter is connected to the optical bus as illustrated in Figure 6.8. We adopt the approach of LumiNOC, which performs both arbitration and data transmission on the same optical bus, referred to as *in-band* arbitration [LBGP14], thus allowing to reuse optical resources.

Implementing a central arbitration unit often led to a challenging layout, energy overheads, and synchronisation problems when communication was performed electrically due to varying distances between the nodes and the arbiter. With optical links, however, these issues are not as severe due to the signal propagation of light and almost distance-independent energy consumption.

At the beginning of the arbitration phase, nodes request the bus by simultaneously sending a REQ to the arbiter on a unique subset of (bus_width/N)- wavelengths. For instance, in Figure 6.8, a bus of 32λ bandwidth and 16 nodes would provide each node

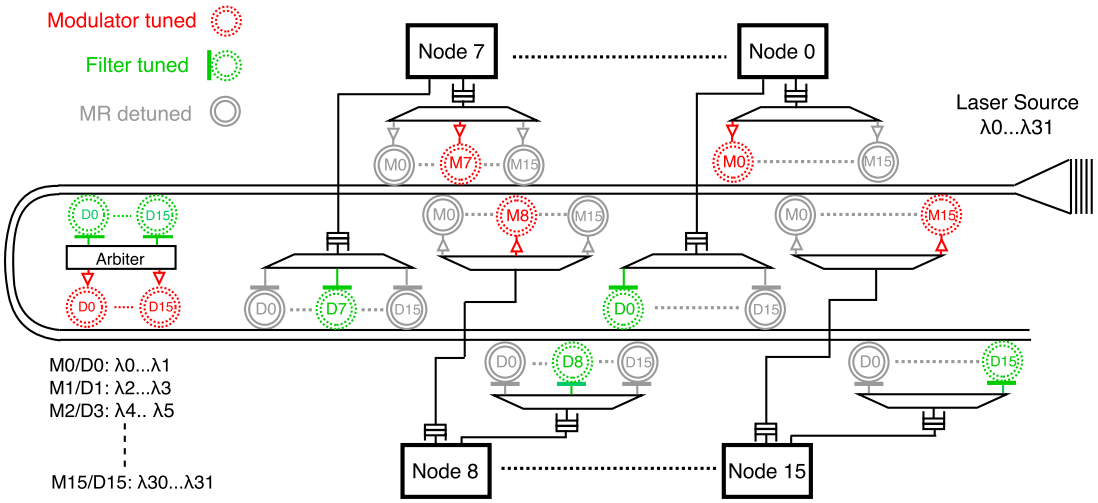


Figure 6.8: Optical Bus During Centralised Arbitration

with 2λ for modulating its REQ (λ_0/λ_1 for Node 0, λ_2/λ_3 for Node 1, etc.). The arbiter has filters to receive REQs from each node tuned to the according λ-subsets. *REQs* sent from the nodes to the arbiter contain fields indicating the packet's destination ID (Dst) and length (Len). The source ID is implicitly known by the arbiter as each node sends on a unique set of λs. For N nodes and S packet sizes, Dst is $log_2(N)$ and Len $log_2(S)$ bits. Once the arbiter received all REQs, it knows all senders, receivers, and packet sizes – all the information needed to compute the subchannel and time slot assignment. This allows for a very compact REQ size that can be modulated quickly with little optical bandwidth.

ACKs are sent from the arbiter to each node on their assigned λ-subsets upon scheduling computation in order to notify senders about the subchannels and time slots on which they are scheduled for transmitting, and receivers about when and to which subchannels they must tune/detune their MR filters. In addition, *all* nodes are informed about when the next arbitration phase begins. If a node is neither a sender nor a receiver in the data transmission phase, it will simply detune its MRs until the next arbitration phase starts.

An ACK contains different fields based on the node's role in the transmission phase:

Senders: ACKs contain 1) a subchannel bitmap with the assigned subchannel(s) set to '1' and 2) the time slot when they have to start sending.

Receivers: ACKs contain 1) a subchannel bitmap, 2) the time slot to tune in, and 3) the packet length used by the receivers to compute the duration of the time slot, i.e. how long their MR filters must stay tuned in.

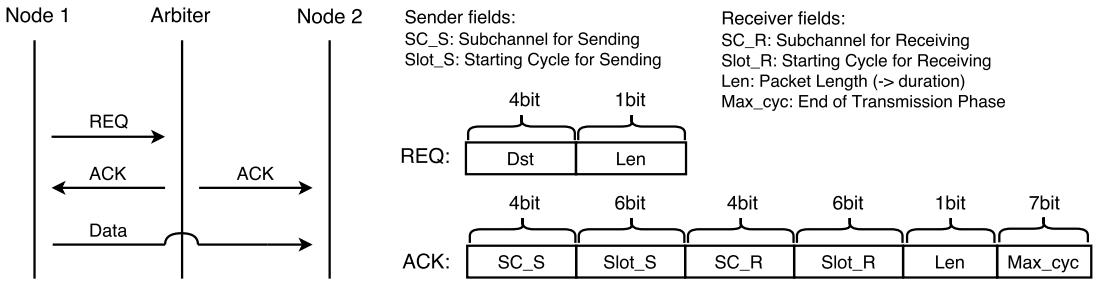


Figure 6.9: Packet exchange in centralised arbitration with all possible REQ/ACK fields for 4 subchannels, 2 packet sizes, 16 nodes, and 32λ bus bandwidth.

Nodes both sending and receiving within the same arbitration round will receive all info appended. Similarly, if a node is receiving *i* packets, *i* receiver fields are appended. The sender fields will always be sent first as each node knows if it is a sender (which allows to interpret the received packet fields). As many receiver fields as needed are appended to the sender fields.

Arbitration Packet Fields Figure 6.9 shows example REQ/ACK packets for a 16-node bus. The shown example ACK encodes a node sending and receiving one packet. While REQs have a constant size, the ACK size depends on whether a node is a receiver, sender, or both in the data transmission phase. With the information contained in the ACKs, every node knows when and on what subchannel(s) it can send, and when and to which subchannel(s) it has to tune in its filters. During the data transmission phases, the arbiter detunes its MRs to avoid filtering the optical signals, and only tunes them in again for the next arbitration phase.

The 'max_cyc' field is appended to indicate when the next arbitration phase starts and will be sent to every node. For instance, for 16 nodes, 4 subchannels, and 32λ bandwidth, this could be up to 78 cycles if each node requests the bus for a 576-bit packet, in which case this field would be 7 bits. In case none of the nodes sends a REQ in an arbitration phase, the arbiter sends an ACK to each node with the 'max_cyc' field set to '0', indicating that the next arbitration phase can start immediately (and will continue doing so if it does not receive a REQ in the following arbitration round).

A separate arbitration unit imposes hardware overheads. A possible arbiter design is depicted in Figure 6.10. Keeping arbitration packets small is not only key to low-latency arbitration, but also reduces buffer area in the arbiter. Moreover, each node is allowed to request the bus for only one packet in each arbitration round to keep the buffer requirements low. Assuming REQ sizes of R bits and N nodes, this would thus

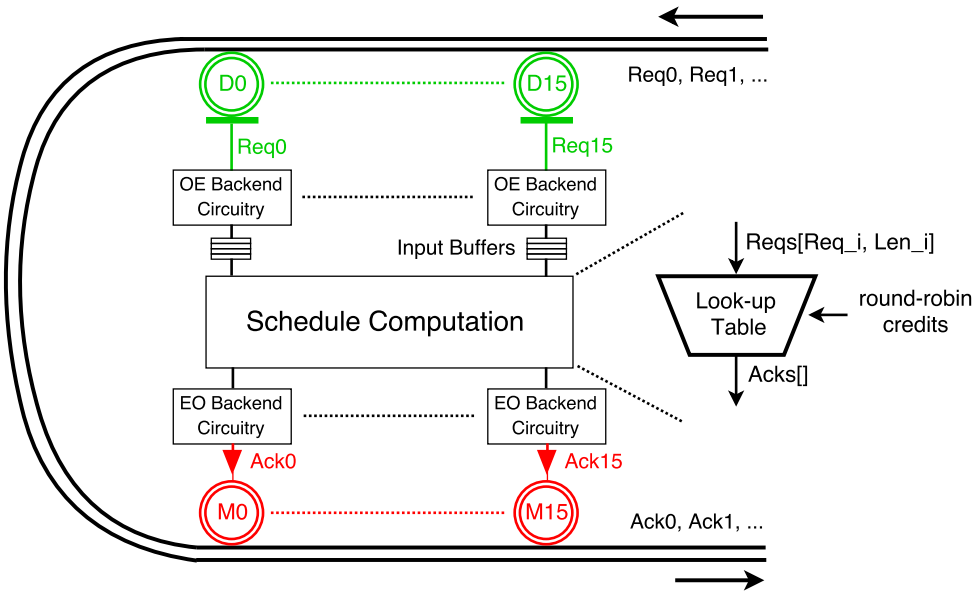


Figure 6.10: Arbiter design. Slots and subchannels are computed based on the incoming requests and assigned based on scheduling/flow control credits.

require (N×R)-bit buffer space, which is negligible for relevant bus sizes (e.g. 8-16 nodes). The scheduling computation unit outputs the ACK packets for each node and must, therefore, provide the same amount of buffers as for the REQs. Receiving and sending data on the optical bus also requires MRs and EO/OE backend circuitry.
Note that, during the arbitration phase, a situation could occur in which two receivers per wavelength are tuned in: the receivers at the arbiter and at the bus nodes. Such a situation would require the laser source to output more power than during the data transmission phase in which only *one* receiver per wavelength is driven at any time (only one receiver on each subchannel). To avoid unnecessary laser power overheads in the arbitration phase, the arbiter will detune its MR filters upon reception of the REQs from the nodes so that only one receiver draws laser power at any time during the arbitration phase, too. This is done in parallel to the scheduling computation and does, therefore, not introduce additional latency overheads.

Distributed Arbitration for Subchannel Scheduling

Distributed arbitration does not require a separate arbiter and in turn saves area and leakage power. Not having a centralised unit, however, makes the implementation of sophisticated arbitration schemes more challenging as more information has to be encoded in the arbitration packets, potentially increasing arbitration latency and energy.

Li et al. [LBGP14] propose LumiNOC with an efficient distributed arbitration mechanism for bus scheduling that leverages the transmission of bitmaps to detect multiple requesters in an arbitration round: each node is synchronised at the beginning of the arbitration phase and transmits a bitmap representing the source addresses on the bus, e.g. for 8 nodes on a bus, the bitmap would consist of 8 bits – one for each node. A node modulates the source bitmap field with its bit set to '1' to request the bus. Each node will receive this arbitration packet and will detect bus contention if more than one bit is set to '1' in the source bitmap field, in which case a dynamic scheduling phase will be entered. Although efficient, LumiNOC's mechanism is merely capable of scheduling nodes in time slots, and not subchannels. Our mechanism provides an efficient extension of LumiNOC for both while aiming to minimise arbitration packet sizes. In particular, it leverages the bitmap creation approach of LumiNOC to propagate information about the senders and subchannels as described in the following.
Like in the centralised approach, each node is assigned to its unique set of λs in the arbitration phase, receives arbitration packets from other nodes on its own λ-set, and modulates its arbitration packets *on each of the other nodes' λ-sets*, broadcasting it on the bus. This is illustrated in Figure 6.11. Arbitration packets are based on 1-hot encoded bitmaps where each bit represents one node. To provide each node with the necessary information to perform scheduling, two arbitration packets are sent subsequently in **two phases**:

1. **Ctrl_1: [Src_Bitmap | Length_Bitmap]**: Each node wishing to send broadcasts Ctrl_1 on the bus with its bit set to '1' in the Src_Bitmap field. At the same position in the Length_Bitmap field, it sets its bit to '1' to indicate a 576-bit packet, and leaves it to '0' for a 64-bit packet. All nodes must be synchronised and must send Ctrl_1 simultaneously so that correctly overlapping bitmaps are created (as described in LumiNOC [LBGP14]).

2. **Ctrl_2: [Src_Bitmap]**: Every node wishing to send modulates a packet containing the same Src_Bitmap field once again right after it sent Ctrl_1, but this time it will only modulate it on the λ-set assigned to its receiver, rather than broadcasting it. This allows receivers to identify their senders because a node A only receives Ctrl_2 from a node B if B wants to send data to A in the following data transmission round. As each receiver already knows the scheduling of the sending nodes from phase 1, it can look at this scheduling to see when it has to tune/detune its MR filters to receive data from the nodes it received Ctrl_2 from.

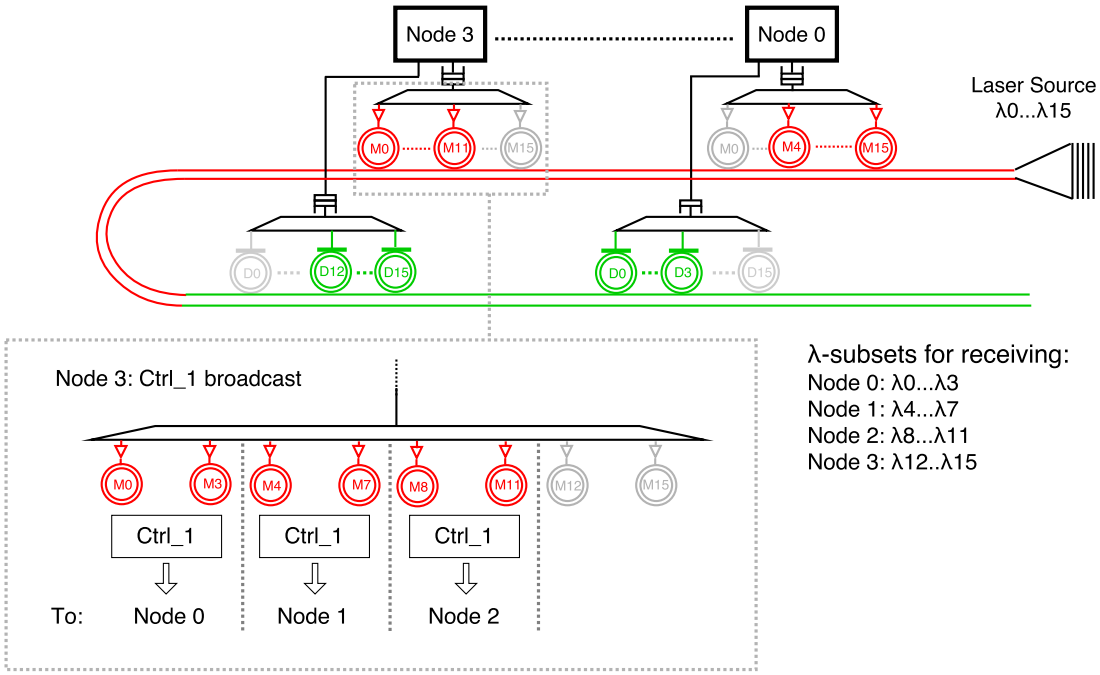


Figure 6.11: Example shared bus (4 nodes, 16λ) during distributed arbitration for subchannel scheduling. Each node is assigned to a distinct λ-subset for receiving arbitration packets. Ctrl_1 is broadcasted on the bus by requesters by modulating a copy of Ctrl_1 on each λ-set (note that, in the figure, the modulators of the last λ-subset are off because this is Node 3's own subset). Ctrl_2 (not shown) is a simple unicast to the future receiver of the packet.

In phase 1, each node receives a bitmap that contains all sending nodes ('Src_Bitmap') and another bitmap to indicate their packet sizes ('Length_Bitmap'), which suffices to compute the scheduling algorithm at each node, allowing each node to know the time slot and subchannels for sending. In addition to that, information is required at each receiving node to know when MR filters have to be tuned/detuned. This information is provided in phase 2.

After phase 1, each node knows the starting slot of each sender, the subchannel(s) it will send on, and the duration from the packet sizes. At this point, however, the receivers are still unknown. This information is obtained in phase 2, where each destination will receive another 'Src_Bitmap'. If one or more bits are set to '1', each receiver can identify its senders as each bit is assigned to one node on the bus. Receivers can now look up the senders' allocated time slot and subchannel(s) in the scheduling results computed after phase 1. If a node does not receive a Ctrl_1 with bits set to '1' it keeps its MRs detuned (as it will not receive packets in this data transmission phase).

This arbitration mechanism should save static power as no centralised arbiter with additional MRs and EO/OE circuitry is required, thereby reducing static optical power requirements. However, broadcasting of arbitration packets and larger packet sizes compared to the centralised arbitration scheme could lead to higher dynamic power which might decrease the benefits of saving static power. Besides, if a NoC had to support more than two packet sizes, this would increase the 'Length_Bitmap' in phase 1, potentially leading to considerable arbitration latencies. In this case, the centralised approach would offer higher efficiency and flexibility.

6.3.3 Evaluation

Methodology

This study compares both distributed and centralised arbitration mechanism for subchannel scheduling to the state-of-the-art time slot only approach LumiNOC [LBGP14] in terms of performance and power consumption.
LumiNOC was shown to significantly outperform a large number of recently proposed ONoCs as well as aggressive electrical baselines. It uses a distributed arbitration approach in which each requesting node broadcasts a bitmap with its respective source address bit set to '1' to request the bus, along with a destination and packet length field to notify the receiver to tune in. Leveraging a bus layout like in Figure 6.5, each requester receives its own arbitration packet and detects, based on the bitmap, whether there are other nodes requesting. If multiple nodes requested the bus (which is detected by multiple bits set to '1' in the source bitmap field), all nodes enter a dynamic scheduling phase in which the requesters – one after another – broadcast an abbreviated arbitration packet followed by the actual data transmission. This abbreviated arbitration packet contains the destination ID and packet length, which is used by the nodes to identify the future receiver and the transmission duration, respectively. To reduce arbitration latency in the case of an uncontested bus, LumiNOC proposes a speculative data transmission approach in which each sender starts sending the data packet right after transmitting its arbitration packet. Upon receiving the arbitration packet, it will either abort data transmission if other nodes want to use the bus too or keep on transmitting in the uncontested case. A comparison to LumiNOC allows to identify both the overheads of more complex arbitration schemes and the benefits of subchannel scheduling.
Both the centralised and distributed arbitration proposals for subchannel scheduling

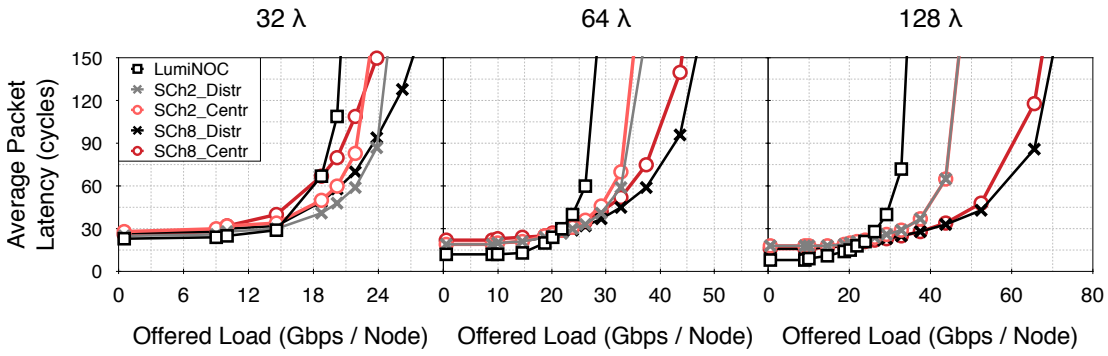
Table 6.1: Experimental Set-up

Modelling tools	HNOCS [BIZCK12] for performance simulation DSENT [SCK+12] for power and area estimations
Frequency and data rates	5 GHz core/router clock, 10 Gb/s modulators/detectors
Packet size / injection	256-bit packets injected with an exponentially-distributed inter-packet gap
Traffic Pattern	Uniform Random Traffic
Technology library	22 nm low-voltage library of DSENT
SiP Technology Parameters	Laser and MR heating power based on Table 5.1
Buffer space	32 bits buffer space for each arbitration packet
Other	1 mm tile dimensions, 10.45 ps/mm signal propagation of light, 1 cycle OE and tuning delay

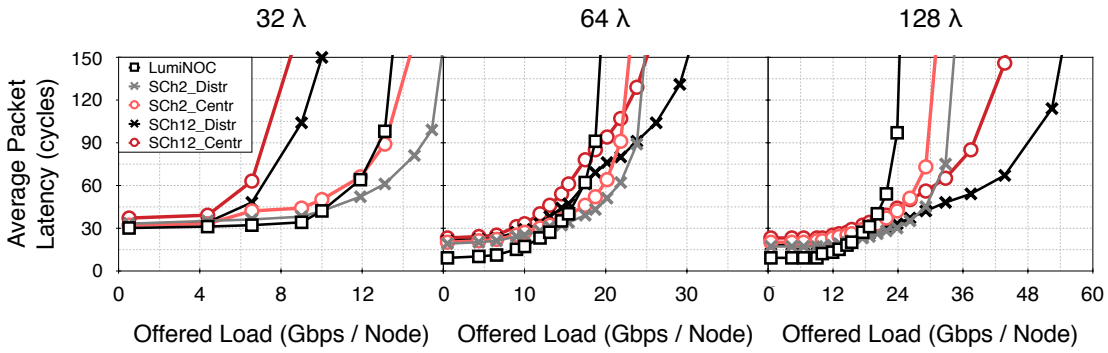
will be compared to LumiNOC for different relevant on-chip bus sizes (i.e. 8, 12, 16) to investigate different design points for their implementation in realistic NoCs and to study scalability. For each bus size, we consider 32λ, 64λ, and 128λ bandwidth on the bus (consisting of multiples of 32λ buses as described in Section 6.2). Nodes are scheduled on the bus based on round-robin scheduling.
Table 6.1 lists a summary of the experimental set-up describing the used modelling tools, traffic generation, and technological assumptions regarding latency and power.

Isolated Bus Evaluation

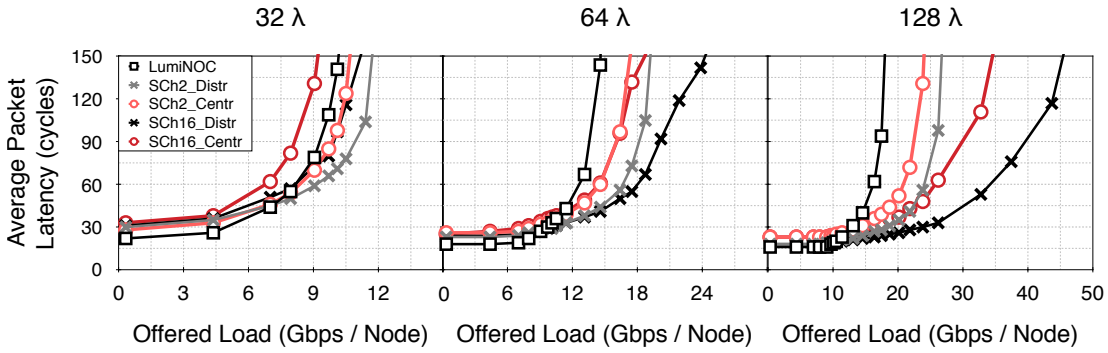
Latency and Throughput Figures 6.12a, 6.12b, and 6.12c illustrate the average packet latency for buses utilising centralised (Centr) and distributed (Distr) arbitration and LumiNOC for varying bus bandwidth (i.e. the number of wavelengths (λ)). 'SCh' indicates the number of subchannels. The proposals were simulated for two subchannels to show the effect of the potential minimum improvement, as well as for N subchannels (N being the number of nodes) which is expected to provide the highest throughput based on the analysis in Section 6.3.
Subchannel scheduling becomes increasingly beneficial for larger numbers of wavelengths and subchannels. Throughput benefits grow along with the bus size. For 12 nodes and 32λ, however, it actually provides less throughput, likely because bandwidth for arbitration is unevenly split between the nodes (32 is not divisible by 12), leading to bandwidth being wasted in the arbitration phase. This has a higher impact on the arbitration mechanisms for subchannel scheduling than on LumiNOC since arbitration phases feature larger arbitration packets. In addition, bandwidth is wasted during



(a) 8 Nodes



(b) 12 Nodes



(c) 16 Nodes

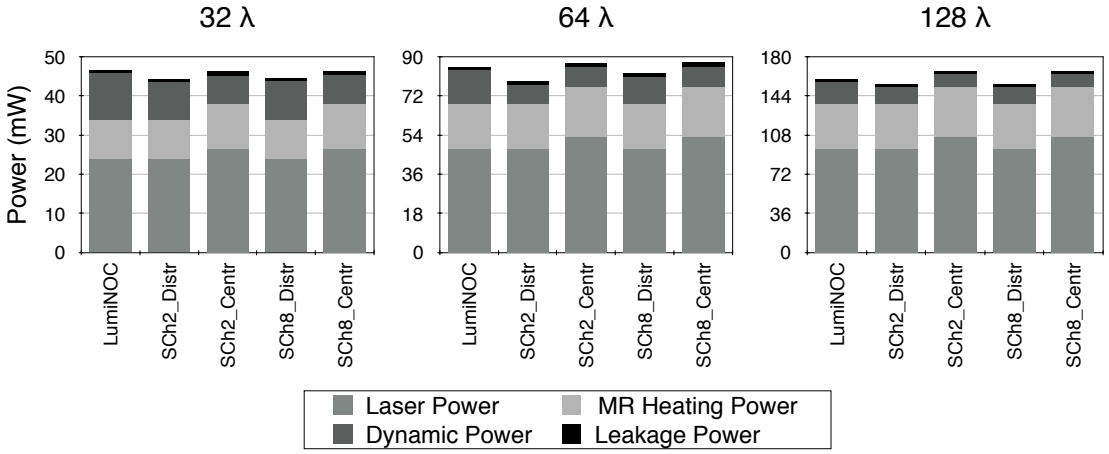
Figure 6.12: Average Packet Latency: In-band Arbitration

the data transmission phase because the number of wavelengths is not divisible by the number of subchannels (32 and 12). Therefore, bus size should be divisible by the number of wavelengths.

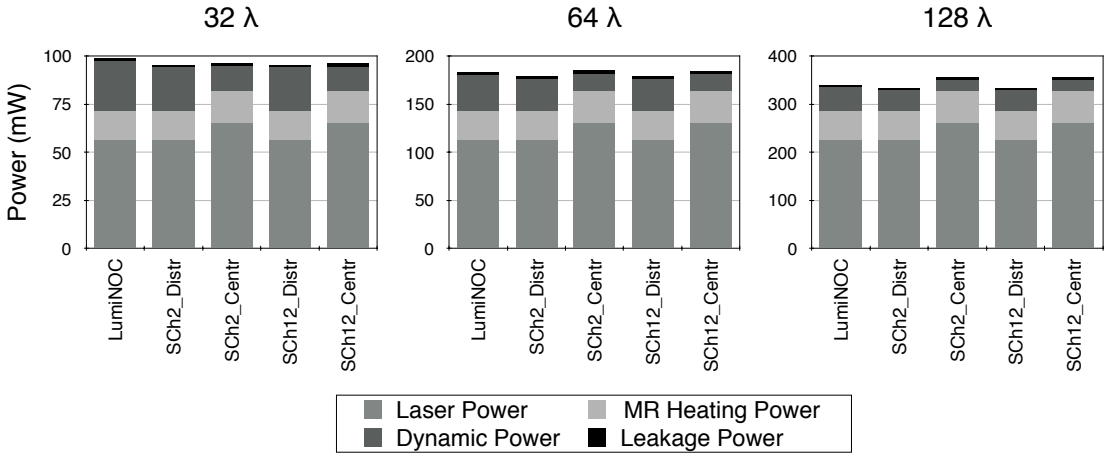
Similarly, the more optical bandwidth is offered to the bus, the higher the throughput gains of subchannel scheduling. Bandwidth is also shared during the arbitration phase, in which each node receives a subset of λs for modulating arbitration packets. Higher

bandwidth thus leads to higher bandwidth per node in the arbitration phase and is beneficial to subchannel arbitration proposals which exhibit larger arbitration packets. The overall impact on arbitration latency is thus higher on the subchannel mechanisms than on LumiNOC, and improves the throughput gains of the former even further.
LumiNOC offers slightly less packet latency for low network loads as its arbitration mechanism is simpler and features a speculative transmission scheme. Although these latency benefits can be up to 30% in the worst case at very low network loads (not including the 12-node bus where bandwidth is not evenly distributed), latency differences shrink as bus bandwidth increases (down to 10%). The benefits of subchannel scheduling increase along with the network loads and eliminate the arbitration overheads for moderate-to-high loads. Subchannel scheduling can improve throughput by $>1.6\times$ for 64λ and $>2\times$ for 128λ for all bus sizes, confirming the assumption that large throughput gains can be achieved; however, it also reveals that buses with subchannel scheduling should be implemented in NoCs so that bus utilisation is kept high and with sufficient bandwidth to avoid latency overheads for low loads.

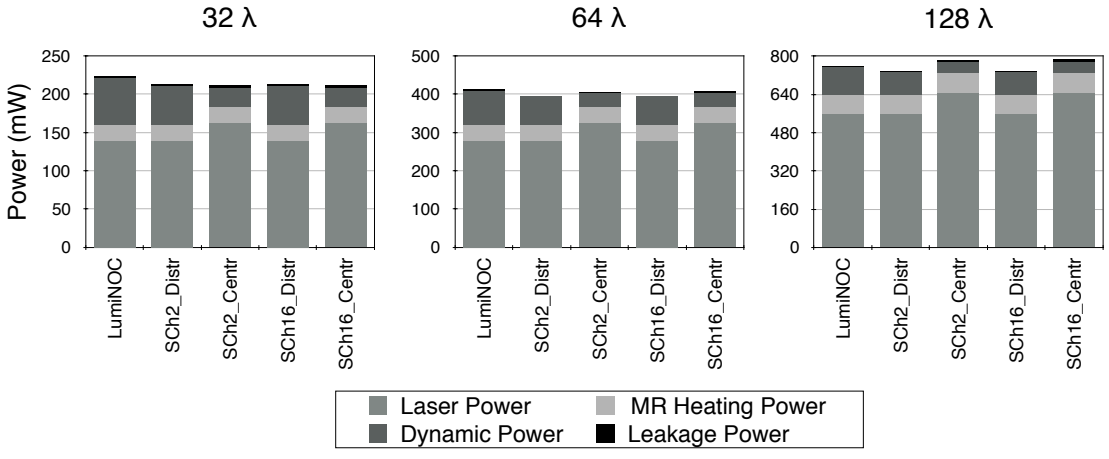
Power Consumption As discussed in Section 6.2, bandwidth on the buses is scaled by implementing multiple physical 32λ buses (logically, all nodes still see the bus as *one* bus). Leakage power includes the static electrical power for buffering arbitration packets and the EO/OE backends at the nodes (and in the arbiter in the centralised approach). Each node requires buffers for storing one REQ and one ACK packet, and the centralised arbiter buffers for N REQs and N ACKs (for N number of nodes). As listed in Table 6.1, pessimistic buffer space of 32 bits is apportioned for each arbitration packet. Dynamic power was captured at the saturation points of LumiNOC.
Figures 6.13a, 6.13b, and 6.13c show the power breakdowns of 8, 12, and 16 nodes, respectively, for different bus bandwidths. Static power consumed at the laser source and for MR heating is the major contributor to the total power. SCh_Centr exhibits slight overheads in terms of both these two metrics due to the additional number of MRs at the arbiter and the associated MR-through losses and heating (at most 5%). Moreover, the circuitry in the centralised arbiter causes higher leakage power. The benefits of SCh_Centr compared to LumiNOC or SCh_Distr is lower dynamic power due to smaller arbitration packets and the absence of broadcasting (apart from the 'max_cyc' field which is small in comparison). This cancels out some of the static power overheads of the centralised arbiter, leading to only very small power overheads overall. In addition, dynamic power scales much better for larger bus sizes (see Figure 6.13c)



(a) 8 Nodes



(b) 12 Nodes



(c) 16 Nodes

Figure 6.13: Power Breakdown: In-band Arbitration

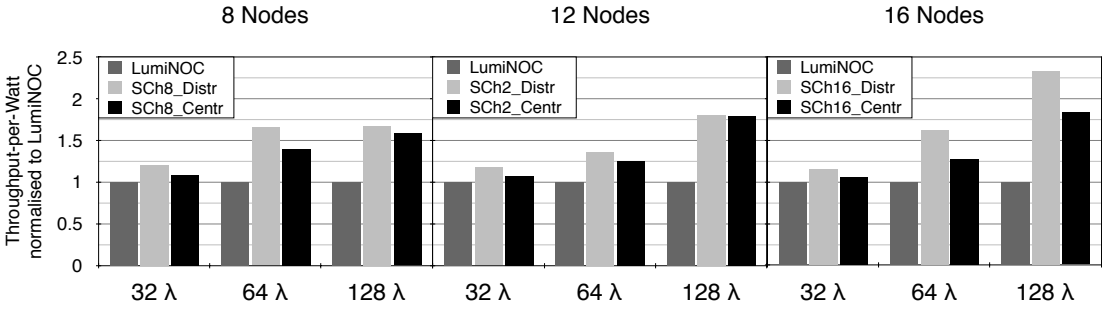


Figure 6.14: Throughput per Watt Comparison of In-band Arbitrated Buses

because arbitration packets are mainly unicast and increase with $log2(N)$ in SCh_Centr (for N nodes), while both LumiNOC and SCh_Distr require broadcasting bitmaps of the source addresses and thus scale with N.

SCh_Distr consumes the least power for 8 nodes where the static power overheads of SCh_Centr are more significant than its dynamic power savings. SCh_Distr also consumes less dynamic power than LumiNOC because senders in LumiNOC must broadcast arbitration packets twice in case of contention (once initially, and one more time prior to data transmission) while senders in SCh_Distr broadcast in phase 1 and unicast in phase 2. Moreover, for synthetic traffic with just *one* packet size, the 'packet length' fields are not needed in both mechanisms, which leads to a smaller arbitration packet size in SCh_Distr as it decreases the arbitration packets to $(dst\_id + Src\_bitmap)$ (LumiNOC) and (Src_bitmap) (SCh_Distr in phase 1). SCh_Distr's dynamic power benefits might thus decrease when more packet sizes must be supported by the NoC.

Throughput-per-Watt TPW is computed by dividing the maximum injection rate per node into the bus prior to network saturation by the consumed power. Figure 6.14 presents the TPW results normalised to LumiNOC for the most competitive design points of the subchannel approaches (i.e. when the number of subchannels equals the number of nodes). For all bus sizes and bandwidths, subchannel scheduling provides higher power efficiency than LumiNOC. In line with the observed throughput gains, both subchannel arbitration approaches become increasingly beneficial to LumiNOC as the number of nodes and bus bandwidth increases.

6.3.4 Discussion

This section showed that arbitration mechanisms implementing subchannel scheduling can be designed efficiently and enable large throughput gains compared to state-of-the-art sequential scheduling. As bandwidth is increased, bus arbitration can be performed faster and subchannel scheduling becomes increasingly beneficial. If power overheads are incurred, they are very small ($< 5\%$), and latency overheads due to higher arbitration complexity are only noticeable for very low injection rates. However, these overheads can become significant if the injection rates and the bandwidth on the bus are low. For applications that are latency critical (and not bandwidth critical), LumiNOC's speculative arbitration approach is superior. If bandwidth is the main requirement, subchannel scheduling is the preferred choice. Alternatively, a mechanism capable of switching between these two bus arbitration schemes dynamically based on the current traffic demands of the NoC would be ideal, especially since many applications exhibit phases of low and high network utilisation [BKSL08].
One corner case that has not been fully captured by this study is the variability of the ACK packet size in the centralised arbitration approach; in particular, the case when one node is receiving a disproportionate number of packets in a data transmission round which would require to append the fields encoding the subchannel and time slot for each sender, and potentially lead to large arbitration overheads in extreme cases (e.g. if a large number of nodes is connected to the bus and each node requests to send to the same destination). As discussed in previous chapters this behaviour is relatively common in shared memory applications. Therefore, analysing such adversarial cases would represent an interesting future study to identify their impact on both arbitration latency and energy.

6.4 In-band vs. Parallel Bus Arbitration

In this Section, we argue that performing arbitration in-band, i.e. stopping data transmission to perform bus arbitration, limits the maximum achievable throughput because of the latency overheads of the arbitration mechanism. Implementing a dedicated bus to perform arbitration independently of data transmission might, at least partially, hide these arbitration overheads [PKM10]. Indeed, Kakoulli et al. [KSKK15] recently evaluated LumiNOC with a dedicated arbitration bus and report promising results.
Although an arbitration bus imposes area overheads, those are expected to be less critical in future CMPs [JBK+09], particularly as shared optical buses already present

a layout-efficient architecture (just one waveguide). Besides, power overheads could actually be kept relatively low since arbitration packets are much smaller than data packets, meaning that much less bandwidth is required on the arbitration bus (which, in turn, translates to less power). In addition, bandwidth on the arbitration bus could be scaled independently from the data bus which is not possible with in-band arbitration where the same bandwidth is used both for arbitration and data transmission.
The throughput gains of parallelising arbitration could be large enough to outweigh these power overheads and provide higher power efficiency overall. In particular, this could be key for the arbitration mechanisms supporting subchannel scheduling in which arbitration latency is slightly higher due to higher complexity.
This section studies the costs of implementing a parallel arbitration bus and its impact on subchannel scheduling, and proposes efficient solutions to parallelise the arbitration mechanisms discussed in the previous section.

6.4.1 Efficient Bus Utilisation With Parallel Arbitration

Parallel bus arbitration should keep the data bus as busy as possible in order to attain high throughput. Kakoulli et al. [KSKK15] evaluated LumiNOC with a parallel arbitration bus on which the abbreviated arbitration flags that are sent in the dynamic scheduling phase are transmitted in parallel to the ongoing data transmission. In particular, senders start modulating flags so that data transmission can begin directly after the previous sender has finished. For instance, when the time it takes to finish bus arbitration is arb_delay and the cycle at which the bus is free again is $t\_busfree$, then bus arbitration should start at cycle $(t\_busfree - arb\_delay)$. Figure 6.15 illustrates this: assuming bus arbitration requires six cycles and data transmission on the data bus is finished after cycle 17, bus arbitration must start in cycle 12 so that the potential senders requesting the bus can start utilising it in cycle 18.
While this hides the arbitration delay for all arbitration approaches, it has an additional side effect on the subchannel arbitration approaches proposed in the previous section: since nodes wait until the last possible moment to perform arbitration, the likelihood of them actually requesting the bus is the highest because the time a node waits for packets to arrive is maximised. This minimises the total number of arbitration phases for a given number of packets in the NoC and should thus provide high throughput. In LumiNOC, the main benefits are only in the dynamic scheduling phase upon collision-detection, at which point the arbitration phase has already started and all nodes have

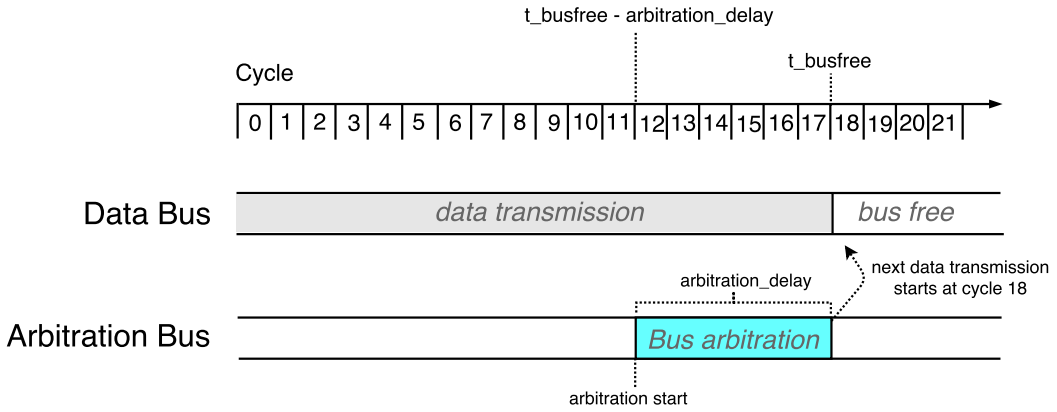


Figure 6.15: Start of arbitration phase on the parallel bus

to wait for it to finish before being able to request the bus again. This is because LumiNOC has basically several small arbitration phases (one for each requester) in the dynamic scheduling phase in which future receivers are notified by the senders, while in the proposed subchannel approaches one (bigger) arbitration phase computes all the scheduling information for each sender-receiver pair in one step.
All arbitration mechanisms discussed in the following (LumiNOC, distributed and centralised arbitration) perform parallel bus arbitration according to this approach.

6.4.2 Parallel Bus: Centralised Arbitration

In both SCh_Distr and SCh_Centr each node is assigned to a distinct λ-subset during arbitration. In SCh_Centr, however, nodes use their subsets to communicate with the centralised arbiter and do not require a broadcast mechanism at each sender. This simplifies the arbitration bus since each node only needs modulators for their λ-subset, and not for the entire set of wavelengths on the bus. Merely the centralised arbiter requires modulators and filters for all wavelengths. Figure 6.16 exemplifies a parallel arbitration bus for centralised arbitration for four nodes and one wavelength per node. The information encoded in the arbitration packets does not differ from in-band arbitration introduced in the previous section.

6.4.3 Parallel Bus: Distributed Arbitration

The design of an efficient parallel arbitration bus requires an analysis of the communication pattern of the arbitration mechanism, and should cause as little overhead as possible. The arbitration mechanisms of LumiNOC and SCh_Distr does not change from

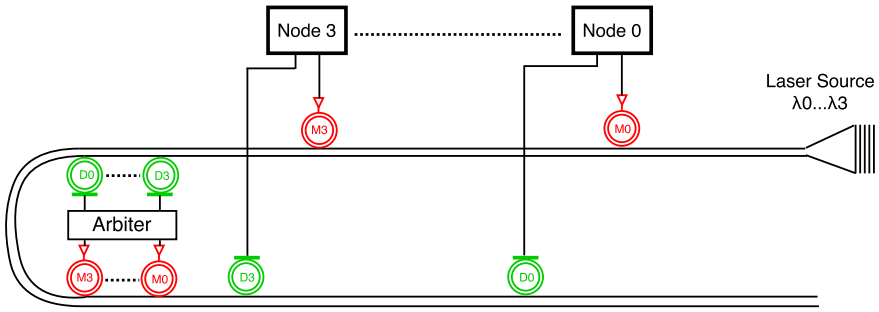


Figure 6.16: Parallel Arbitration Bus: Centralised Arbitration

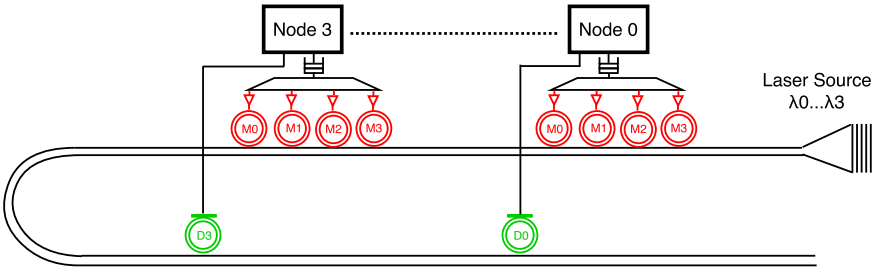


Figure 6.17: Parallel Arbitration Bus: Distributed Arbitration

the in-band arbitration case introduced in the previous section since the required information at each node to compute the scheduling remains the same. Both LumiNOC and SCh_Distr have to broadcast the arbitration flags by modulating them on the λ-subsets of each node connected to the bus. Therefore, modulators are required on the entire optical bandwidth of the arbitration bus; however, other than on the data bus, each node only needs MR filters on their own λ-subset during arbitration, leading to a total of $((N+1)\times\lambda)$-MRs on the bus (for N nodes). This decreases the total number of MRs and MR-through losses compared to a shared optical bus, and, in turn, reduces laser and MR heating power. Figure 6.17 illustrates an example of this arbitration bus design for four nodes and one wavelength per node.

As described above, each arbitration round starts at cycle $(t\_busfree - arb\_delay)$ at which point all nodes start broadcasting their arbitration packets to correctly produce the bitmap-overlapping that detects multiple requesters. After the arbitration phase, each node knows about the senders and receivers, as well as their packet lengths, which allows each node to compute the number of clock cycles the data bus will be utilised, and based on that the starting point of the next arbitration round.

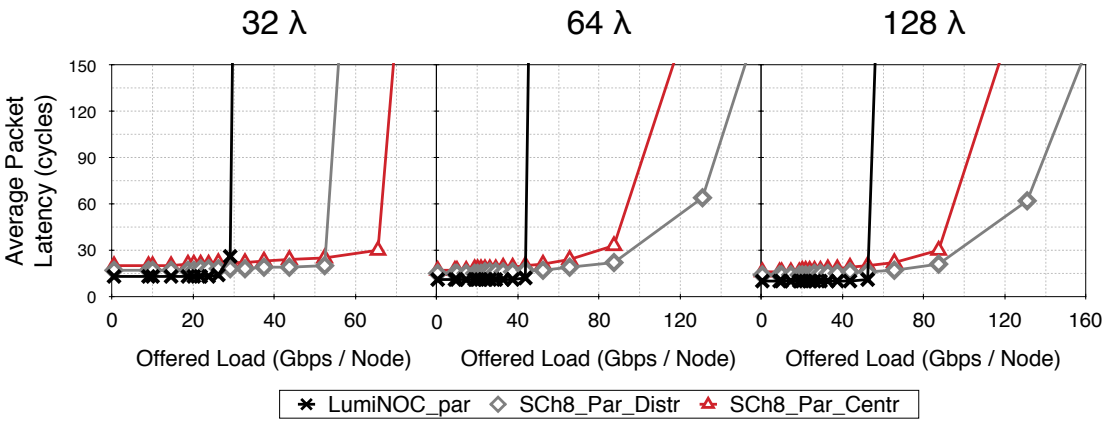
6.4.4 Evaluation

Methodology

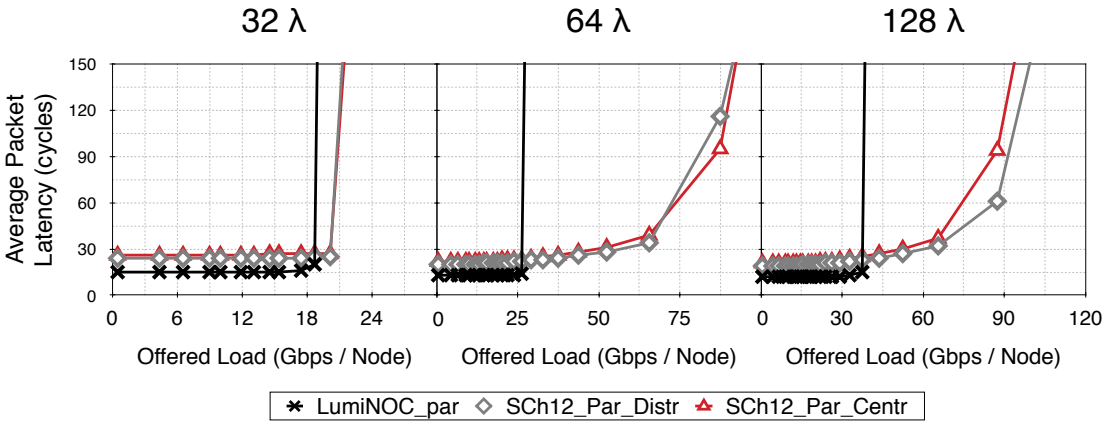
This study uses the same experimental set-up as the in-band arbitration study in the previous section (see Table 6.1). The arbitration bus provides *two* wavelengths to each node for transmitting arbitration packets, which is an efficient design point for a low-power arbitration bus that does not incur large latency overheads (for the investigated design points). Like in the previous section, buses of different sizes (8, 12, and 16 nodes) and bandwidths (32λ, 64λ, 128λ) are considered to study a range of design points and scalability.
The designs with parallel buses that utilise subchannel scheduling are denoted with 'SCh_Centr' and 'SCh_Distr' for centralised and distributed arbitration, respectively, and are compared to the sequential scheduling in LumiNOC. The number after 'SCh' indicates the number of subchannels used (e.g. SCh8_Centr for 8 subchannels in the centralised approach), which we restrict to the most efficient design points revealed in the previous section (the number of nodes equals the number of subchannels in most cases). In the studies comparing the parallel arbitration approaches to the in-band arbitration approaches, '_Par' and '_InB' is added to the names to denote parallel and in-band arbitration, respectively.

Latency and Throughput

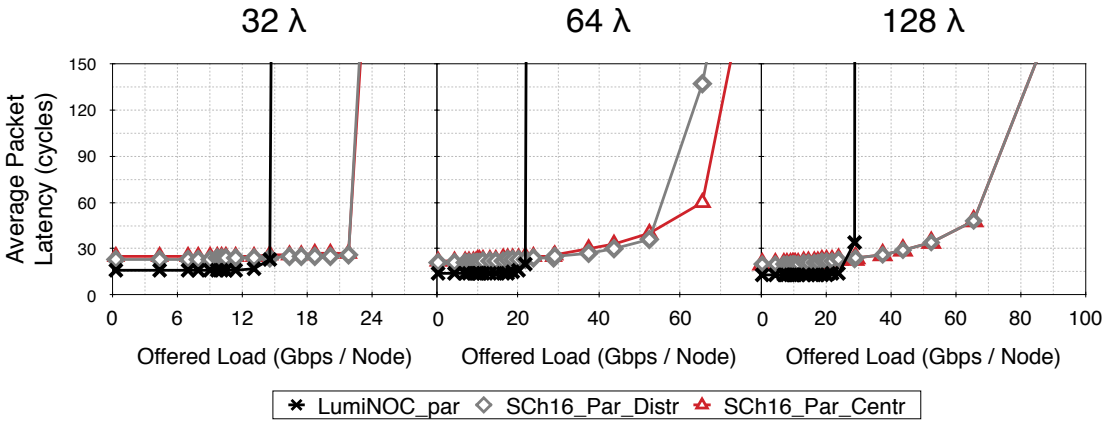
Figures 6.18a, 6.18b, and 6.18c depict the average packet latency for varying injection rates and bus bandwidth for 8, 12, and 16 nodes, respectively. Generally, we observe the same trends as with in-band arbitration: subchannel scheduling offers large throughput gains, particularly with increasing number of nodes and bus bandwidth. For 64λ and 128λ, throughput is more than doubled for all considered bus sizes. In addition to that, the latency benefits of LumiNOC for low network loads are decreased compared to the in-band case, suggesting that the latency overheads of more complex arbitration schemes can be reduced by performing bus arbitration in parallel rather than in-band. SCh_Distr can sustain higher injection rates than SCh_Centr for 8 nodes; however, both show very similar latency curves for 12 and 16 nodes since SCh_Distr relies on transmitting source bitmaps in the arbitration phase which grow linearly with the number of nodes – its performance benefits compared to SCh_Centr therefore shrink.



(a) 8 Nodes



(b) 12 Nodes



(c) 16 Nodes

Figure 6.18: Average Packet Latency: Parallel Bus Arbitration

Power Consumption

Parallel Arbitration Approaches The first concern of utilising a separate arbitration bus is how much power overheads it incurs in comparison to the data bus. Compared to in-band arbitration, an arbitration bus entails power overheads that include

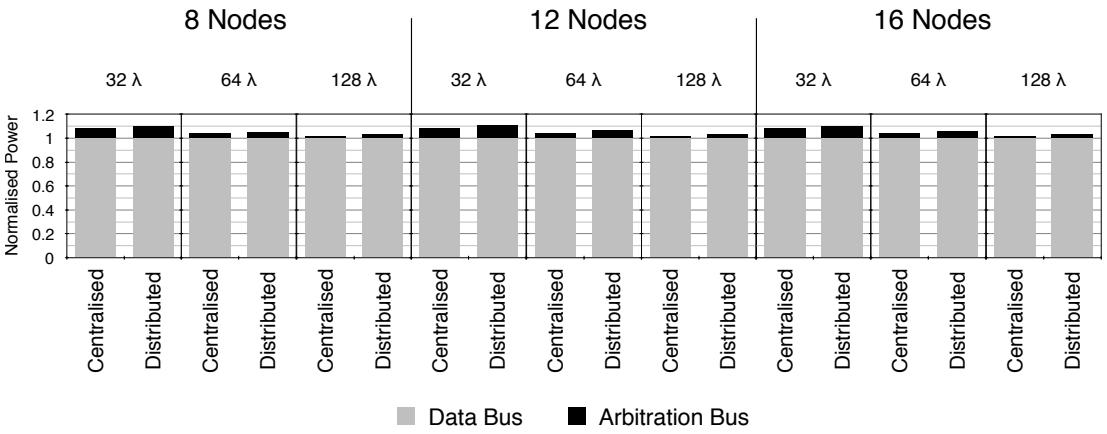


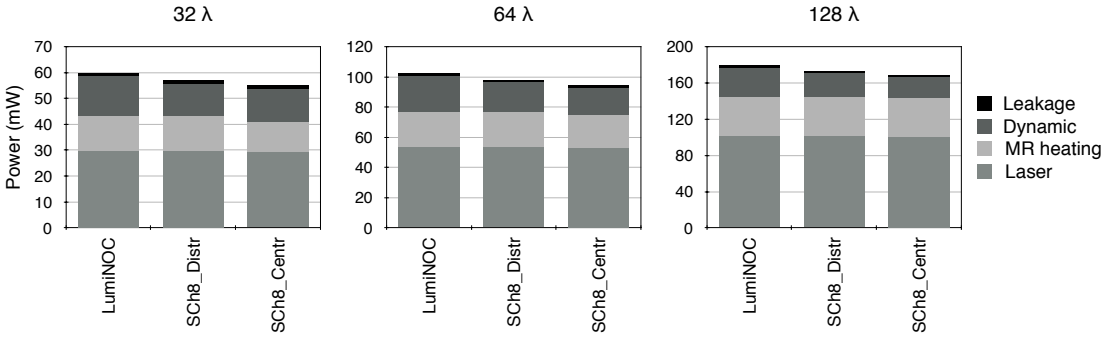
Figure 6.19: Power Overheads of the Parallel Arbitration Bus

leakage power in the EO/OE backend circuitries required for the modulators and receivers, laser and MR heating power. Figure 6.19 illustrates the static overheads for the considered bus sizes and widths. We omit dynamic power in these charts since the arbitration mechanisms do not change compared to the in-band case, i.e. the same arbitration packets are exchanged, leading to the same dynamic power. The charts normalise power consumption to the data bus, i.e. the data bus equals one and the arbitration bus represents its overheads when added to the data bus.

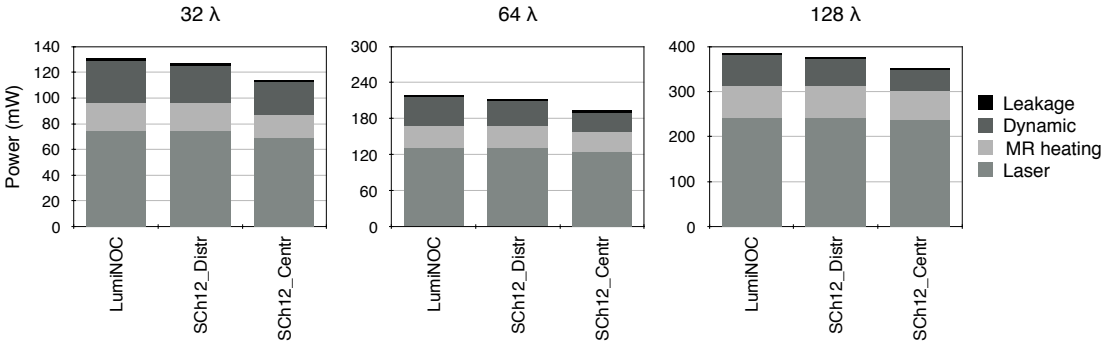
The contribution of the arbitration bus to the total power is less than 10% in all cases. As bandwidth on the data bus and bus size increases, so does laser and MR heating power of the data bus which renders the power on the arbitration bus insignificant in relation to the total power.

Differences between LumiNOC, SCh_Distr, and SCh_Centr in terms of power consumption stem from the different arbitration bus designs (distributed vs. centralised), and differences in dynamic power from the different arbitration packet sizes and exchange patterns in the arbitration stage. Figures 6.20a, 6.20b, and 6.20c show the power breakdown of the different bus designs for the bus sizes under investigation. Dynamic power has been extracted before the saturation point of LumiNOC.

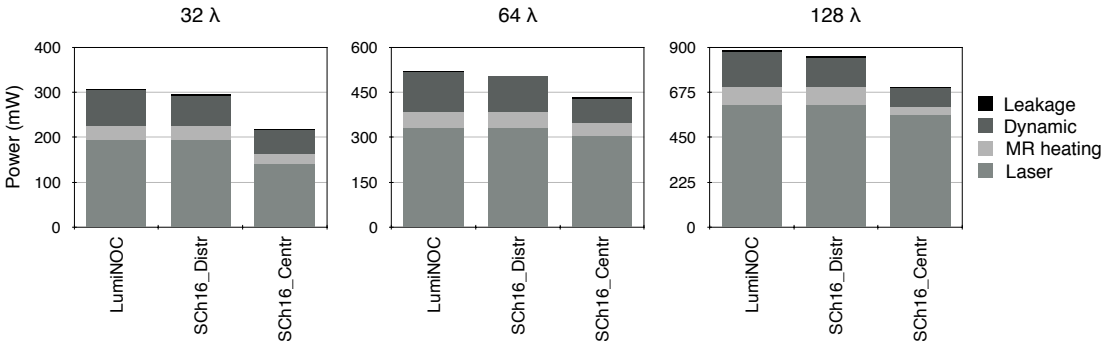
Although requiring separate arbitration circuitry, SCh_Centr consumes the least power out of all approaches because its arbitration bus has fewer MRs and lower optical path losses due to less MR-through loss on the data path. In LumiNOC and SCh_Distr, each node has modulators for each wavelength on the bus, which leads to a large number of MR passings of the optical signal, which, in turn, leads to increasing laser power as the number of nodes increases. In SCh_Centr, on the other hand, each node has modulators only on its λ-subset and only the arbiter has modulators and filters for each



(a) 8 Nodes



(b) 12 Nodes

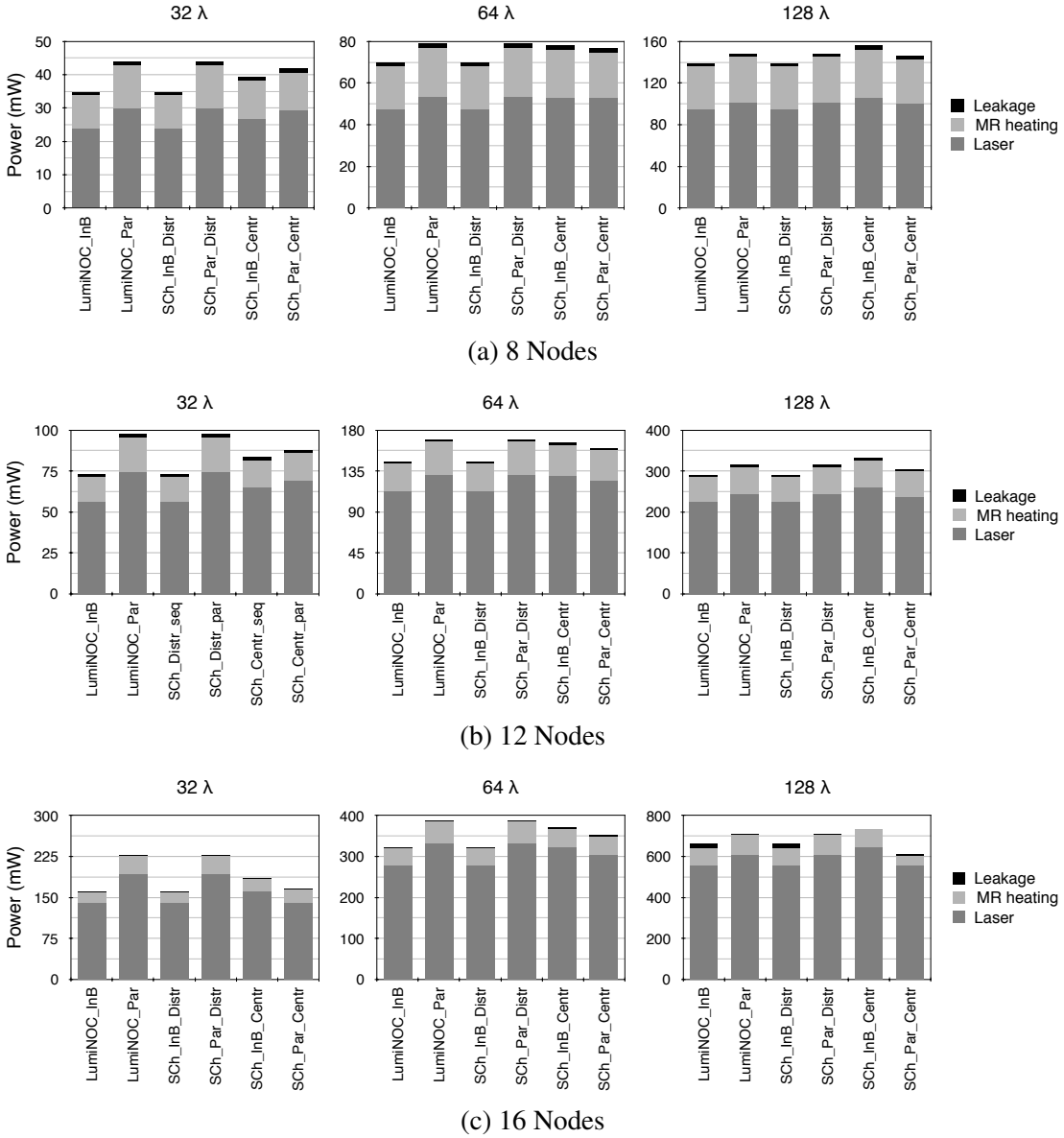


(c) 16 Nodes

Figure 6.20: Power Breakdown: Parallel Bus Arbitration

λ on the bus. The number of MRs and MR-through losses in SCh_Centr is thus lower and scales better, which reduces laser and MR heating power. In addition, as observed in the previous section, fewer arbitration packets are exchanged in SCh_Centr, leading to lower dynamic power. These savings outweigh the power incurred by the circuitry overheads of the centralised arbiter.

SCh_Distr saves power compared to LumiNOC for the same reasons it does with in-band arbitration: fewer arbitration packets with smaller packet sizes are exchanged when only one packet size is considered. For two packet sizes, these savings are likely



(a) 8 Nodes

(b) 12 Nodes

(c) 16 Nodes

Figure 6.21: Static Power Comparison: In-band vs. Parallel Arbitration

to decrease since LumiNOC only requires to exchange a field of $log_2(pkt\_sizes)$, while SCh_Distr requires to exchange a bitmap.

In-Band vs. Parallel Arbitration Approaches Figures 6.21a, 6.21b, and 6.21c depict a comparison of the in-band and parallel arbitration designs. LumiNOC and SCh_Distr utilise the same arbitration bus, therefore the overheads of parallel arbitration are the same. The general trend can be observed that with increasing number of wavelengths on the data bus, the overheads of the arbitration bus lose significance and

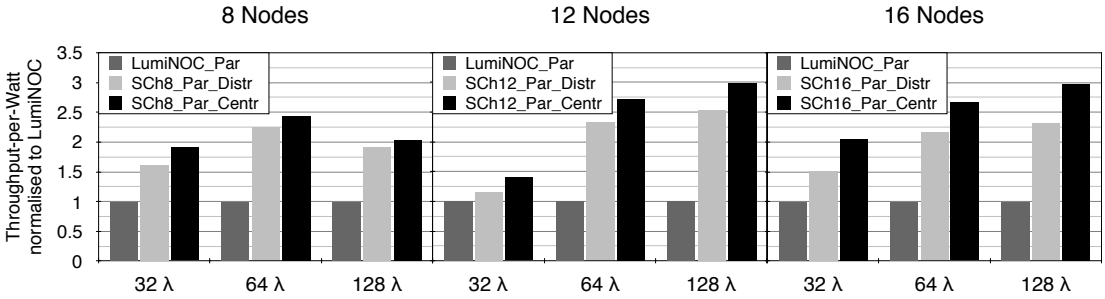


Figure 6.22: Throughput per Watt: Parallel Bus Arbitration Approaches

become relatively smaller. In the worst case (32λ/16 nodes), the power overheads of the parallel arbitration bus are 27%; however, they can be as small as 5% (e.g. 128λ/8 nodes).

The power results of SCh_Centr seem unintuitive at first because, with a parallel arbitration bus, the power consumption is actually lower in most cases. This is because the arbitration bus has 2λ per node, which results in fewer MRs in the arbiter (in most cases). For instance, for 16 nodes, a total of 32λ are on the arbitration bus, which would lead to the same number of MRs in the arbiter as in the 32λ in-band bus case. Effectively, in these cases, less bandwidth is available for the arbitration phase on the parallel bus than in the in-band case (e.g. for 8 nodes and 32λ, each node has a λ-subset of 4λ). Clearly, power overheads would be imposed if the parallel arbitration bus had the same number of wavelengths per node as in the in-band case.

This observation reveals another weak point of performing arbitration in-band: bandwidth for the arbitration round is fixed and dependent on the bandwidth required for data transmission, which leads to an inflexible design and over-provisions the bus arbitration phase in which small arbitration packets may not need that much bandwidth. Although the authors of LumiNOC propose to use some of the bandwidth during arbitration for credit return of the flow control mechanism, this is unlikely to fully utilise the entire bandwidth efficiently if, for instance, 128λ are needed on the bus. Parallel arbitration, on the other hand, offers a good opportunity to reduce these costs as it allows to adjust the bandwidth on the arbitration bus independent from the data bus.

Throughput-per-Watt

Parallel Arbitration Approaches Figure 6.22 shows the TPW for the different parallel arbitration buses normalised to LumiNOC. Dynamic power is reported at the saturation point of the different buses. Generally, parallelised subchannel scheduling

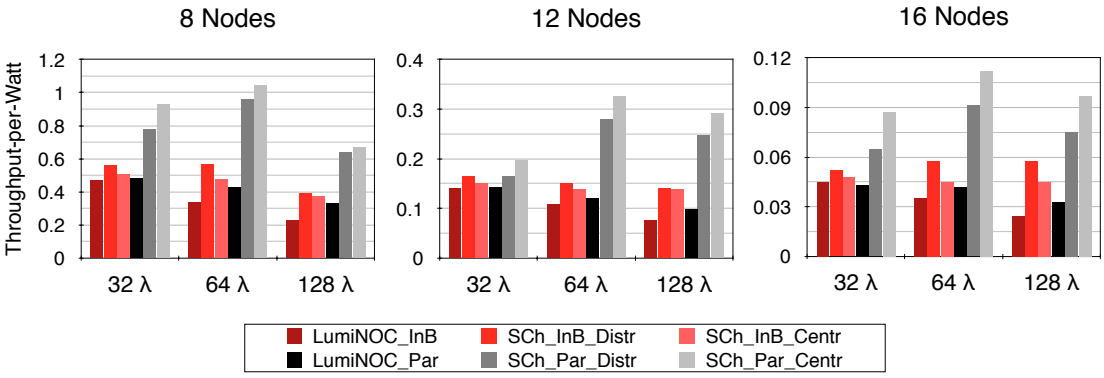


Figure 6.23: Throughput-per-Watt: In-band vs. Parallel Arbitration

improves power efficiency tremendously compared to the sequential scheduling of LumiNOC. In most cases, TPW is more than doubled, with up to 3× the TPW for 16 nodes/128λ for SCh16_Centr. The proposed arbitration mechanisms supporting subchannel scheduling not only improve throughput, but also consume slightly less power, making it the overall superior design choice. The trends with regard to bandwidth scaling on the data bus are similar to the in-band case: the higher the bandwidth on the data bus, the greater the benefits of subchannel scheduling.

In-band vs. Parallel Arbitration Approaches Figure 6.23 compares the TPW of the different arbitration approaches to their in-band counterparts. All of the designs are more power-efficient when bus arbitration is parallelised. With regard to subchannel scheduling, TPW improvements are considerable, especially as the number of nodes and bandwidth on the buses increases. Another indication of the efficiency of subchannel scheduling can be observed when comparing SCh_InB_Centr and SCh_InB_Distr to LumiNOC_par: in all cases, subchannel scheduling with in-band arbitration outperforms LumiNOC although its bus arbitration is parallelised.
For 32λ and 12/16 nodes, the power overheads of the parallel arbitration bus outweigh the throughput gains in LumiNOC and SCh_InB_Distr. This suggests that, if the bandwidth on the data bus is too low, parallelising bus arbitration does not improve power efficiency in these cases. Parallel arbitration is, therefore, particularly beneficial if high bandwidth is provided on the data bus.

Area Overheads

Implementing a parallel arbitration bus naturally imposes certain hardware overheads. Figure 6.24 shows these overheads for the considered bus sizes and widths. Note that

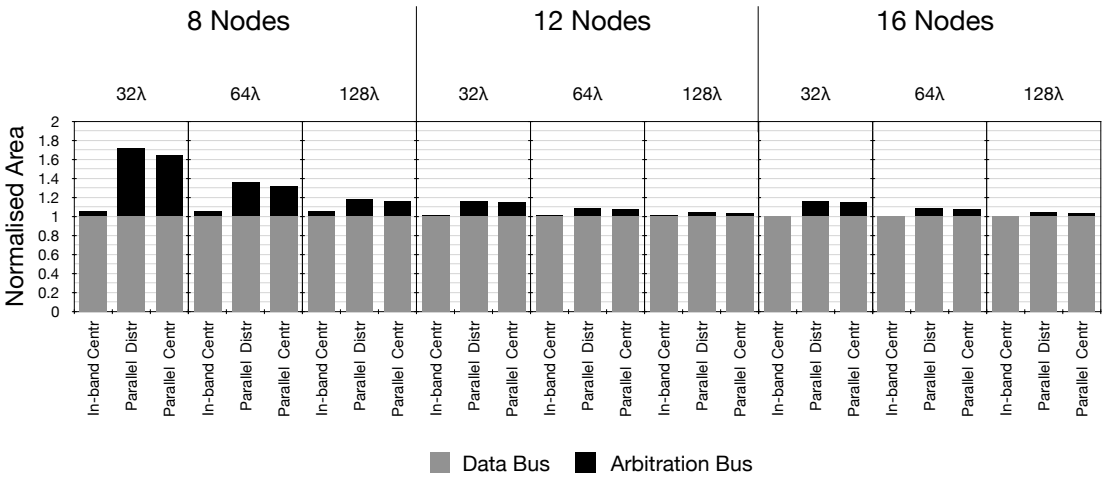


Figure 6.24: Area Overheads: In-band vs. Parallel Arbitration

the area results are normalised to the data bus in each case, i.e. the overheads represent the area for the arbitration bus. Area values were extracted with DSENT using its 22 nm technology library for electronic components, 5 μm waveguide pitch and 10 μm^2 MR area. For the in-band buses, we only consider those with centralised arbitration, in which case the area refers to the resource requirements of the arbiter. In distributed in-band buses, there is no area overhead for arbitration (buffers for REQs/ACKs are negligible compared to SiP components in terms of area). In fact, the obtained results during this study showed that area required for the electrical backends and buffers in the arbiter is negligible in all cases.

The overheads for parallelising arbitration are the highest for the 32λ buses because the data bus also consists of just one waveguide and has in total relatively few MRs compared to 64λ and 128λ buses. In these two cases, one and three more 32λ buses are required, respectively, which renders the area overheads of the control bus less significant. For 12 and 16 nodes, the control bus becomes insignificant compared to the data bus for 64λ and 128λ.

6.4.5 Discussion

Performing bus arbitration in parallel to data transmission on a separate arbitration bus is a more efficient design point compared to in-band arbitration where wavelengths are re-used for both arbitration and data transfer. This is particularly the case with increasing bandwidth and number of nodes on the data bus as it de-emphasises the overheads

of the arbitration bus. The benefits with regard to power efficiency of parallel bus arbitration decreases compared to in-band arbitration for low bandwidth on the data bus. Bandwidth on the arbitration bus can be provisioned flexibly and does not depend on the bandwidth of the data bus like in the in-band case. This allows for light-weight designs since 1) arbitration packets are smaller than data packets and thus require less bandwidth and 2) arbitration delay can partly be hidden by performing arbitration simultaneous to data transmission. Finally, performing bus arbitration in parallel amplifies the throughput gains of subchannel scheduling with in-band arbitration compared to LumiNOC. This is because, while bus arbitration takes longer for subchannel scheduling compared to LumiNOC, it has a less limiting factor on throughput and latency when parallelised.

6.5 Scaling up to Larger NoCs

The previous sections have revealed the superiority of subchannel scheduling both for in-band and parallel arbitration on a single, shared optical bus. A study investigating how these improvements translate to NoCs of realistic sizes that rely on these buses would allow to estimate their overall impact and could provide interesting insights.

6.5.1 Topology

This section studies the impact of utilising subchannel scheduling with the different arbitration schemes when implemented in the topology proposed by Li et al. [LBGP14], which is depicted in Figure 6.25. In this NoC, shared buses are implemented in the rows and columns, and dimension-order XY-routing is performed if destinations cannot be reached within one hop (i.e. if the destination is not in the same row or column). In that case, a packet would be sent on the row bus to an intermediate node that resides in the same column as the destination.
The results of the previous study indicate considerable throughput gains by performing bus arbitration in parallel. Rather than using this additional throughput to improve the overall NoC performance, this can also be leveraged to implement low-power designs by using fewer buses with core clustering. In order to identify to which extent this can be beneficial, we study an additional NoC layout shown in Figure 6.26. In this NoC, two nodes are clustered at one router. Compared to the NoC in Figure 6.25, this allows to 1) halve the number of shared buses in the rows and 2) halve the number

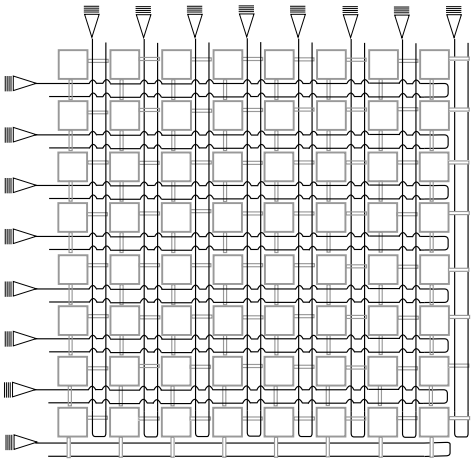


Figure 6.25: 64-node NoC without clustering: shared buses are placed in rows and columns

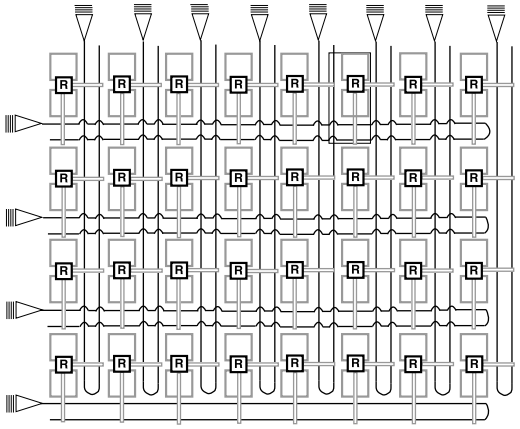


Figure 6.26: 64-node NoC with clustering: two nodes are grouped at each router

of nodes connected to a shared bus in the columns, which allows to reduce power and resources. This clustered NoC implements the buses with the highest reported throughput, i.e. parallel-arbitrated and subchannel-scheduled, in order to identify whether the throughput gains of these buses can enable the power reductions of core clustering while sustaining performance levels.

6.5.2 Evaluation

Methodology

The experimental set-up is the same as in the previous sections (see Table 6.1); however, the NoCs are stressed with three different synthetic traffic patterns: uniform random, bit complement, and tornado traffic. Bandwidth on the buses is studied for 64λ and 128λ, with 2λ per node on the arbitration buses for the parallel arbitrated buses. Although state-of-the-art NoC router designs enable a packet traversal delay of merely one clock cycle, they run at low-GHz frequencies (typically 1-2 GHz) [PKC+12] [HDV+11]. Routers clocked at 5 GHz, as assumed in this study, require deeper pipelines with more stages, which increases the router traversal delay. In addition, the routers in the NoCs investigated in this study have a fairly high radix (a router can theoretically receive a packet from each node connected to the bus simultaneously), which further complicates the router architecture [JP09]. Therefore, we assume a more pessimistic router traversal delay of three clock cycles. Leakage power in the routers was estimated with typical NoC router buffer specifications of 7 virtual

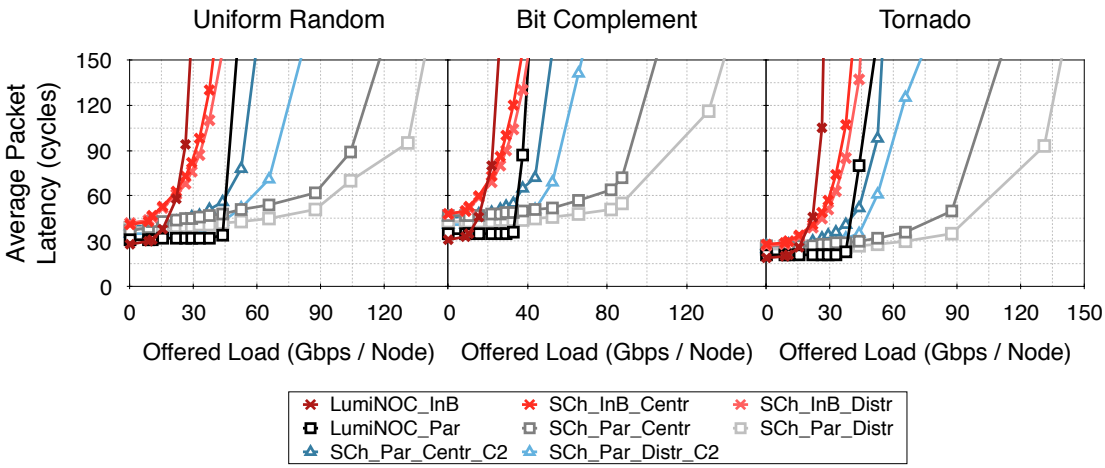
Table 6.2: Bus and NoC Configuration and Description

Bus Name	Arbitration	Scheduling	NoC Topology
LumiNOC_InB	In-band, LumiNOC	sequential only	No clustering (Fig.6.25)
SCh_InB_Centr	In-band, centralised	subchannel	No clustering (Fig.6.25)
SCh_InB_Distr	In-band, distributed	subchannel	No clustering (Fig.6.25)
LumiNOC_Par	Parallel, LumiNOC	sequential only	No clustering (Fig.6.25)
SCh_Par_Centr	Parallel, centralised	subchannel	No clustering (Fig.6.25)
SCh_Par_Distr	Parallel, distributed	subchannel	No clustering (Fig.6.25)
SCh_Par_Centr_C2	Parallel, centralised	subchannel	2 Nodes per router (Fig.6.26)
SCh_Par_Distr_C2	Parallel, distributed	subchannel	2 Nodes per router (Fig.6.26)

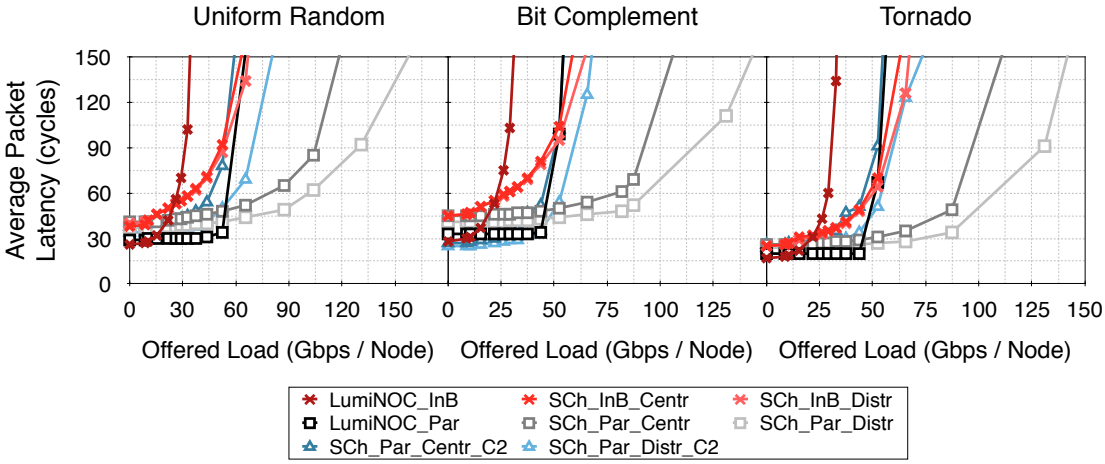
channels per input port, each of which 5 flits deep with a flit size of 64 bits [LBGP14] (in addition to the buffer space of 32 bits for each arbitration packet). We study the NoCs with all the shared bus proposals in this section, which are summarised in Table 6.2. For all buses utilising subchannel scheduling, we chose the configuration with the highest throughput, which is when the number of nodes connected to the bus equals the number of subchannels (as revealed in the previous sections).
We study NoCs with 64 (8×8) and 256 (16×16) to evaluate scalability. For these NoC sizes, buses in the rows and columns connect 8 and 16 nodes for the topology in Figure 6.25, respectively. In the topology in Figure 6.26 that performs clustering, column buses connect 4/8 nodes and row buses connect 8/16 nodes for 64/256 nodes, respectively.

Performance

Figures 6.27a and 6.27b depict the average packet latency for 64 nodes with 64λ and 128λ bus widths, respectively. NoCs utilising buses with in-band arbitration approaches (in red) saturate significantly earlier compared to buses with parallel arbitration for all traffic patterns and bus widths. Moreover, they even saturate earlier than the NoCs in which clustering is applied (in blue), although the latter implement fewer total buses – further underlining the throughput benefits of performing bus arbitration in parallel.
These trends persist as the NoC is scaled up to 256 nodes (Figures 6.28a and 6.28b), apart from when LumiNOC_Par is compared to subchannel approaches with in-band arbitration, which can sustain higher network loads without requiring a parallel arbitration bus. The throughput and latency benefits of subchannel scheduling, as well as parallel bus arbitration, thus successfully translate to realistic NoC implementations with up to 256 nodes.



(a) 64λ Data Bus Bandwidth



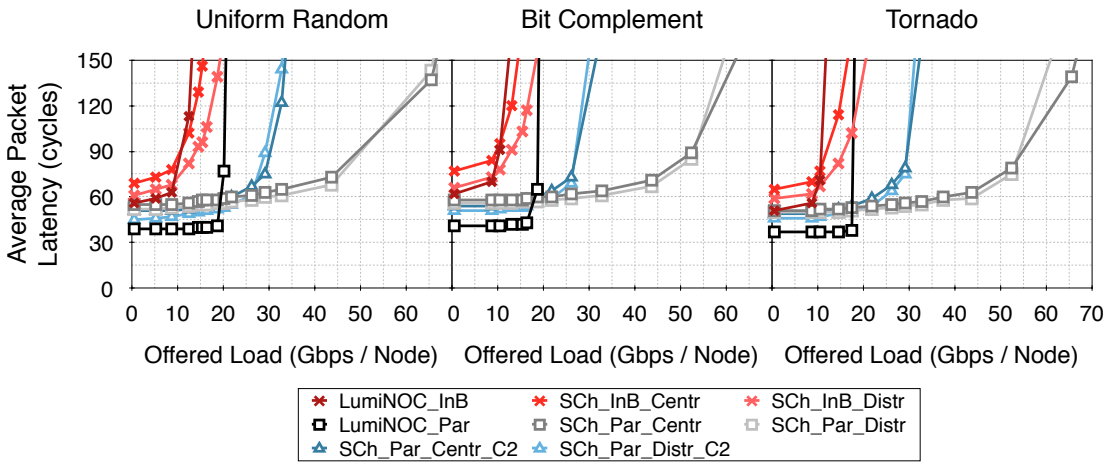
(b) 128λ Data Bus Bandwidth

Figure 6.27: Average Packet Latency for 64 Nodes

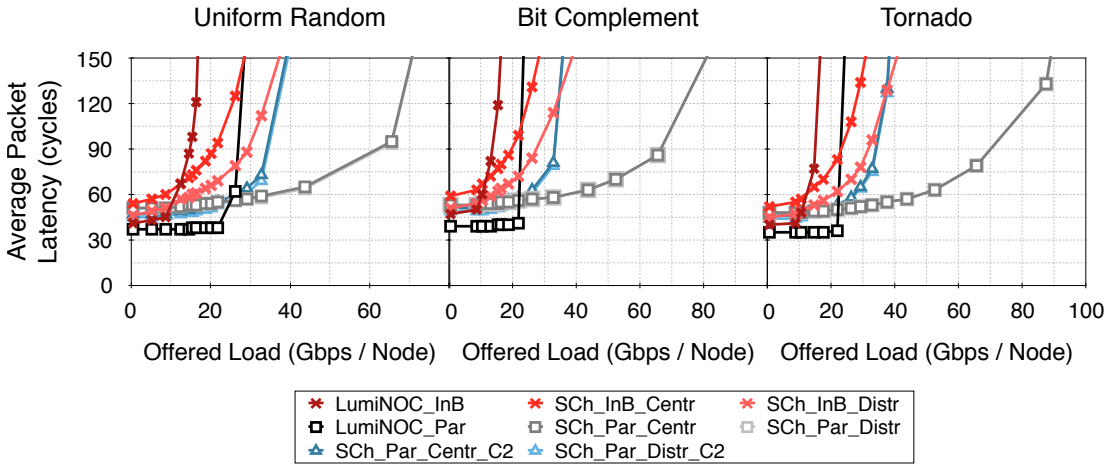
However, the latency overheads compared to LumiNOC for low injection rates translate to the realistic NoC, too. The number of hops through such a NoC is important in terms of latency since the arbitration overheads are imposed at each bus traversal. A topology consisting of subchannel scheduled buses only should thus minimise the average hop count for latency critical applications that exhibit low injection rates. Alternatively, a mechanism switching between LumiNOC and the proposed subchannel schemes could be implemented (as discussed above).

Power

Static power breakdowns of each NoC are shown in Figure 6.29a and Figure 6.29b for 64 and 256 nodes, respectively. Laser and MR heating power dominate the power



(a) 64λ Data Bus Bandwidth



(b) 128λ Data Bus Bandwidth

Figure 6.28: Average Packet Latency for 256 Nodes

budgets of all NoCs. At 256 nodes, most of the power is consumed by the laser source, which is in line with the power results of 16-node shared buses in isolation. In general, the power results of the shared buses in the previous section translate to NoCs implementing them. Core clustering reduces the total static power consumption significantly: for both network sizes, the total static power is more than halved by applying clustering. This is because of two reasons: first, the number of shared buses in the rows is halved compared to the non-clustering case. Second, the number of nodes connected to the buses in the columns is halved, too, which more than halves the power consumed on a shared bus due to significant reductions in both MR-through loss and MR heating. Dynamic power plays a minor role in all NoCs since all data communication is performed optically and is thus low-energy. In addition, the utilised topology allows to

reach any node within two hops, i.e. a packet requires at most 3 router traversals and 2 optical link traversals. As discussed earlier, the consumed dynamic power is very similar in the in-band and parallel case since the exchanged arbitration packets do not differ. To put the dynamic power into relation to the static power: our results showed that SCh_InB_Distr with 64λ bus bandwidth consumes 0.256 W dynamic power close to its saturation point, which is just 18% of the total power. Dynamic power plays a minor role in all NoC utilising in-band arbitration, mainly because of the earlier saturation points. Only for SCh_Par_Distr and SCh_Par_Centr, which can sustain the highest network loads, dynamic power becomes significant as the injection rate approaches the saturation points ($\sim$40%). In general, as bus bandwidth and NoC size increases, the contribution of dynamic power decreases and most of the power is consumed at the laser source and for MR heating – even with core clustering.

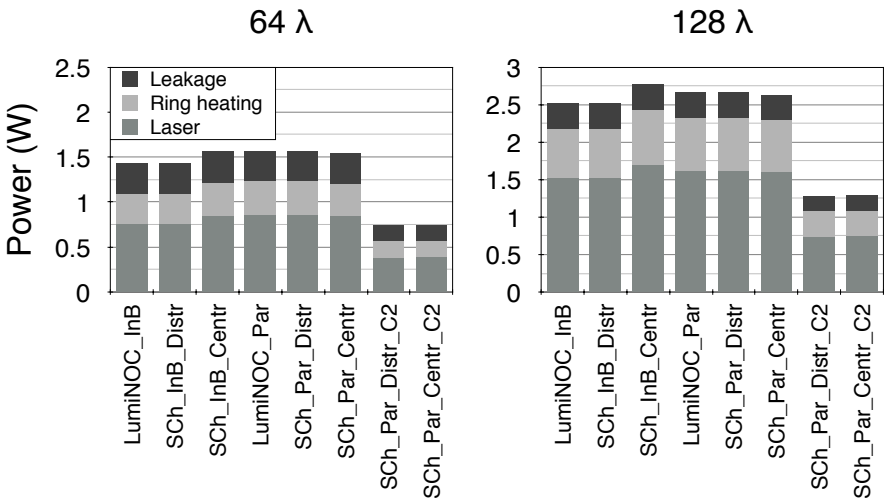
Throughput per Watt

Figures 6.30a and 6.30a present the TPW results for 64 and 128 nodes, respectively. The results confirm the findings in the isolated bus studies: improving the throughput of shared optical buses through subchannel scheduling and parallel bus arbitration improves power efficiency, which then translates to higher order topologies of realistic NoCs. For 64 nodes, TPW is improved by more than 2$\times$ on average by implementing subchannel scheduling with parallel bus arbitration for both 64λ and 128λ. Similar trends can be observed as the network size increases to 256 nodes. Clustering does not noticeably reduce power efficiency compared to the non-clustering NoC for both network sizes, which confirms the assumption that the throughput improvements on the buses can be utilised for both high-throughput NoCs and low-power NoCs. In fact, the large power savings provided by core clustering makes the clustered NoCs with the parallel arbitrated buses the most power-efficient design.

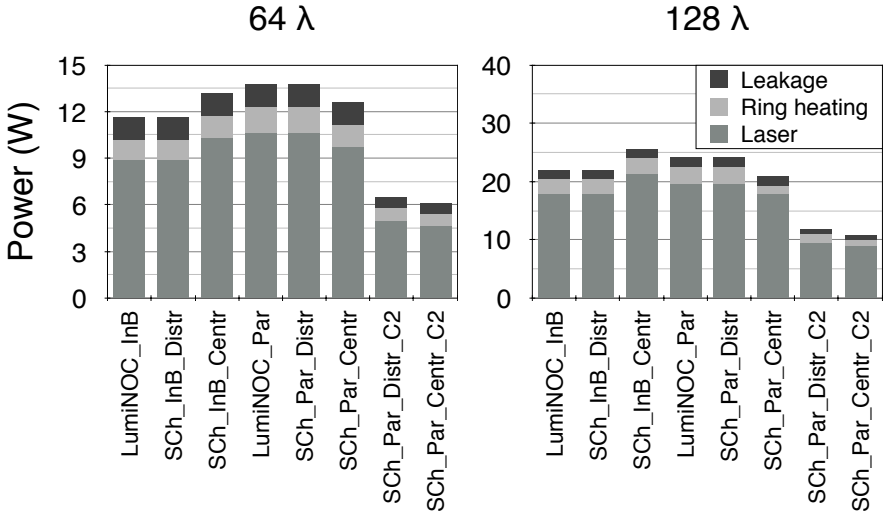
6.6 Summary

This chapter analysed the suitability of shared optical buses as on-chip interconnects, proposed a novel use of optical buses by splitting them into subchannels, leveraged the use of subchannels with innovative bus scheduling and arbitration techniques to improve throughput and power efficiency, and evaluated the trade-offs between in-band arbitration and parallel arbitration on a separate arbitration bus.

MR-through loss is the critical contributor to the path losses and, in turn, laser power,



(a) 64 Nodes

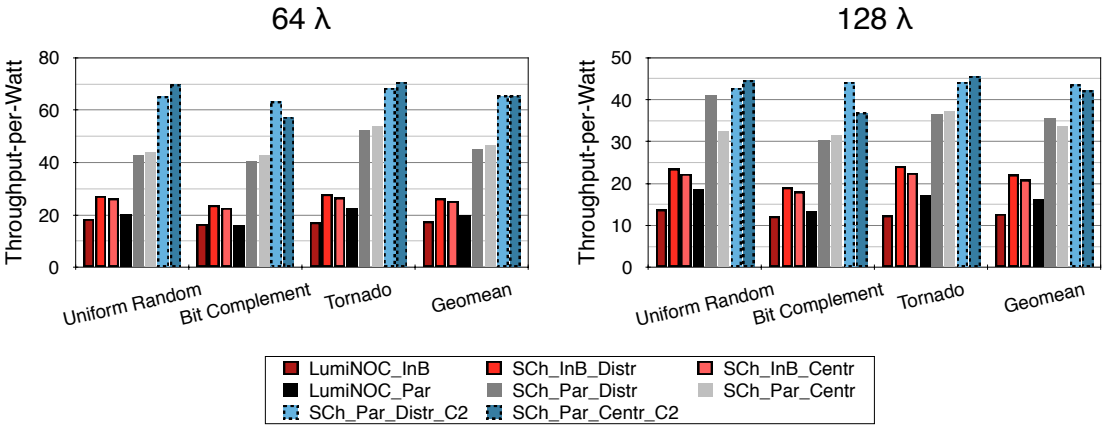


(b) 128 Nodes

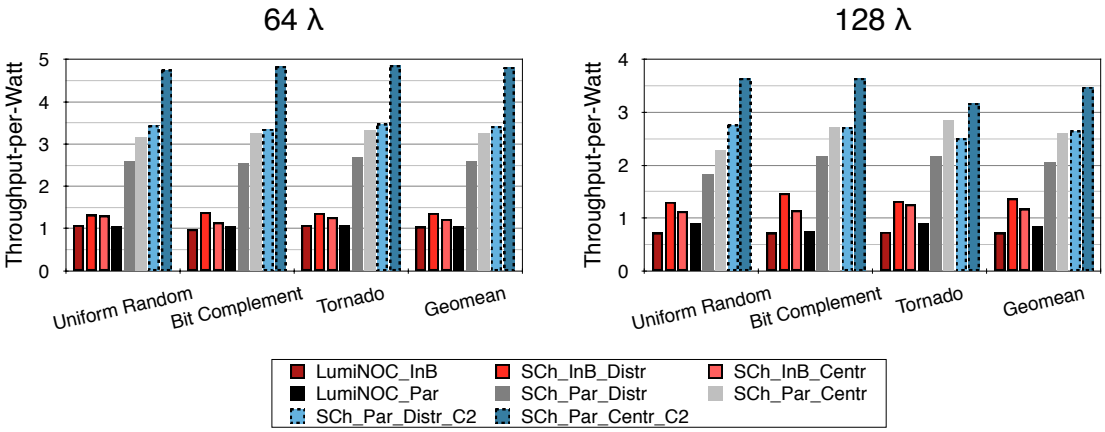
Figure 6.29: Static Power Breakdown

where the latter dominates the power budget of shared optical buses for current SiP technology parameters; however, future device speculations forecasting a decrease of MR-through loss by a factor of 10 would have a large impact on laser power. While already a highly-efficient architecture, shared optical buses would then represent a supreme design choice for on-chip interconnects.

As opposed to the state-of-the-art approach that schedules requesting nodes sequentially on the bus, utilising the possibility of tuning MRs individually allows to schedule requesting nodes both sequentially and in parallel on subchannels. This chapter



(a) 64 Nodes



(b) 128 Nodes

Figure 6.30: Throughput per Watt

illustrated that efficient bus arbitration mechanisms can be designed to implement sub-channel scheduling with low power and latency overheads, while enabling significantly higher bus utilisation and, in turn, throughput.

Implementing a parallel arbitration bus to perform bus arbitration simultaneous to data transmission provides large throughput gains and very little power overheads, which results in much higher power efficiency overall. Moreover, it allows flexible scaling of bandwidth on the arbitration bus, independent from the data bus. Improvements in terms of power efficiency were shown to carry over to higher-order topologies that implement these buses as a backbone. Being a modular building block, the shared buses can then be used to either improve throughput or to lower power by decreasing the total number of buses in the NoC.

Although offering higher power efficiency, subchannel scheduling increases arbitration complexity which leads to latency overheads for low injection rates compared to timeslot-only approaches. This should be kept in mind if buses are considered for implementation in latency-critical NoCs with low traffic demands. Ideally, a mechanism would adapt the scheduling scheme to the current traffic demands and switch between LumiNOC for low, and subchannel scheduling for high injection rates.

Chapter 7

Conclusion

7.1 Introduction

The pace at which SiPs have matured and the opportunities they provide paved the way for an exciting new research field – optical networks-on-chip – that is widely considered the prime candidate to maintain future performance and power scaling of CMPs in the face of the end of Moore's law. To enable a widespread adoption of ONoCs in the near future, research on both the technology and architectural levels is required. This thesis contributes to the architectural level of ONoCs by exploring novel, more efficient ways of integrating optical links into the on-chip communication fabric.

This chapter reviews the primary findings of this thesis (Section 7.2) and discusses the significance of the contributions to the realm of NoCs and CMPs (Section 7.3). Finally, the presented work offers many intriguing opportunities for future studies, which are summarised in Section 7.4.

7.2 Summary of Contributions

Before even considering ONoCs, both currently available SiP devices and electrical interconnects need to be studied and compared (as Chapter 2 has). For a low-voltage 22 nm technology library, utilising optical links is only beneficial for sufficiently large distances due to energy and latency overheads caused by EO/OE conversions; however, if the trend of increasing number of cores and die sizes maintains, optical interconnects will become sooner or later the preferred choice. Until then, completely discarding

electrical interconnects leads to inefficiencies and combining both interconnect technologies in a topology represents the most beneficial approach.

Chapter 4 showed that novel WRONoC topologies can further reduce static optical power by offering fewer path losses and MRs for optical switching. Amon demonstrates this by reducing static optical power by 21% compared to the state of the art. Performance in these NoCs is largely determined by the destination-reservation mechanism, and parallelising control packet exchange to data transmission can halve latency without any power overheads. In order to reduce their susceptibility to traffic hotspots, implementing multiple ejection channels per switch is an efficient solution that can offer 50% latency reduction (compared to just using one ejection channel per node) while incurring less than 0.1% power overheads. Finally, leveraging MR heating to select the injection waveguide dynamically enables to reduce the number of injection channels into the NoC, which in turn decreases power by up to 60%.

Chapter 5 revealed that implementing a higher number of low-bandwidth optical links lowers laser power compared to utilising fewer high-bandwidth links due to lower overall path losses. Based on this finding, 'Lego' was proposed, a novel hybrid NoC topology that supplies the NoC with a higher number of low-bandwidth links paired with an electrical 2D mesh for local traffic so that optical links are only used for larger distances at which serialisation delay can be hidden. This approach offers high bisection bandwidth without large laser power overheads, competitive latency, and reduced dynamic power (more than 50% on realistic workloads). Although imposing area overheads up to 75%, TPW is more than doubled.

Chapter 6 studies the shared optical bus which aims to efficiently utilise the optical bandwidth letting all connected nodes use it in a TDM fashion. MR-through loss is the critical loss parameter for current SiP technologies and advanced devices with lower loss would make the shared bus a supreme candidate for on-chip communication.

Rather than scheduling simultaneously requesting nodes sequentially on the entire bandwidth, scheduling them both in time slots and on subchannels decreases the total bus utilisation time and thus improves throughput. Both centralised and distributed arbitration mechanisms were shown to be suitable to implement subchannel scheduling. Although increasing arbitration complexity, TPW is improved by more than 50% compared to the state-of-the-art sequentially-scheduled mechanism LumiNOC without power overheads. For very low injection rates, however, LumiNOC offer less latency as its arbitration mechanism is less complex and faster.

Performing bus arbitration in-band (i.e. on the same bus as data transmission) is less

power efficient than implementing a separate bus for parallel arbitration because the latter improves throughput by more than 50% while the power overheads of the separate bus are at most 10%. In addition, when implemented in a NoC, the throughput and power efficiency gains achieved by subchannel scheduling carry over to a similar extent when many buses are put together into a larger NoC.

7.3 Concluding Remarks

The low maturity of SiP technologies enforces some limitations to the discussed architectural approaches: current laser and MR heating power requirements limit the bandwidth that can be offered to the NoC. Therefore, on-chip bandwidth should be considered scarce rather than abundant, and designs must always aim to utilise the available bandwidth as efficiently as possible, be it in all-optical or hybrid NoCs, or NoCs consisting of shared buses. As SiP technologies mature, these limitations may be less severe, but such advances are still speculative at this point.

In fact, for the current state of SiPs, it is questionable whether designers will consider all-optical on-chip communication in the near future as electrical interconnects are still superior for short distances. Overheads for EO/OE conversions, laser and MR heating power cannot be fully resolved by NoC architectures alone, and require of advances in laser efficiencies, receiver sensitivities, device losses, athermal MRs, or adaptive laser sources.

Hybrid NoCs are likely the next step to occur to support a NoC with optical links for large distances where electrical links become inefficient. Based on the results in this thesis, the most power-efficient architecture to implement this is the shared optical bus with a parallel arbitration bus, especially with advanced SiP devices offering less MR-through loss. Such designs offer very high bandwidth at low power budgets, which is particularly useful in CMPs with high traffic demands.

The trend of increasing die sizes and number of cores combined with the end of Moore's law will likely speed up the rise of SiPs for on-chip communication and may make its adoption inevitable to maintain performance and power scaling in CMPs in the future. Once manufacturing of SiP devices reaches a point of maturity at which optical links can be reliably integrated into NoCs at large scales, architectures like those presented in this thesis will determine the most efficient design.

7.4 Future Work

The contributions of this thesis provide numerous opportunities for future considerations in the field of optical NoC architectures as different key technologies advance.

7.4.1 Adaptive Laser Sources

Although the architectures presented in this thesis improve the power efficiency of the state of the art, static optical power overheads make electrical NoC favourable for low injection rates, which are common in multi-threaded applications. While significant improvements in SiP devices are necessary to reduce the laser power of a static off-chip laser source to insignificant levels, adaptive on-chip lasers have recently emerged as a very promising alternative. With reported laser switch on/off times no higher than 2 ns [CACP+12], those would allow to proactively switch the laser on only when data transmission occurs and thereby to save most of the laser power (62-92%) for small latency overheads (2-6%) for realistic workloads [DH15b]. While previous work has demonstrated such adaptive laser control mechanisms only for simple topologies or crossbars, the architectures proposed in this thesis offer many opportunities, too.
WRONoCs like Amon are ideal to adopt adaptive lasers possibly without any latency degradation since the latency required for controlling and switching on/off the laser could technically be hidden: a control mechanism could snoop acknowledgement packets on the control network and turn on the laser pro-actively for a certain sender. While this approach would be hardly feasible if just one laser source provides the light to all senders (like in the LPDN introduced in this study), multiple lasers could be coupled into the chip into lower branches of the LPDN so that the senders being turned on/off can be controlled in a more fine-grained fashion. A static laser is then only required for the control network, which requires significantly less laser power than the data network. Although increasing the number of lasers and moving them into the chip, recent studies have shown that this leads to large power savings at the laser source and overall higher energy efficiency [DH15b].
Lego provides similar opportunities as it is based on R-SWMR buses: each sender has to broadcast a control packet to notify the destinations to tune in/out prior to data transmission. A control signal turning on/off the laser source on the data bus could be transmitted in parallel, thereby allowing to hide the laser control delay. Just like in Amon, only the control bus – which requires much less laser power than the data bus – would then need a static laser.

Shared buses could efficiently integrate such adaptive laser mechanisms, too. In-band arbitrated buses, however, require the laser to be always on since both bus arbitration and data transmission is performed on the same bus. If arbitrated in parallel, on the other hand, only the arbitration bus must be provided with light at all times and the data bus can be shut down if unused.
Adaptive lasers could not only be used to shut down the data bus in idle phases, it could also be used to dynamically scale the bandwidth of a bus in the design presented in Chapter 6: since data buses consist of multiple 32λ buses, some data buses could be shut down if current communication demands do not require high bandwidth. This would allow to highly-efficient bandwidth utilisation of the optical links, but also increase the number of lasers since each 32λ bus must be provided with its separate adaptive laser.

7.4.2 Combining Different Architectures

This thesis tackled challenges of different architectural approaches to integrate optical interconnects into NoCs. These discussed designs, however, are by no means mutually exclusive. In fact, they could be combined to mitigate each others' shortcomings.
One obvious drawback of Amon, particularly compared to electrical NoCs, is its latency overheads for local traffic patterns. Imposing the latency on the control network and for EO/OE data conversion just to transmit data to a neighbouring node is inefficient both in terms of latency and power. Therefore, combining Amon with electrical interconnects dedicated for localised traffic – like in Lego – would make sense. In addition, this would allow to take load off the optical NoC, which could be leveraged to reduce the optical resources and in turn power consumption. A study evaluating the extent to which power could be saved and the implications on performance could provide interesting insights.
Combining Lego with subchannel scheduled buses would be ideal since Lego's electrical NoC could take some load off the buses by reducing the number of connected nodes, which would provide large laser power savings due to fewer MR-through losses. In addition, arbitration latency could be further hidden by using the shared bus only for large on-chip distances and the electrical NoC otherwise – as in Lego's distance-based approach. Chapter 6 exposed that subchannel scheduling improves throughput and power efficiency, and that bus arbitration should be performed on a parallel control bus rather than in-band. The results also showed that LumiNOC is actually more efficient

in terms of latency if injection rates are low since scheduling nodes only in time slots and not subchannels enables less arbitration complexity. Ideally, a mechanism would decide dynamically on whether to use sequential or subchannel scheduling based on the current communication demands. Our evaluation results could give a guideline on the injection rate at which it is beneficial to switch from LumiNOC to subchannel scheduling. This mechanism would be especially efficient for realistic application traffic since programs often exhibit computation intensive and communication intensive phases, and would likely result in large performance gains as it would utilise each arbitration and scheduling mechanism when it is the most efficient.

7.4.3 Further Simulation Studies

Although this thesis rigorously evaluated all NoCs under investigation regarding all important figures of merit, additional simulation studies exist that could further explore the suitability of the proposed approaches to different domains and scales.
For instance, Chapter 5 considered Lego only for 64 nodes. Implementing its distance-based approach of combining electrical and optical links for larger scale NoCs could improve scalability significantly, especially because our results suggest that using the electrical mesh with increasing distances provided the highest power efficiency.
Similarly, Chapter 4 identified that Amon does not scale above 64 nodes in a straightforward manner due to excessive laser power; however, studying other scaling techniques, such as core clustering, could evaluate whether Amon can provide sufficient bandwidth for multiple cores connected to an optical switch. For instance, it could be more efficient overall to scale Amon to 64 cores by using a 32-node Amon with a clustering of two. This approach would reduce power consumption and might be sufficient to satisfy performance demands of realistic workloads.
Aside from studying different scales, the workloads could also be varied. Although it has widely been argued that bandwidth demands on the NoC is expected to increase in the future, the SPLASH-2 and PARSEC benchmarks do not confirm this observation. In fact, average injection rates in these workloads are very low. Other application workloads outside the high-performance computing domain, such as server or cloud traces, might have very different NoC usage and could make optical NoCs more favourable.
Although LumiNOC outperforms aggressive 2D mesh baselines on realistic traffic, it would still be interesting to study how NoCs implementing the proposed buses perform under realistic workloads. This is particularly the case for the distributed arbitration approach in which two packet sizes would lead to larger arbitration packets since a

length bitmap field must be appended. Coherence protocols with multicast messages could expose a weakness of subchannel scheduling if one node has to contend for the bus multiple times for small coherence packets, thereby experiencing larger arbitration delays with minimum throughput improvements. However, extending our approach to allow a node to request the bus for multiple packets per arbitration round could provide significant improvements to LumiNOC again since subchannel scheduling could be leveraged to transmit multicast packets in parallel. Allowing multiple packets per sender per arbitration round while keeping control packets small is probably the most challenging task to be tackled by future studies, but might result in large performance benefits.

7.4.4 Optical Interconnects at the Interposer Level

2.5D integrated chips in which multiple dies are placed on an interposer – a large separate die – have gained high interest as it enables heterogeneous integration of different processors/accelerators and DRAM on the same chip [GPAY17]. Interconnecting the different dies placed on the interposer is a novel research field on its own, and a number of recent studies envision this communication to occur optically with SiPs since distances between nodes on an interposer are much larger than within a single die [GPABY16][GPAY17].

Constraints of the interposer interconnects connecting dies through an interposer differ from NoCs for intra-die communication (discussed in this thesis). First, the number of nodes an interposer interconnect must connect is typically much lower (e.g. recent studies consider chips interconnecting 16 dies on the interposer [GPAY17]). Second, area constraints are significantly relaxed: interposers are $\sim$900 mm^2 in area [KJL15][GPAY17] and can be used nearly exclusively by the interconnect.

Applying and evaluating interconnects for 2.5D integrated chips with the architectures proposed in this thesis could represent an interesting future study, particularly as the laser power and MR heating requirements of SiPs do likely not change compared to intra-die NoCs and the benefits of the proposed architectures are maintained. For instance, designs like Lego that have higher area requirements could benefit from the relaxed area constraints of interposers and turn out to be an even more efficient design choice for connecting dies on an interposer than it is for intra-die communication.

Bibliography

[And14] Mark A Anders. High-performance energy-efficient noc fabrics: Evolution and future challenges. In *Eighth IEEE/ACM International Symposium on Networks-on-Chip*, pages i–i. IEEE, 2014.

[BBB+11] Nathan Binkert, Bradford Beckmann, Gabriel Black, Steven K Reinhardt, Ali Saidi, Arkaprava Basu, Joel Hestness, Derek R Hower, Tushar Krishna, Somayeh Sardashti, et al. The gem5 simulator. *ACM SIGARCH Computer Architecture News*, 39(2):1–7, 2011.

[BCB+14] Keren Bergman, Luca P Carloni, Aleksandr Biberman, Johnnie Chan, and Gilbert Hendry. *Photonic network-on-chip design*. Springer, 2014.

[BCM+14] Rajeev Balasubramonian, Jichuan Chang, Troy Manning, Jaime H Moreno, Richard Murphy, Ravi Nair, and Steven Swanson. Near-data processing: Insights from a micro-46 workshop. *IEEE Micro*, 34:36–42, 2014.

[BDHVV+12] Wim Bogaerts, Peter De Heyn, Thomas Van Vaerenbergh, Katrien De Vos, Shankar Kumar Selvaraja, Tom Claes, Pieter Dumon, Peter Bienstman, Dries Van Thourhout, and Roel Baets. Silicon microring resonators. *Laser & Photonics Reviews*, 6(1):47–73, 2012.

[BEA+08] Shane Bell, Bruce Edwards, John Amann, Rich Conlin, Kevin Joyce, Vince Leung, John MacKay, Mike Reif, Liewei Bao, John Brown, et al. Tile64-processor: A 64-core soc with mesh interconnect. In *IEEE International Solid-State Circuits Conference. Digest of Technical Papers.*, pages 88–598. IEEE, 2008.

[BGB+07] Matthieu Briere, Bruno Girodias, Youcef Bouchebaba, Gabriela Nicolescu, Fabien Mieyeville, Frédéric Gaffiot, and Ian O'Connor. System

level assessment of an optical noc in an mpsoc platform. In *Proceedings of the conference on Design, automation and test in Europe*, pages 1084–1089. EDA Consortium, 2007.

[BH07] Luiz André Barroso and Urs Hölzle. The case for energy-proportional computing. *Computer*, 40(12), 2007.

[BIZCK12] Yaniv Ben-Itzhak, Eitan Zahavi, Israel Cidon, and Avinoam Kolodny. Hnocs: modular open-source simulator for heterogeneous nocs. In *International Conference on Embedded Computer Systems*, pages 51–57. IEEE, 2012.

[BJO+09] Christopher Batten, Ajay Joshi, Jason Orcutt, Anatol Khilo, Benjamin Moss, Charles W Holzwarth, Miloš A Popovic, Hanqing Li, Henry I Smith, Judy L Hoyt, et al. Building many-core processor-to-dram networks with monolithic cmos silicon photonics. *IEEE Micro*, 29(4), 2009.

[BKSL08] Christian Bienia, Sanjeev Kumar, Jaswinder Pal Singh, and Kai Li. The parsec benchmark suite: characterization and architectural implications. In *Proceedings of the 17th international conference on Parallel architectures and compilation techniques*, pages 72–81. ACM, 2008.

[Bor13] Shekhar Borkar. Exascale computing–a fact or affliction. *Keynote presentation at IPDPS'13*, 2013.

[BP14] Shirish Bahirat and Sudeep Pasricha. Meteor: hybrid photonic ring-mesh network-on-chip for multicore architectures. *Transactions on Embedded Computing Systems*, 13:116, 2014.

[BPH+11] Aleksandr Biberman, Kyle Preston, Gilbert Hendry, Nicolás Sherwood-Droz, Johnnie Chan, Jacob S Levy, Michal Lipson, and Keren Bergman. Photonic network-on-chip architectures using multilayer deposited silicon materials for high-performance chip multiprocessors. *Journal on Emerging Technologies in Computing Systems*, 7:7, 2011.

[BRBS16] Anja Von Beuningen, Luca Ramini, Davide Bertozzi, and Ulf Schlichtmann. Proton+: a placement and routing tool for 3d optical networks-on-chip with a single optical layer. *ACM Journal on Emerging Technologies in Computing Systems*, 12(4):44, 2016.

[BRNB16] Meisam Bahadori, Sébastien Rumley, Dessislava Nikolova, and Keren Bergman. Comprehensive design space exploration of silicon photonic interconnects. *Journal of Lightwave Technology*, 34(12):2975–2987, 2016.

[BRSB13] Anja Boos, Luca Ramini, Ulf Schlichtmann, and Davide Bertozzi. Proton: An automatic place-and-route tool for optical networks-on-chip. In *Proceedings of the International Conference on Computer-Aided Design*, pages 138–145. IEEE, 2013.

[BS11] Wim Bogaerts and SK Selvaraja. Compact single-mode silicon hybrid rib/strip waveguide with adiabatic bends. *IEEE Photonics Journal*, 3(3):422–432, 2011.

[BSK+10] Scott Beamer, Chen Sun, Yong-Jin Kwon, Ajay Joshi, Christopher Batten, Vladimir Stojanović, and Krste Asanović. Re-architecting dram memory systems with monolithically integrated silicon photonics. In *ACM SIGARCH Computer Architecture News*, volume 38, pages 129–140. ACM, 2010.

[BSP+16] Brent Bohnenstiehl, Aaron Stillmaker, Jon Pimentel, Timothy Andreas, Bin Liu, Anh Tran, Emmanuel Adeagbo, and Bevan Baas. A 5.8 pj/op 115 billion ops/sec, to 1.78 trillion ops/sec 32nm 1000-processor array. In *IEEE Symposium on VLSI Circuits*, pages 1–2. IEEE, 2016.

[CACP+12] Rodolfo E Camacho-Aguilera, Yan Cai, Neil Patel, Jonathan T Bessette, Marco Romagnoli, Lionel C Kimerling, and Jurgen Michel. An electrically pumped germanium laser. *Optics express*, 20(10):11316–11320, 2012.

[CAJ15] Chao Chen, José L Abellán, and Ajay Joshi. Managing laser power in silicon-photonic noc through cache and noc reconfiguration. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34(6):972–985, 2015.

[CCH+07] Guoqing Chen, Hui Chen, Mikhail Haurylau, Nicholas A Nelson, David H Albonesi, Philippe M Fauchet, and Eby G Friedman. Predictions of cmos compatible on-chip optical interconnect. *Integration, the VLSI journal*, 40(4):434–446, 2007.

[CHB+10] Johnnie Chan, Gilbert Hendry, Aleksandr Biberman, Keren Bergman, and Luca P Carloni. Phoenixsim: A simulator for physical-layer analysis of chip-scale photonic interconnection networks. In *Proceedings of the Conference on Design, Automation and Test in Europe*, pages 691–696. European Design and Automation Association, 2010.

[CHTB11] L Chen, E Hall, L Theogarajan, and J Bowers. Photonic switching for data center applications. *IEEE Photonics journal*, 3(5):834–844, 2011.

[CKA09] Mark J Cianchetti, Joseph C Kerekes, and David H Albonesi. Phastlane: a rapid transit optical routing network. *ACM SIGARCH Computer Architecture News*, 37:441–450, 2009.

[Com14] International Roadmap Commitee. International technology roadmap for semiconductors. `www.itrs2.net`, 2014. [Online; accessed 20-08-2017].

[CZC+14] Chao Chen, Tiansheng Zhang, Pietro Contu, Jonathan Klamkin, Ayse K Coskun, and Ajay Joshi. Sharing and placement of on-chip laser sources in silicon-photonic nocs. In *Eighth IEEE/ACM International Symposium on Networks-on-Chip*, pages 88–95. IEEE, 2014.

[DH14] Yigit Demir and Nikos Hardavellas. Ecolaser: an adaptive laser control for energy-efficient on-chip photonic interconnects. In *Proceedings of the 2014 international symposium on Low power electronics and design*, pages 3–8. ACM, 2014.

[DH15a] Yigit Demir and Nikos Hardavellas. Parka: Thermally insulated nanophotonic interconnects. In *Ninth IEEE/ACM International Symposium on Networks-on-Chip*, page 1. ACM, 2015.

[DH15b] Yigit Demir and Nikos Hardavellas. Towards energy-efficient photonic interconnects. In *SPIE OPTO*, pages 93680T–93680T. International Society for Optics and Photonics, 2015.

[DH16a] Yigit Demir and Nikos Hardavellas. Energy-proportional photonic interconnects. *ACM Transactions on Architecture and Code Optimization*, 13(4):54, 2016.

[DH16b] Yigit Demir and Nikos Hardavellas. Slac: Stage laser control for a flattened butterfly network. In *IEEE International Symposium on High Performance Computer Architecture*, pages 321–332. IEEE, 2016.

[DLF+09] Po Dong, Shirong Liao, Dazeng Feng, Hong Liang, Dawei Zheng, Roshanak Shafiiha, Cheng-Chih Kung, Wei Qian, Guoliang Li, Xuezhe Zheng, et al. Low v pp, ultralow-energy, compact, high-speed silicon electro-optic modulator. *Optics express*, 17(25):22484–22490, 2009.

[DNSD13] Reetuparna Das, Satish Narayanasamy, Sudhir K Satpathy, and Ronald G Dreslinski. Catnap: energy proportional multiple network-on-chip. In *ACM SIGARCH Computer Architecture News*, volume 41, pages 320–331. ACM, 2013.

[DPS+14] Yigit Demir, Yan Pan, Seukwoo Song, Nikos Hardavellas, John Kim, and Gokhan Memik. Galaxy: A high-performance energy-efficient multi-chip architecture using photonic interconnects. In *Proceedings of the 28th ACM international conference on Supercomputing*, pages 303–312. ACM, 2014.

[FPC+06] Alexander W Fang, Hyundai Park, Oded Cohen, Richard Jones, Mario J Paniccia, and John E Bowers. Electrically pumped hybrid algainas-silicon evanescent laser. *Optics express*, 14(20):9203–9210, 2006.

[FSB+15] Shaoqi Feng, Kuanping Shang, Jock T Bovington, Rui Wu, Binbin Guan, Kwang-Ting Cheng, John E Bowers, and SJ Ben Yoo. Athermal silicon ring resonators clad with titanium dioxide for 1.3 μm wavelength operation. *Optics express*, 23(20):25653–25660, 2015.

[GCL13] Biswajeet Guha, Jaime Cardenas, and Michal Lipson. Athermal silicon microring resonators with titanium oxide cladding. *Optics express*, 21(22):26557–26563, 2013.

[GK10] Paul Gratz and Stephen W Keckler. Realistic workload characterization and analysis for networks-on-chip design. In *The 4th workshop on chip multiprocessor memory systems and interconnects*, pages 1–10, 2010.

[GLM+11] Michael Georgas, Jonathan Leu, Benjamin Moss, Chen Sun, and Vladimir Stojanović. Addressing link-level design tradeoffs for integrated photonic interconnects. In *Custom Integrated Circuits Conference*, pages 1–8. IEEE, 2011.

[GMS+14] Michael Georgas, BR Moss, Chen Sun, Jeffrey Shainline, JS Orcutt, Mark Wade, Y-H Chen, Kareem Nammari, JC Leu, A Srinivasan, et al. A monolithically-integrated optical transmitter and receiver in a zero-change 45nm soi process. In *Symposium on VLSI Circuits Digest of Technical Papers*, pages 1–2. IEEE, 2014.

[GPABY16] Paolo Grani, Roberto Proietti, Venkatesh Akella, and SJ Ben Yoo. Photonic interconnects for interposer-based 2.5 d/3d integrated systems on a chip. In *Proceedings of the Second International Symposium on Memory Systems*, pages 377–386. ACM, 2016.

[GPAY17] Paolo Grani, Roberto Proietti, Venkatesh Akella, and SJ Ben Yoo. Design and evaluation of awgr-based photonic noc architectures for 2.5 d integrated high performance computing systems. In *IEEE International Symposium on High Performance Computer Architecture*, pages 289–300. IEEE, 2017.

[GTER11] FY Gardes, DJ Thomson, NG Emerson, and GT Reed. 40 gb/s silicon photonics modulator for te and tm polarisations. *Optics express*, 19(12):11804–11814, 2011.

[HB14] Martijn JR Heck and John E Bowers. Energy efficient and energy proportional optical interconnects for multi-core processors: Driving the need for on-chip sources. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(4):1–12, 2014.

[HCC+06] Mikhail Haurylau, Guoqing Chen, Hui Chen, Jidong Zhang, Nicholas A Nelson, David H Albonesi, Eby G Friedman, and

Philippe M Fauchet. On-chip optical interconnect roadmap: Challenges and critical directions. *IEEE Journal of Selected Topics in Quantum Electronics*, 12(6):1699–1705, 2006.

[HDV+11] Jason Howard, Saurabh Dighe, Sriram R Vangal, Gregory Ruhl, Nitin Borkar, Shailendra Jain, Vasantha Erraguntla, Michael Konow, Michael Riepen, Matthias Gries, et al. A 48-core ia-32 processor in 45 nm cmos using on-die message-passing and dvfs for performance and power scaling. *IEEE Journal of Solid-State Circuits*, 46(1):173–183, 2011.

[HGK10] Joel Hestness, Boris Grot, and Stephen W Keckler. Netrace: dependency-driven trace-based network-on-chip simulation. In *Proceedings of the Third International Workshop on Network on Chip Architectures*, pages 31–36. ACM, 2010.

[HJH14] Parisa Khadem Hamedani, Natalie Enright Jerger, and Shaahin Hessabi. Qut: A low-power optical network-on-chip. In *Eighth IEEE/ACM International Symposium on Networks-on-Chip*, pages 80–87. IEEE, 2014.

[Ho06] R Ho. Wire scaling and trends," a presentation at mto darpa meeting. *Sun Microsystems Laboratories, Jackson Hole, WY*, 2006.

[JBK+09] Ajay Joshi, Christopher Batten, Yong-Jin Kwon, Scott Beamer, Imran Shamim, Krste Asanovic, and Vladimir Stojanovic. Silicon-photonic clos networks for global on-chip communication. In *3rd ACM/IEEE International Symposium on Networks-on-Chip*, pages 124–133. IEEE Computer Society, 2009.

[JP09] Natalie Enright Jerger and Li-Shiuan Peh. On-chip networks. *Synthesis Lectures on Computer Architecture*, 4(1):1–141, 2009.

[KAH11] Somayyeh Koohi, Meisam Abdollahi, and Shaahin Hessabi. All-optical wavelength-routed noc based on a novel hierarchical topology. In *Proceedings of the Fifth ACM/IEEE International Symposium on Networks-on-Chip*, pages 97–104. ACM, 2011.

[KC11] Yu-Hsiang Kao and H Jonathan Chao. Blocon: A bufferless photonic clos network-on-chip architecture. In *Fifth IEEE/ACM International Symposium on Networks on Chip*, pages 81–88. IEEE, 2011.

[KH12] Somayyeh Koohi and Shaahin Hessabi. Scalable architecture for a contention-free optical network on-chip. *Journal of Parallel and Distributed Computing*, 72(11):1493–1506, 2012.

[KII+03] S Kamei, M Ishii, M Itoh, T Shibata, Y Inoue, and T Kitagawa. 64× 64-channel uniform-loss and cyclic-frequency arrayed-waveguide grating router module. *Electronics Letters*, 39(1):83–84, 2003.

[KJL15] Ajaykumar Kannan, Natalie Enright Jerger, and Gabriel H Loh. Enabling interposer-based disintegration of multi-core processors. In *48th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*, pages 546–558. IEEE, 2015.

[KK16] Matthew Kennedy and Avinash Karanth Kodi. Clap-net: Bandwidth adaptive optical crossbar architecture. *Journal of Parallel and Distributed Computing*, 2016.

[KMP+10] George Kurian, Jason E Miller, James Psota, Jonathan Eastep, Jifeng Liu, Jurgen Michel, Lionel C Kimerling, and Anant Agarwal. Atac: a 1000-core cache-coherent processor with on-chip optical network. In *Proceedings of the 19th international conference on Parallel architectures and compilation techniques*, pages 477–488. ACM, 2010.

[KNK+13] Brian R Koch, Erik J Norberg, Byungchae Kim, John Hutchinson, Jae-Hyuk Shin, Gregory Fish, and Alexander Fang. Integrated silicon photonic laser sources for telecom and datacom. In *National Fiber Optic Engineers Conference*, pages PDP5C–8. Optical Society of America, 2013.

[KSC+12] George Kurian, Chen Sun, Chia-Hsin Owen Chen, Jason E Miller, Jurgen Michel, Lan Wei, Dimitri A Antoniadis, Li-Shiuan Peh, Lionel

Kimerling, Vladimir Stojanovic, and Anant Agarwal. Cross-layer energy and performance evaluation of a nanophotonic manycore processor system using real application workloads. In *26th International Parallel & Distributed Processing Symposium*, pages 1117–1130. IEEE, 2012.

[KSKK15] Elena Kakoulli, Vassos Soteriou, Charalambos Koutsides, and Kyriacos Kalli. Design of high-performance, power-efficient optical nocs using silica-embedded silicon nanophotonics. In *33rd IEEE International Conference on Computer Design*, pages 1–8. IEEE, 2015.

[LBCB10] Benjamin G Lee, Aleksandr Biberman, Johnnie Chan, and Keren Bergman. High-performance modulators and switches for silicon photonic networks-on-chip. *IEEE Journal of Selected Topics in Quantum Electronics*, 16(1):6–22, 2010.

[LBGP14] Cheng Li, Mark Browning, Paul V Gratz, and Samuel Palermo. Luminoc: A power-efficient, high-performance, photonic network-on-chip. *Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 33(6):826–838, 2014.

[LBLO+14] Sébastien Le Beux, Hui Li, Ian O'Connor, Kazem Cheshmi, Xuchen Liu, Jelena Trajkovic, and Gabriela Nicolescu. Chameleon: Channel efficient optical network-on-chip. In *Proceedings of the conference on Design, Automation & Test in Europe*, page 304. European Design and Automation Association, 2014.

[LBTO+11] Sébastien Le Beux, Jelena Trajkovic, Ian O'Connor, Gabriela Nicolescu, Guy Bois, and Pierre Paulin. Optical ring network-on-chip (ornoc): Architecture and design methodology. In *Design, Automation & Test in Europe Conference & Exhibition*, pages 1–6. IEEE, 2011.

[LLK+10] Cedric F Lam, Hong Liu, Bikash Koley, Xiaoxue Zhao, Valey Kamalov, and Vijay Gill. Fiber optic communication technologies: What's needed for datacenter network operations. *IEEE Communications Magazine*, 48(7), 2010.

[LLR+07] Liu Liao, A Liu, D Rubin, J.A.B.J. Basak, Y.A.C.Y. Chetrit, H.A.N.H. Nguyen, R.A.C.R. Cohen, N.A.I.N. Izhaky, and M.A.P.M. Paniccia.

40 gbit/s silicon optical modulator for high-speed applications. *Electronics letters*, 43(22):1196–1197, 2007.

[LNP+13] Junghee Lee, Chrysostomos Nicopoulos, Sung Joo Park, Madhavan Swaminathan, and Jongman Kim. Do we need wide flits in networks-on-chip? In *Computer Society Annual Symposium on VLSI*, pages 2–7. IEEE, 2013.

[LQJ+15] Zhongqi Li, Amer Qouneh, Madhura Joshi, Wangyuan Zhang, Xin Fu, and Tao Li. Aurora: A cross-layer solution for thermally resilient photonic network-on-chip. *IEEE Transactions on VLSI Systems*, 23(1):170–183, 2015.

[LSCA+10] Jifeng Liu, Xiaochen Sun, Rodolfo Camacho-Aguilera, Lionel C Kimerling, and Jurgen Michel. Ge-on-si laser operating at room temperature. *Optics letters*, 35(5):679–681, 2010.

[LSZP14] Yangyang Liu, Jeffrey M Shainline, Xiaoge Zeng, and Miloš A Popović. Ultra-low-loss cmos-compatible waveguide crossing arrays based on multimode bloch waves and imaginary coupling. *Optics letters*, 39(2):335–338, 2014.

[LYM15] Jiwei Liu, Jun Yang, and Rami Melhem. Gasolin: global arbitration for streams of data in optical links. In *International Parallel and Distributed Processing Symposium*, pages 93–102. IEEE, 2015.

[LZT+12] Guoliang Li, Xuezhe Zheng, Hiren Thacker, Jin Yao, Ying Luo, Ivan Shubin, Kannan Raj, John E Cunningham, and Ashok V Krishnamoorthy. 40 gb/s thermally tunable cmos ring modulator. In *IEEE 9th International Conference on Group IV Photonics*, pages 1–3. IEEE, 2012.

[LZY+11] Guoliang Li, Xuezhe Zheng, Jin Yao, Hiren Thacker, Ivan Shubin, Ying Luo, Kannan Raj, John E Cunningham, and Ashok V Krishnamoorthy. 25gb/s 1v-driving cmos ring modulator with integrated thermal tuning. *Optics Express*, 19(21):20435–20443, 2011.

[MK10] Randy Morris and Avinash Karanth Kodi. Exploring the design of 64- and 256-core power efficient nanophotonic interconnect. *IEEE Journal of Selected Topics in Quantum Electronics*, 16:1386–1393, 2010.

[MKK+10] Jason E Miller, Harshad Kasture, George Kurian, Charles Gruenwald, Nathan Beckmann, Christopher Celio, Jonathan Eastep, and Anant Agarwal. Graphite: A distributed parallel simulator for multicores. In *16th International Symposium on High Performance Computer Architecture*, pages 1–12. IEEE, 2010.

[MKL12] Randy Morris, Avinash Karanth Kodi, and Ahmed Louri. Dynamic reconfiguration of 3d photonic networks-on-chip for maximizing performance and improving fault tolerance. In *Proceedings of the 2012 45th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 282–293. IEEE Computer Society, 2012.

[MM09] Thomas Moscibroda and Onur Mutlu. A case for bufferless routing in on-chip networks. In *ACM SIGARCH Computer Architecture News*, volume 37, pages 196–207. ACM, 2009.

[MNM+12] Gianlorenzo Masini, Adit Narasimha, Attila Mekis, Brian Welch, Carl Ogden, Colin Bradbury, Chang Sohn, Dan Song, Dany Martinez, Dennis Foltz, et al. Cmos photonics for optical engines and interconnects. In *Optical Fiber Communication Conference and Exposition, and the National Fiber Optic Engineers Conference*, pages 1–3. IEEE, 2012.

[MYH+14] Xiang Ma, Jiyang Yu, Xingcheng Hua, Chao Wei, Yi Huang, Longzhi Yang, Defeng Li, Qinfen Hao, Peng Liu, Xiaoqing Jiang, et al. Lioesim: a network simulator for hybrid opto-electronic networks-on-chip analysis. *Journal of Lightwave Technology*, 32(22):3699–3708, 2014.

[MYW+10] Kwai Hung Mo, Yaoyao Ye, Xiaowen Wu, Wei Zhang, Weichen Liu, and Jiang Xu. A hierarchical hybrid optical-electronic network-on-chip. In *2010 IEEE Computer Society Annual Symposium on VLSI*, pages 327–332. IEEE, 2010.

[NFA11] Christopher Nitta, Matthew Farrens, and Venkatesh Akella. Addressing system-level trimming issues in on-chip nanophotonic networks. In *17th International Symposium on High Performance Computer Architecture*, pages 122–131. IEEE, 2011.

[OMS+12] Jason S Orcutt, Benjamin Moss, Chen Sun, Jonathan Leu, Michael Georgas, Jeffrey Shainline, Eugen Zgraggen, Hanqing Li, Jie Sun, Matthew Weaver, et al. Open foundry platform for high-performance electronic-photonic integration. *Optics express*, 20(11):12222–12232, 2012.

[ON12] Ian O'Connor and Gabriela Nicolescu. *Integrated optical interconnect architectures for embedded systems*. Springer Science & Business Media, 2012.

[OORVYB16] Marta Ortín-Obón, Luca Ramini, Víctor Viñals-Yúfera, and Davide Bertozzi. A tool for synthesizing power-efficient and custom-tailored wavelength-routed optical rings. In *2nd Asia and South Pacific Design Automation Conference*. IEEE, 2016.

[OOTR+17] Marta Ortín-Obón, Mahdi Tala, Luca Ramini, Víctor Viñals-Yufera, and Davide Bertozzi. Contrasting laser power requirements of wavelength-routed optical noc topologies subject to the floorplanning, placement, and routing constraints of a 3-d-stacked system. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2017.

[PB06] M Paniccia and J Bowers. First electrically pumped hybrid pumped hybrid silicon laser silicon laser. `https://www.intel.com/content/dam/www/public/us/en/documents/technology-briefs/intel-labs-hybrid-silicon-laser-announcement.pdf`, 2006. [Online; accessed 20-08-2017].

[PD10] Sudeep Pasricha and Nikil Dutt. *On-chip communication architectures: system on chip interconnect*. Morgan Kaufmann, 2010.

[PDB14] Ritesh Parikh, Reetuparna Das, and Valeria Bertacco. Power-aware nocs through routing and topology reconfiguration. In *51st ACM/EDAC/IEEE Design Automation Conference*, pages 1–6. IEEE, 2014.

[PKC+12] Sunghyun Park, Tushar Krishna, Chia-Hsin Chen, Bhavya Daya, Anantha Chandrakasan, and Li-Shiuan Peh. Approaching the theoretical limits of a mesh noc with a 16-node chip prototype in 45nm soi. In

Proceedings of the 49th Annual Design Automation Conference, pages 398–405. ACM, 2012.

[PKD+10] Petar Pepeljugoski, Jeffrey Kash, Fuad Doany, Daniel Kuchta, Laurent Schares, Clint Schow, Marc Taubenblatt, Bert Jan Offrein, and Alan Benner. Low power and high density optical interconnects for future supercomputers. In *Optical Fiber Communication (OFC), collocated National Fiber Optic Engineers Conference, 2010 Conference on (OFC/NFOEC)*, pages 1–3. IEEE, 2010.

[PKK+09] Yan Pan, Prabhat Kumar, John Kim, Gokhan Memik, Yu Zhang, and Alok Choudhary. Firefly: illuminating future network-on-chip with nanophotonics. In *ACM SIGARCH Computer Architecture News*, volume 37, pages 429–440. ACM, 2009.

[PKM10] Yan Pan, John Kim, and Gokhan Memik. Flexishare: Channel sharing for an energy-efficient nanophotonic crossbar. In *6th International Symposium on High Performance Computer Architecture*, pages 1–12. IEEE, 2010.

[PKM11] Yan Pan, John Kim, and Gokhan Memik. Featherweight: low-cost optical arbitration with qos support. In *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 105–116. ACM, 2011.

[Pra10] Subodh Prabhu. *OCIN_TSIM-a DVFS aware simulator for NoC design space exploration and optimization*. PhD thesis, Texas A & M University, 2010.

[PRG+16] Andrea Peano, Luca Ramini, Marco Gavanelli, Maddalena Nonato, and Davide Bertozzi. Design technology for fault-free and maximally-parallel wavelength-routed optical networks-on-chip. In *IEEE/ACM International Conference on Computer-Aided Design*. IEEE, 2016.

[PSDLL11] Kyle Preston, Nicolas Sherwood-Droz, Jacob S Levy, and Michal Lipson. Performance guidelines for wdm interconnects based on silicon microring resonators. In *CLEO: Science and Innovations*. Optical Society of America, 2011.

[PTDS15] Eldhose Peter, Arun Thomas, Anuj Dhawan, and Smruti R Sarangi. Coldbus: A near-optimal power efficient optical bus. In *22nd International Conference on High Performance Computing*, pages 275–284. IEEE, 2015.

[PTDS16] Eldhose Peter, Arun Thomas, Anuj Dhawan, and Smruti R Sarangi. Active microring based tunable optical power splitters. *Optics Communications*, 359:311–315, 2016.

[Ram11] Carl Ramey. Tile-gx100 manycore processor: Acceleration interfaces and architecture. In *Hot Chips 23 Symposium*, 2011.

[RBP+17] Sébastien Rumley, Meisam Bahadori, Robert Polster, Simon D Hammond, David M Calhoun, Ke Wen, Arun Rodrigues, and Keren Bergman. Optical interconnects for extreme scale computing systems. *Parallel Computing*, 64:65–80, 2017.

[RBW+16] Sébastien Rumley, Meisam Bahadori, Ke Wen, Dessislava Nikolova, and Keren Bergman. Phoenixsim: Crosslayer design and modeling of silicon photonic interconnects. In *Proceedings of the 1st International Workshop on Advanced Interconnect Solutions and Technologies for Emerging Computing Systems*, page 7. ACM, 2016.

[RGBB13] Luca Ramini, Paolo Grani, Sandro Bartolini, and Davide Bertozzi. Contrasting wavelength-routed optical noc topologies for power-efficient 3d-stacked multicore processors using physical-layer analysis. In *Design, Automation & Test in Europe Conference & Exhibition*, pages 1589–1594. EDA Consortium, 2013.

[RGF+14] Luca Ramini, Paolo Grani, Hervé Tatenguem Fankem, Alberto Ghiribaldi, Sandro Bartolini, and Davide Bertozzi. Assessing the energy break-even point between an optical noc architecture and an aggressive electronic baseline. In *Design, Automation and Test in Europe Conference and Exhibition (DATE)*, pages 1–6. IEEE, 2014.

[RJS15] Cezar RW Reinbrecht, Martha Johanna, and Altamiro Amadeu Susin. Phicit: Improving hierarchical networks-on-chip through 3d silicon photonics integration. In *Proceedings of the 28th Symposium on Integrated Circuits and Systems Design*, page 28. ACM, 2015.

[RTB14] Luca Ramini, Mahdi Tala, and Davide Bertozzi. Exploring communication protocols for optical networks-on-chip based on ring topologies. In *Asia Communications and Photonics Conference*, pages ATh3A–165. Optical Society of America, 2014.

[S+15] Chen Sun et al. *Silicon-photonics for VLSI systems*. PhD thesis, Massachusetts Institute of Technology, 2015.

[SBC08] Assaf Shacham, Keren Bergman, and Luca P Carloni. Photonic networks-on-chip for future generations of chip multiprocessors. *Transactions on Computers*, 57:1246–1260, 2008.

[SCK+12] Chen Sun, Chia-Hsin Owen Chen, George Kurian, Lan Wei, Jason Miller, Anant Agarwal, Li-Shiuan Peh, and Vladimir Stojanovic. Dsent-a tool connecting emerging photonics with electronics for opto-electronic networks-on-chip modeling. In *Sixth IEEE/ACM International Symposium on Networks on Chip*, pages 201–210. IEEE, 2012.

[SZZ+14] Thomas E Sarvey, Yang Zhang, Yue Zhang, Hanju Oh, and Muhannad S Bakir. Thermal and electrical effects of staggered micropin-fin dimensions for cooling of 3d microsystems. In *Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, pages 205–212. IEEE, 2014.

[TLY+16] Jian Tang, Ming Li, Ye Yang, Shuqian Sun, Nuannuan Shi, Wei Li, and Ninghua Zhu. High-speed tunable broadband microwave photonics phase shifter based on an active microring resonator. In *25th Wireless and Optical Communication Conference*, pages 1–3. IEEE, 2016.

[TZ14] Yvain Thonnart and Mounir Zid. Technology assessment of silicon interposers for manycore socs: Active, passive, or optical? In *Eighth IEEE/ACM International Symposium on Networks-on-Chip*, pages 168–169. IEEE, 2014.

[UMC+10] Aniruddha N Udipi, Naveen Muralimanohar, Niladrish Chatterjee, Rajeev Balasubramonian, Al Davis, and Norman P Jouppi. Rethinking dram design and organization for energy-constrained multi-cores. In *ACM SIGARCH Computer Architecture News*, volume 38, pages 175–186. ACM, 2010.

[Van10] Dana M Vantrease. *Optical tokens in many-core processors*. PhD thesis, UNIVERSITY OF WISCONSIN–MADISON, 2010.

[Var99] András Varga. Using the omnet++ discrete event simulation system in education. *IEEE Transactions on Education*, 42(4):11–pp, 1999.

[VBSL09] Dana Vantrease, Nathan Binkert, Robert Schreiber, and Mikko H Lipasti. Light speed arbitration and flow control for nanophotonic interconnects. In *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture*, pages 304–315. ACM, 2009.

[VHR+07] Sriram Vangal, Jason Howard, Gregory Ruhl, Saurabh Dighe, Howard Wilson, James Tschanz, David Finan, Priya Iyer, Arvind Singh, Tiju Jacob, Shailendra Jain, Siram Venkataraman, Yatin Hoskote, and Nitin Bordkar. An 80-tile 1.28 tflops network-on-chip in 65nm cmos. In *International Solid-State Circuits Conference. Digest of Technical Papers*, pages 98–99. IEEE, 2007.

[VHR+08] Sriram R Vangal, Jason Howard, Gregory Ruhl, Saurabh Dighe, Howard Wilson, James Tschanz, David Finan, Arvind Singh, Tiju Jacob, Shailendra Jain, et al. An 80-tile sub-100-w teraflops processor in 65-nm cmos. *IEEE Journal of solid-state circuits*, 43(1):29–41, 2008.

[VLJW13] Anouk Van Laer, Timothy Jones, and Philip M Watts. Full system simulation of optically interconnected chip multiprocessors using gem5. In *Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference (OFC/NFOEC), 2013*, pages 1–3. IEEE, 2013.

[VSG+12] Stavros Volos, Ciprian Seiculescu, Boris Grot, Naser Khosro Pour, Babak Falsafi, and Giovanni De Micheli. Ccnoc: Specializing on-chip interconnects for energy efficiency in cache-coherent servers. In *Sixth IEEE/ACM International Symposium on Networks on Chip*, pages 67–74. IEEE, 2012.

[VSM+08] Dana Vantrease, Robert Schreiber, Matteo Monchiero, Moray McLaren, Norman P Jouppi, Marco Fiorentino, Al Davis, Nathan Binkert, Raymond G Beausoleil, and Jung Ho Ahn. Corona: System implications of emerging nanophotonic technology. In *ACM SIGARCH*

Computer Architecture News, volume 36, pages 153–164. IEEE Computer Society, 2008.

[VTL+16] Pascal Vivet, Yvain Thonnart, Romain Lemaire, Cristiano Santos, Edith Beigné, Christian Bernard, Florian Darve, Didier Lattard, Ivan Miro-Panadès, Denis Dutoit, Fabien Clermidy, S. Cheramy, Abbas Sheibanyrad, Frédéric Pétrot, Eric Flamand, Jean Michailos, Alexandre Arriordaz, Lee Wang, and Juergen Schloeffel. A 4 x 4 x 2 homogeneous scalable 3d network-on-chip circuit with 326 mflit/s 0.66 pj/b robust and fault tolerant asynchronous 3d links. *IEEE Journal of Solid-State Circuits*, 2016.

[WGWS15] Yang Wang, Shitao Gao, Ke Wang, and Efstratios Skafidas. Ultra-broadband and low-loss optical power splitter based on tapered silicon waveguides. In *IEEE Optical Interconnects Conference (OI)*, pages 80–81. IEEE, 2015.

[WML+13] Philip Wolf, Philip Moser, Gunter Larisch, Werner Hofmann, Hui Li, James A Lott, Chien-Yao Lu, Shun L Chuang, and Dieter Bimberg. Energy-efficient and temperature-stable high-speed vcsels for optical interconnects. In *15th International Conference on Transparent Optical Networks*, pages 1–5. IEEE, 2013.

[WOT+95] Steven Cameron Woo, Moriyoshi Ohara, Evan Torrie, Jaswinder Pal Singh, and Anoop Gupta. The splash-2 programs: Characterization and methodological considerations. In *ACM SIGARCH Computer Architecture News*, pages 24–36. ACM, 1995.

[WXY+14] Xiaowen Wu, Jiang Xu, Yaoyao Ye, Zhehui Wang, Mahdi Nikdast, and Xuan Wang. Suor: Sectioned undirectional optical ring for chip multiprocessor. *ACM Journal on Emerging Technologies in Computing Systems*, 10(4):29, 2014.

[XYM12] Yi Xu, Jun Yang, and Rami Melhem. Channel borrowing: an energy-efficient nanophotonic crossbar architecture with light-weight arbitration. In *Proceedings of the 26th ACM international conference on Supercomputing*, pages 133–142. ACM, 2012.

[YGS16] SJ Ben Yoo, Binbin Guan, and Ryan P Scott. Heterogeneous 2d/3d photonic integrated microsystems. *Microsystems & Nanoengineering*, 2, 2016.

[YXH+13] Yaoyao Ye, Jiang Xu, Baihan Huang, Xiaowen Wu, Wei Zhang, Xuan Wang, Mahdi Nikdast, Zhehui Wang, Weichen Liu, and Zhe Wang. 3-d mesh-based optical network-on-chip for multiprocessor system-on-chip. *Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32(4):584–596, 2013.

[ZAU+15] Amir Kavyan Kavyan Ziabari, Jose L Abellán, Rafael Ubal, Chao Chen, Ajay Joshi, and David Kaeli. Leveraging silicon-photonic noc for designing scalable gpus. In *Proceedings of the 29th ACM on International Conference on Supercomputing*, pages 273–282. ACM, 2015.

[ZKS+13] Arslan Zulfiqar, Pranay Koka, Herb Schwetman, Mikko Lipasti, Xuezhe Zheng, and Ashok Krishnamoorthy. Wavelength stealing: an opportunistic approach to channel sharing in multi-chip photonic interconnects. In *Proceedings of the 46th Annual IEEE/ACM International Symposium on Microarchitecture*, pages 222–233. ACM, 2013.

[ZL10] Xiang Zhang and Ahmed Louri. A multilayer nanophotonic interconnection network for on-chip many-core communications. In *Proceedings of the 47th Design Automation Conference*, pages 156–161. ACM, 2010.

[ZPL+11] Xuezhe Zheng, Dinesh Patil, Jon Lexau, Frankie Liu, Guoliang Li, Hiren Thacker, Ying Luo, Ivan Shubin, Jieda Li, Jin Yao, et al. Ultra-efficient 10gb/s hybrid integrated silicon photonic transmitter and receiver. *Optics Express*, 19(6):5172–5186, 2011.