ARCHITECTURE AND MAC PROTOCOLS FOR AWG–BASED WDM SINGLE–HOP NETWORKS

by

Chun Fan

Dissertation Presented in Partial Fullfillment
of the Requirement for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY

July 2004

ARCHITECTURE AND MAC PROTOCOLS FOR AWG–BASED WDM SINGLE–HOP NETWORKS

by

Chun Fan

has been approved

July 2004

APPROVED:

_____, Chair

_____

_____

_____

_____

Supervisory Committee

ACCEPTED:

_____

Department Chair

_____

Dean, Graduate College

ABSTRACT

Single–hop WDM networks based on a central Passive Star Coupler (PSC) have received a great deal of attention as promising solutions for metropolitan and local area networks which face increasing amount of traffic. This work explores a relatively new class of single–hop networks based on an Arrayed Waveguide Grating (AWG). The AWG is a wavelength routing device that enables wavelength reuses at all of the input and output ports, which greatly improves the throughput–delay performances of the network using only a limited number of wavelengths. In addition to improving wavelength efficiency, this work also examines reliability and multicasting, which are critical features for todays WDM network. The first proposal, a AWG||PSC network, combines the AWG in parallel with an PSC. The parallel architecture provides a number of protection features for the network. The second proposal, a $FT^{\Lambda} - FR^{\Lambda}$ AWG network, uses arrays of fix–tuned transmitters and fix–tuned receivers at each node, providing particularly efficient multicast performance. The results indicate that these AWG–based networks give significant throughput–delay performances compared to PSC networks.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Page

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Three major levels of networks form the telecommunications and internet infrastructure. Individual users are connected by access networks to the metropolitan area networks (MAN) which are interconnected by long-haul wide area networks. The rapid growth of the internet has pushed the demand for bandwidth at all three levels. Access networks are providing ever increasing amount of bandwidth with advanced gigabit Ethernet LANs, broadband DSL and cable-modem access, and next generation UTMS wireless systems. Wave division multiplexing (WDM) is already widely deployed in the long-haul backbone networks. These very high–speed optical backbone networks are built by interconnecting the point–to–point backbone links with optical add–drop multiplexers and optical cross connects, which are controlled by multi-protocol optical lambda switching, optical burst switching, and optical packet switching mechanisms. These technologies provides abundant capacity to handle the growth of data traffic into the near foreseeable future. MAN which connects the high-speed access networks to the WDM backbone networks typically use SDH/SONET technology which is based on circuit switching. In SDH/SONET networks, the high cell overhead and the optical-electronic-optical (OEO) conversion required at each node adds significant processing burden and complexity. MAN are expected to become the bottleneck in the future.

With the network intelligence increasingly residing at the periphery, many have envisioned IP over WDM to remove data overhead of SDH/SONET and ATM networks. Single-hop WDM network is one favored proposal for future MAN due to their minimum hop distance, high-bandwidth efficiency (no bandwidth is wasted due to packet forwarding), and inherent transparency. These advantages have been explored through numerous proposals for single-hop WDM networks based on the passive-star-coupler (PSC) as the central broadcasting device. The central theme in most of these proposals is the design of media access control (MAC) protocols which maximizes throughput and minimizes delay WDM networks. Although these measures are central to a networks performance, a number of other issues are also critical.

Survivability of the network is one paramount issue. Despite SDH/SONET's large overhead, one of its important features is resilience through its ability to redirect the flow of data under fiber and node failures. Specifically, single–hop network operation is immune from node failures since nodes do not have to forward traffic. But all single–hop networks — either PSC or AWG based — suffer from a *single point of failure*: If the central hub fails the network connectivity is entirely lost due to missing alternate paths. Note that this holds also for all multi-hop networks whose logical topology is embedded on a physical single–hop network. Therefore, protection of (physical) single–hop networks is required to ensure survivability

Today's networks also face increase in high bandwidth and multidestination applications such as video conference, video on demand, image distributions. Multicasting is the simultaneous transmission of information from one source to multiple destinations. It is bandwidth efficient because it eliminates the necessity for the source to send multiple times to each individual receiver which consumes more transmission and wavelength resources.

Thus supporting multicast is an important feature of future WDM single–hop networks.

Wavelength efficiency is another topic that is seldom addressed by many proposed single–hop network proposals. Efficiency means that the WDM network components and the wavelengths are utilized so as to achieve a large network throughput and small network delays. As mentioned before, DWDM provides large bandwidth by exploiting hundreds of wavelengths with large arrays of fixed–tuned transceivers. However, the costly backbone DWDM technologies and the large transceiver arrays required are not attractive for the very cost–sensitive metro area networks. Wavelengths are typically scarce in metro area networks because of (1) the limited number of wavelengths provided by the cost–effective coarse WDM technology, and (2) the limited tuning range of the fast tunable electro–optic transceivers required for dynamic wavelength allocation.

Our single-hop WDM network solutions to these three major issues are based on arrayed-waveguide-grating (AWG) as the central network device. We propose and analyze two different networks. The first network is an AWG||PSC network, which offers protection and wavelength efficiency compared to single PSC networks. The second network is an AWG stand–alone network with nodal transceiver arrays, which is an efficient architecture for multicasting. We present the media access control (MAC) protocols for each of these networks and examine the network performance impact of the added features. The significance of these two networks is to discover and examine the critical features in network architecture, nodal transceiver architecture, and MAC protocols in an effort to improve the reliability and performance of single–hop WDM networks.

CHAPTER 2

Properties of the PSC and the AWG

Single–hop networks come in two flavors: *broadcast* networks and *switched* networks. In the 90's much research has been focused on the design and evaluation of MAC protocols for single–hop WDM networks that are based on a passive star coupler (PSC), see for instance [46]. These networks form broadcast networks in which each wavelength is distributed to all destination nodes. Recently, arrayed-waveguide grating (AWG) based single–hop networks have attracted much interest [4, 32, 41, 51]. By using a wavelength–routing AWG instead of a PSC as central hub each wavelength is not broadcast but routed to a different AWG output port resulting in *switched* single–hop networks. These switched single–hop networks allow each wavelength to be used at all AWG input ports simultaneously without resulting in channel collisions at the AWG output ports. The resulting spatial wavelength reuse dramatically improves the throughput–delay performance of single–hop networks [41].

For this reason, AWG based networks have recently begun to attract significant attention. The AWG is a wavelength routing device which allows for spatial wavelength reuse, i.e., the entire set of wavelengths can be simultaneously applied at each AWG input port without resulting in collisions at the AWG output ports. This spatial wavelength reuse has been demonstrated to significantly improve the network performance for a fixed set of

wavelengths compared to PSC based networks [41].

1. Passive Star Coupler–PSC

The passive star coupler (PSC) is a passive broadcasting device. In an $N \times N$ PSC, a signal coming from any input port is equally divided among the N output ports. The theory and construction of the PSC are detailed in [53, 59]. The PSC can be constructed using $\frac{N}{2} log_2 N 2 \times 2$ couplers.

Star topology networks based on the PSC as the central broadcast device require a lower power budget compared to networks with a linear bus topology or a tree topology. Ignoring the excess loss, the optical input power is equally distributed to the output ports. In the case of a $N \times N$ star coupler the resulting splitting loss equals $10 log_{10} N$(dB).

The broadcast property of the PSC makes it an ideal device for distributing information to all nodes in WDM networks. These advantages have led to numerous proposals for PSC–based broadcast–and–select networks.

The broadcasting feature of the PSC also has its drawbacks. Broadcasting information to unintended nodes may lead to added processing burden for the nodes. In a PSC network, each wavelength can only be used by one input port at a time. A collision occurs if a wavelength is used by more than one input port at the same time, resulting in a corrupted signal. Since each wavelength provides exactly one channel between a source–destination pair, expanding the transmission capacity of a PSC network requires more wavelengths.

2. Arrayed Waveguide Grating–AWG

The arrayed–waveguide grating (AWG) is a passive wavelength routing device. Dragone *et al.* [12, 13] discuss the construction and the properties of the AWG. Several works

[21, 42, 60] discuss the applications of AWG in multiplexing, demultiplexing, add–drop multiplexing, and routing. In the proposed AWG||PSC network, we use the AWG as a router. The crosstalk performance of AWG routers and the feasibility of AWG routers have been studied in [2, 29].

In our networks, we exploit two features of the AWG: (1) wavelength reuse, and (2) periodic wavelength routing in conjunction with utilizing multiple free spectral ranges (FSRs). Wavelength reuse allows the same wavelengths to be used simultaneously at all of the AWG input ports. So, with a $D \times D$ AWG (D input ports and D output ports), each wavelength can be reused D times. Periodic wavelength routing and the utilization of multiple FSRs allow each input–output port pair to be connected by multiple wavelengths. We let R denote the number of utilized FSRs. For $R = 1$, there is only one wavelength connecting each input–output port pair. For $R = 2$, there are exactly two wavelengths connecting each input–output port pair.

The wavelength reuse and period routing properties of the AWG are illustrated in Fig. 1. Four wavelengths are simultaneously applied at both AWG input ports of a 2×2 AWG. The AWG routes every second wavelength to the same output port. Fig. 1 shows two FSRs, allowing two simultaneous transmissions between each AWG input–output port pair. The number of wavelengths is $\Lambda = R \cdot D$ at each AWG port. From Fig. 1, we also see that in order for a signal from one input port to reach all of the output ports at the same time, a multi–wavelength or broadband light source is required.

Here we point out that the number of nodes N in a metropolitan or local area network is typically larger than D. Combiners are used to connect groups of the transmitters to the input ports of the AWG and splitters are used to connect groups of receivers to the output ports of the AWG. With a given number of nodes, there is more than one way to construct a

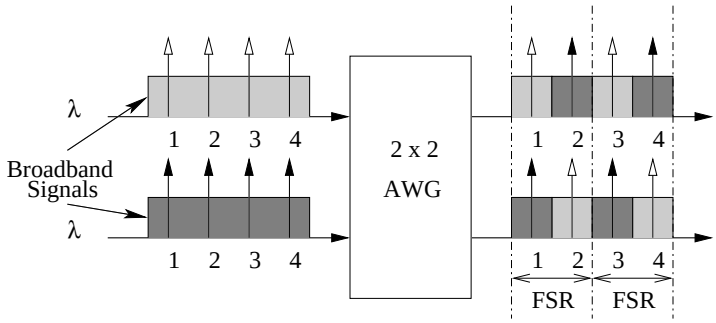

Figure 1. Periodic wavelength routing of an AWG

network by varying the parameters of the AWG and the combiners/splitters. For example, we can connect 16 nodes to a 4×4 AWG using four 4×1 combiners and four 1×4 splitters. Or we can connect the 16 nodes using a 2×2 AWG using two 8×1 combiners and two 1×8 splitters. Assuming $\Lambda = 4$ wavelengths, the first case results in one FSR per input–output port pair, i.e., $R = 1$. The second case results in two FSR's per input–output port pair, i.e., $R = 2$.

CHAPTER 3

Survey of Previous Work

Single–hop networks based on one PSC as the central broadcasting device have been studied extensively since WDM technology was first proposed for optical networks. The studies [8, 9, 11, 15, 19, 31, 34, 35, 37, 44, 47, 48, 58] represent a sample of the numerous proposals of MAC protocols and analysis of throughput–delay performances associated with various PSC based network architectures. Typically, in these single–hop networks, all users are connected to the central PSC. These networks typically employ a pre–transmission coordination Medium Access Control (MAC) protocol, where a control packet is broadcasted to all nodes before the actual data packet is transmitted [39]. After successful pre–transmission coordination, the data packet is transmitted directly from the source to the destination without any intermediate forwarding, thus avoiding any nodal re–transmission burden compared to multi–hop networks. In the recently proposed single–hop networks, both the pre–transmission coordination and the data packet transmission are conducted over the PSC.

1. Protection of Single–hop Networks

Protection of single–hop networks has received only little attention so far [22, 52]. While the passive nature of the PSC and AWG makes the network fairly reliable, it does not eliminate the inherent single point of failure. Clearly, two protection options which come to mind are conventional 1+1 or 1:1 protection. In these cases, the network would consist of two PSCs or two AWGs in parallel. This kind of (homogeneous) protection is rather inefficient: While in the 1+1 protection the backup device is used to carry duplicate data traffic, in the 1:1 protection the backup device is not used at all during normal operation. In the work by Hill *et al.* the central hub of the single–hop WDM network consists of r working AWGs which are protected by n identical standby AWGs. These standby wavelength routers are activated only in case of failure, thus implementing a conventional homogeneous $n : r$ protection scheme. Sakai *et al.* [52] study a dual–star structure where 2 AWGs back up each other in a 1:1 fashion.

2. Wavelength Reuse

The main constraint of using one PSC is that each wavelength provides only one channel of communication between a pair of nodes at any one instance in time. As discussed, wavelengths are precious in metropolitan and local area networks due to cost considerations and tunable transceiver limitations. One of the ways to increase the transmission efficiency, i.e., increase capacity without increasing the number of wavelengths, is to reuse the same set of wavelengths in the network. A number of strategies have been examined over the years. Kanan *et al.* [30] introduce a two level PSC star so that the same set of wavelengths can be reused by each star cluster. Janoska and Todd [26] propose a hierarchical arrangement

of linking multiple local optical networks to a remote optical network. Recently, the use of the wavelength routing AWG as the central hub in single–hop networks has received more attention. The spatial wavelength reuse of the AWG overcomes the channel resource limitations of single hop networks. The photonic feasibility aspects of the single–hop WDM networks based on a uniform–loss cyclic frequency AWG with nodes consisting of individual transceivers as well as transceiver arrays have been demonstrated in [32, 49]. General design principles for networks based on AWGs are studied, for instance, in [1, 23, 57]. Chae *et al.* [6] use an AWG to link multiple PSC networks in series. Again the same set of wavelengths are reused by each star cluster. Banerjee *et al.* [1] and Glance *et al.* [17] outline network architectures based on AWG routers for wavelength reuse. Maier *et al.* [39] provide a functional MAC protocol for an AWG–based single–hop network with spatial wavelength reuse.

3. Transceiver Array Networks

Most of the MAC proposals for PSC networks have been made on 3 main variations of receiver and transmitter configurations: (*i*) fix–turned transmitter (FT) with tunable receivers (TR), (*ii*) tunable transmitters (TT) with fix–tuned receivers (FR), and (*iii*) tunable transmitters (TT) with tunable receivers (TR). Unicasting and multicasting in a single–hop AWG based metro WDM network with decentralized media access control are also studied in [40, 41]. The network considered in [40, 41] also employs a single fast–tunable transmitter and a single fast–tunable receiver at each node. One characteristic of these networks is that only one receiver, either FR or TR, is assigned to data packet reception. So the challenge for multicasting in these networks is to scheduled the idle receivers to maximize throughput and minimize delay. Since Jue and Mukherjee proposed the partitioning of multicast to

improve receiver utilization in [28], partitioning and scheduling algorithms have been one of the themes for improving multicast performance in networks with one data receiver [35, 33]. Hamad *et.al.*in [20] surveys various scheduling and multicast algorithms. One observation of these proposals is that single data receiver at each node present limitations in multicast performance without complex algorithms.

While this single fast–tunable transceiver node architecture is conceptually very appealing and has a number of advantages, such as low power consumption and small foot print, fast–tunable transceivers are generally a less mature technology than fixed–tuned transceiver arrays. More specifically fast–tunable transmitters have just recently been experimentally proven to be feasible in a cost–competitive manner [55], while fast tunable optical filter receivers with acceptable channel crosstalk remain a technical challenge at the photonics level. In contrast, arrays of fixed–tuned transmitters and receivers are better understood [7, 25], more mature, more reliable, and commercially available. Several papers [25, 36] discuss the performance characteristics of transceivers arrays of various sizes. Receiver arrays can be constructed using either waveguides or photonic devices [7]. The drawbacks of fixed transceiver arrays are increased power consumption and larger footprint.

At the MAC protocol level, transceiver arrays have a number of distinct advantages. The transmitter arrays allow for high–speed signaling over the AWG in contrast to the low-speed signaling through the spectral slicing of broadband light sources [41] which suffer from a small bandwidth–distance product. The receiver arrays, on the other hand, relieve the receiver bottleneck caused by multicast traffic, that is transmitted over the large number of wavelength channels obtained from spatial wavelength reuse on the AWG.

The use of transmitter arrays and receiver arrays in WDM networks has not received nearly the attention of networks with tunable transmitters and tunable receivers. The

LAMBDANET [18] is one of the earlier PSC networks proposed by Goodman *et al.* which uses one fix–tuned transmitter and N fixed tuned receivers at each node. The use of receiver arrays enabled point–to–multi-point communication in the LAMBDANET. McKinnon *et al.* studied the performance of a single–hop ATM switch based on the PSC and using tunable transmitter and fixed–tuned receivers in [43]. The use of full fixed–tuned transmitter arrays and receiver arrays connected via the PSC has been studied in [62, 16]. Desai and Ghose [10] compared the performance of single–hop PSC networks using fixed–tuned transmitter and receiver arrays with networks that has single tunable transmitter and/or tunable receiver using trace simulation and found that tuning latencies severely degrade the performance of the network and that network with fixed–tuned transmitters and receivers outperform network with tunable transceivers.

SONATA [3, 4] is a national–scale network based on an AWG. In SONATA individual nodes (terminals) are connected to passive optical networks (PONs) which in turn are connected to the AWG. SONATA employs a centralized network controller to arbitrate the access of the terminals to the shared wavelength channels and wavelength converter arrays at central AWG to balance the load between PON pairs.

CHAPTER 4

AWG‖PSC Network

To improve wavelength efficiency and address the issue of protection for single–hop WDM networks, we propose a network consists of one AWG and one PSC in parallel, which we subsequently call the AWG‖PSC network. The AWG‖PSC network enables highly efficient data transport by (*i*) spatially reusing all wavelengths at all AWG ports, and (*ii*) using those wavelengths *continuously* for data transmission. Under normal operation, i.e., both AWG and PSC are functional, the AWG‖PSC network uniquely combines the respective strengths of both devices and provides *heterogeneous protection* in case either device fails. The presented MAC protocols are devised for the three different operating modes: (*i*) "both AWG and PSC functional" (*AWG–PSC mode*), (*ii*) "PSC failed" (*AWG–only mode*), and (*iii*) "AWG failed" (*PSC–only mode*). We find that the throughput of the AWG‖PSC network in the AWG–PSC mode is significantly larger than the throughput of the stand–alone AWG network plus the throughput of a stand–alone PSC network. Moreover, over a wide operating range the AWG‖PSC network achieves a better throughput–delay performance than a network consisting of either two load sharing PSCs in parallel or two load sharing AWGs in parallel.

1. Architecture

Fig. 2 shows the architecture of the proposed AWG‖PSC network. The PSC and

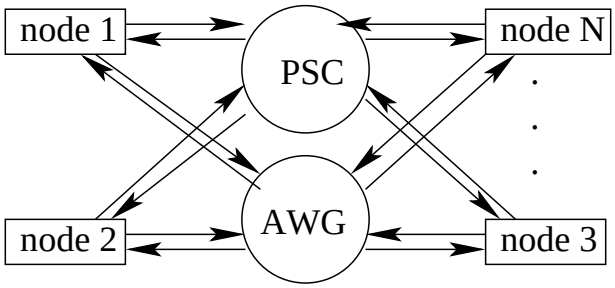

Figure 2. AWG‖PSC Network architecture

the AWG operate in parallel. The nodal architecture is depicted in Fig. 3. In star networks

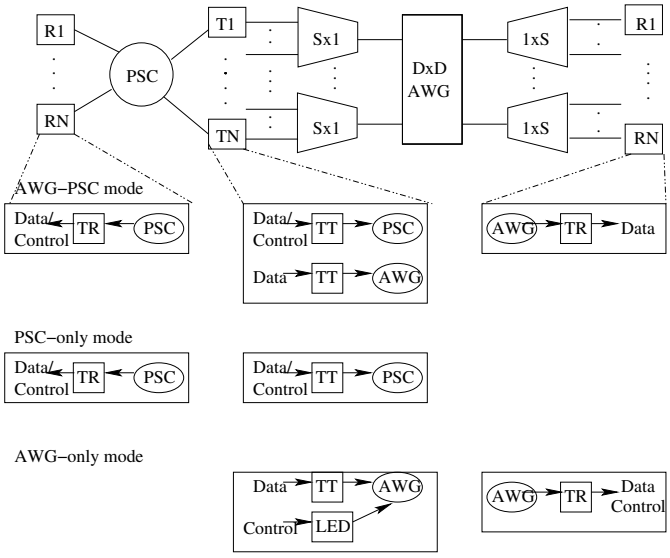

Figure 3. AWG‖PSC node architecture

without redundant fiber back–up, each node is connected by one pair of fibers, one for the

transmission of data, and one for the reception of data. In our network we deploy one–to–one fiber back–up for improved path protection and survivability, that is, each node is connected to the AWG‖PSC network by two pairs of fibers.

Each node is equipped with two fast tunable transmitters (TT), two fast tunable receivers (TR), each with a tuning range of $\Lambda = R \cdot D$ wavelengths, and one off–the–shelf broadband light emitting diode (LED). Due to the extensive spatial wavelength reuse, the tuning range (number of wavelengths) can be rather small. This allows for deploying electro–optic transceivers with negligible tuning times. One TT and one TR are attached directly to one of the PSC's input ports and output ports, respectively. The TT and TR attached to the PSC are henceforth referred to as *PSC TT* and *PSC TR*, respectively. The second TT and TR are attached to one of the AWG's input ports and output ports via an $S \times 1$ combiner and a $1 \times S$ splitter, respectively. These are referred to as *AWG TT* and *AWG TR*.

We note that an alternative architecture to the PSC TT–TR is to equip each node with a tunable PSC transmitter and two fixed–tuned PSC receivers, one tuned to the node's home channel and the other tuned to the control channel. The drawback of this architecture is the lack of data channel flexibility resulting in inefficient channel utilization. In addition, with our approach all wavelength channels can be used for data transmission, whereas with a fixed control channel one wavelength is reserved exclusively for control. Studies in [37, 56] have shown that, by allowing a node to receive data on any free channel, the TT–TR architecture has smaller delays and higher channel utilizations compared to the TT–FR architecture.

The LED is attached to the AWG's input port via the same $S \times 1$ combiner as the AWG TT. The LED is used for broadcast of control packets by means of spectral slicing over

the AWG when the network is operating in AWG–only mode (discussed in more detail in Section 2). Two pairs of TTs and TRs allow the nodes to transmit and receive packets over the AWG and the PSC simultaneously. This architecture also enables transceiver back–up for improved nodal survivability.

2. MAC Protocols

We describe MAC protocols for the normal operating mode as well as the various back–up modes. We define two levels of back–up. The first level is the back–up of the central network components, i.e., the PSC or the AWG. Because the AWG and the PSC operate in parallel, the two devices naturally back–up each other. We have three different modes of operation: (*i*) *AWG–PSC mode*, with both AWG and PSC functional, (*ii*) *PSC–only mode*, with AWG down, and (*iii*) *AWG–only mode*, with PSC down. We present the MAC protocols for all three operating modes. The network's throughput and delay performance for each of the three operating modes is examined in Section 4. The second level of back–up makes use of the two TT/TR's at each node to enable transceiver back–up at the node level. The MAC protocol for transceiver back–up is presented in Appendix A.

2.1. AWG–PSC Mode. In a highly flexible environment where both transmitter and receiver are tunable, wavelength access is typically controlled by reservation protocols, see the survey [39] and references therein. That is, prior to transmitting a given data packet the source node sends a control packet to inform the corresponding destination node.

The wavelength assignment and timing structure are shown in Fig. 4. With a transceiver tuning range of Λ wavelengths, the PSC provides a total of Λ wavelength channels. The length of a PSC frame is F slots. The slot length is equal to the transmission time

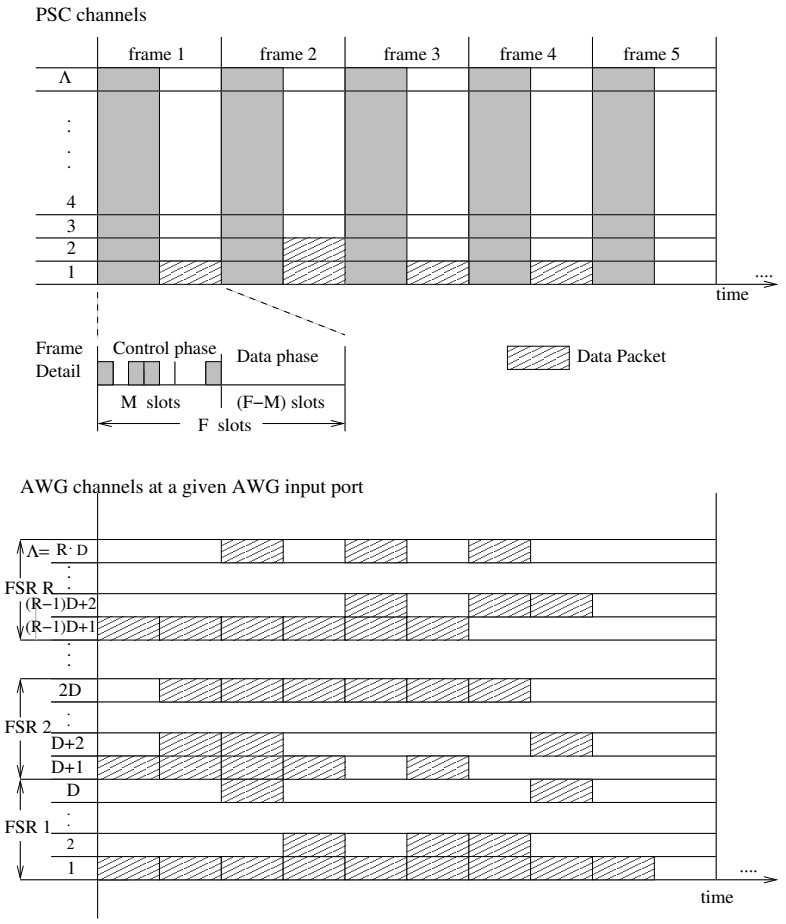

Figure 4. AWG–PSC mode timing structure

of a control packet (which is discussed shortly). Each PSC frame is divided into a control phase and a data phase. During the control phase, all of the nodes tune their PSC TR to a preassigned wavelength. (One of the wavelength channels on the PSC is used as control channel during the first M slots in a frame; in the remaining slots this channel carries data.)

Given N nodes in the network, if node i, $1 \leq i \leq N$, has to transmit a packet to node j, $i \neq j$, $1 \leq j \leq N$, node i randomly selects one of the M control slots and transmits a control packet in the slot. The slot is selected using a uniform distribution to ensure fairness. Random control slot selection, as opposed to fixed reservation slot assignment,

also makes the network upgradeable without service disruptions and scalable.

The nodes transmit their data packets only after knowing that the corresponding control packets have been successfully transmitted and the corresponding data packets successfully scheduled. All nodes learn of the result of the control channel transmission after the one–way end–to–end propagation delay (i.e., half the round–trip time). A control packet collision occurs when two or more nodes select the same control slot. A node with a collided control packet enters the backlog state and retransmits the control packet in the following frame with probability p.

The control packet contains three fields: destination address, length of the data packet, and the type of service. Defining the type of service enables circuit–switching. Once a control packet requesting a circuit is successfully scheduled, the node is automatically assigned a control slot in the following frame. This continues until the node releases the circuit and the control slot becomes available for contention.

A wide variety of algorithms can be employed to schedule the data packets (corresponding to successfully transmitted control packets) on the wavelength channels provided by the AWG and the PSC. To avoid a computational bottleneck in the distributed scheduling in the nodes in our very high–speed optical network, the scheduling algorithm must be simple. Therefore, we adopt a first–come–first–served and first–fit scheduling algorithm with a frame timing structure on the AWG. The frames on the AWG are also F slots long, as the PSC frames. However, unlike the PSC frames, the AWG frames are not subdivided into control and data phase. Instead, the entire AWG frame is used for data. With this algorithm, data packets are assigned wavelength channels starting with the earliest available frame on the lowest FSR on the AWG. Once all the FSRs on the AWG are assigned for that frame, assignment starts on the PSC beginning with the lowest wavelength. Once all

the AWG FSRs and PSC wavelengths are assigned in the earliest available frame, assignment starts for the next frame, again beginning with the lowest FSR on the AWG, and so forth. This continues until the scheduling window is full. The unassigned control packets are discarded and the nodes retransmit the control packets with probability p in the next frame. A node with a collided control packet or a data packet that did not get scheduled (even though the corresponding control packet was successfully transmitted) continues to retransmit the control packet, in each PSC frame with probability p, until the control packet is successfully transmitted and the corresponding data packet scheduled.

The nodes avoid receiver collision by tuning their PSC TR to the preassigned control wavelength during the control phase of each frame and executing the same wavelength assignment (scheduling) algorithm. Each node maintains the status of all the receivers in the network. Also, since both the PSC TR and the AWG TR may receive data simultaneously, in the case when two data packets are addressed to the same receiving node in the same frame, the receivers may be scheduled for simultaneous reception of data from both transmitting nodes. In case there are more than two data packets destined to the same receiving node, transmission for the additional packet(s) has to be scheduled for future frame(s).

We note that we consider unicast traffic throughout this paper. However, we do point out that the AWG||PSC network provides a flexible infrastructure for efficient multicasting. A multicast with receivers at only one AWG output port can be efficiently conducted over the AWG, with the splitter distributing the traffic to all attached receivers. A multicast with receivers at several AWG output ports, on the other hand, might be more efficiently conducted over the PSC (to avoid repeated transmissions to the respective AWG output ports).

2.2. PSC–only Mode. The network operates in the PSC–only mode when the AWG fails. A node scheduled to receive a data packet over the AWG detects AWG failure if the scheduled data packet fails to arrive after the propagation delay. The node then signals other nodes by sending a control packet in the following frame. The network changes from AWG–PSC mode to PSC–only mode after the successful transmission of this control packet.

In this mode, each frame has a control phase and a data phase as illustrated in Fig. 5. During the control phase, all of the nodes with data packets transmit their control packets

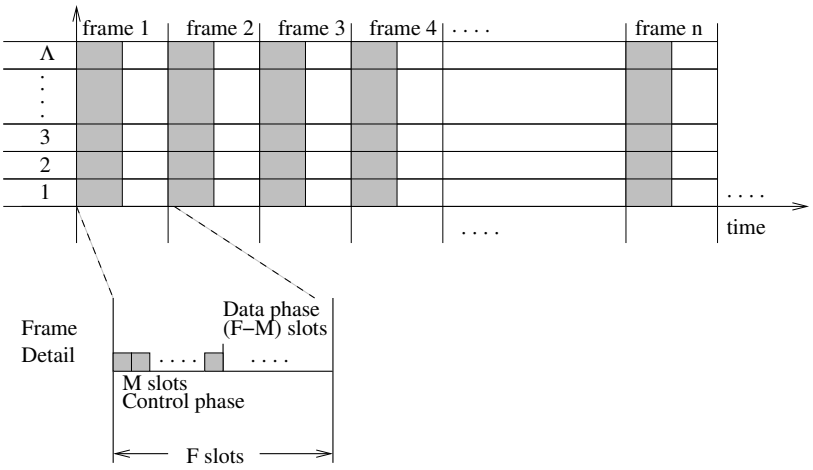

Figure 5. *PSC–only mode* frame structure

in one of the M slots during the control phase. Nodes with collided packets retransmit their control packets following a back–off schedule similar to that of the AWG–PSC mode. The nodes that have successfully transmitted the control packet are assigned the earliest slot starting with the lowest available wavelength. Once the scheduling window is full, the control packets corresponding to unscheduled data packets are discarded and the corresponding nodes retransmit the control packets with probability p in the following frame.

2.3. AWG–only Mode. The network operates in the AWG–only mode when the PSC fails. Since all of the nodes have their PSC TR tuned to the control channel during the control phase of each frame, PSC failure is immediately known by all nodes and the network transitions from AWG–PSC mode to AWG-only mode.

Transmitting and receiving control packets over the AWG are more complicated compared to the PSC. First, recall that a multi–wavelength or a broadband light source is required to transmit a signal from one input port to all output ports (see Fig. 1). Thus, in the AWG–only mode the LED is used to broadcast the control packets by means of spectral slicing. Second, the transmission of control packets follows a timing structure consisting of cycles to prevent receiver collision of spectral slices. For example (see Fig. 1), if two nodes that are attached to different input ports broadcast control packets using their broadband light source, the wavelength routing property of the AWG slices the signals and sends a slice from each of the broadband signals to each output port. The TR at each node can only pick from one of the wavelengths at each output port to receive the control packet, resulting in receiver collision for the second control packet. Therefore, only the group of nodes attached to the same AWG input port via a common combiner is allowed to transmit control packets in a given frame. In the following frame, the next group of nodes attached to another combiner transmits control packets. This continues until all of the nodes have had a chance to transmit a control packet, and the cycle then starts over. Therefore, with a $D \times D$ AWG, a cycle consist of D frames. The control packet transmission cycle and the frame structure are depicted in Fig. 6. Methods for frame and cycle synchronization are beyond the scope of this paper (see for instance [5, 24] for techniques for distributed slot synchronization in WDM networks).

Control packets collide when two or more nodes attached to the same combiner select

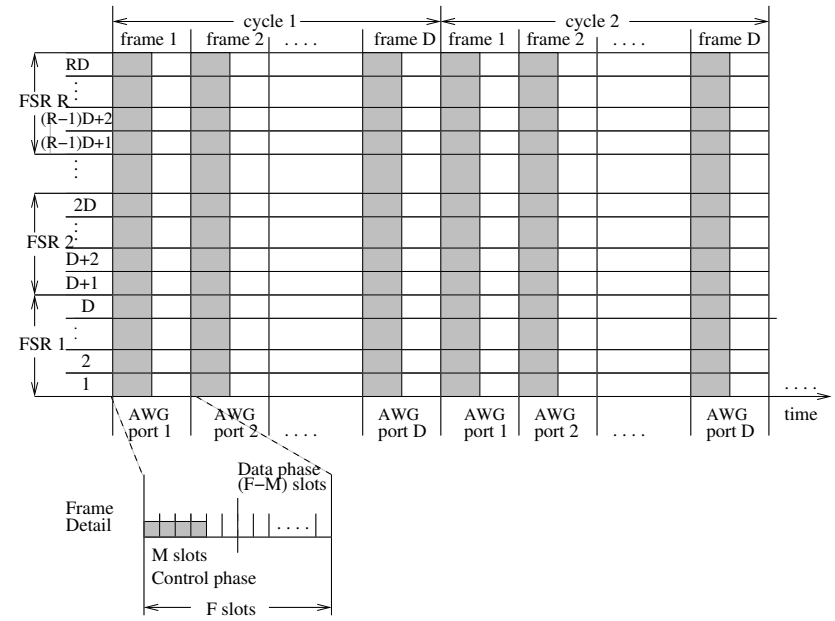

Figure 6. *AWG–only mode* frame structure

the same control slot. Nodes with collided control packets retransmit the control packets in the next transmission cycle with probability p.

In the AWG–only mode we distinguish data packet transmission without spatial wavelength reuse and data packet transmission with spatial wavelength reuse. If the scheduling window for data packets is one frame, then nodes can transmit data packets only in one frame out of the D frames in a cycle, which means that there is effectively no wavelength reuse. Full spatial wavelength reuse requires a scheduling window of at least D frames.

3. Analysis

In this section we develop a probabilistic model for the AWG‖PSC network.

3.1. System Model. We make the following assumptions in the modeling of the proposed network and MAC protocols.

- *Fixed data packet size*: Data packets have a fixed size of $F/2$ slots. Both the control phase and the data phase on the PSC are $F/2$ slots long, i.e., $M = F - M = F/2$. On the AWG, each frame accommodates two data packets, as illustrated in Fig. 4. With a degree of D and R utilized FSRs (and a corresponding transceiver tuning range of $\Lambda = D \cdot R$), the AWG provides Λ wavelength channels at each of its D ports, for a total of $D^2 \cdot R$ wavelength channels. Thus, the AWG can accommodate at most $2 \cdot D^2 \cdot R$ data packets per frame.

- *Uniform unicast traffic*: A data packet is destined to any one of the N nodes, including the originating node, with equal probability $1/N$. (In our simulations, see Section 4, a node does not transmit to itself. We find that the assumption made in our analytical model that a node transmits to itself with probability $1/N$ gives very accurate results.)

- *Scheduling window*: The scheduling window is generally one frame. (For the AWG–only mode we consider a scheduling window of one frame as well as a scheduling window of one cycle.) In the AWG–PSC mode and the PSC–only mode, a node with collided control packet or with successfully transmitted control packets but no resources (for data packet scheduling) in the current frame retransmits its control packet in the following frame with probability p. In the case of the AWG–only mode, a node with collided control packet or with no transmission resources retransmits in the following cycle with probability p_A.

- *Nodal states and traffic generation*: There are two nodal states: idle and backlogged. A node with no data packet in its buffer is defined as idle and generates a new data

packet with probability σ at the beginning of a frame. Let η denote the number of nodes in this idle state. A node is backlogged if it has (i) a control packet that has failed in the control packet contention, or (ii) a successful control packet but no transmission resources for scheduling the corresponding data packet. The number of backlogged nodes equals $N - \eta$. Backlogged nodes retransmit their control packets with probability p in a frame. If a node has successfully transmitted a control packet and the corresponding data packet has been successfully scheduled, then the node is considered idle and generates a new packet with probability σ in the following frame.

- *Receiver Collision*: We ignore receiver collisions in our analysis. In our simulations in Section 4, on the other hand, we take receiver collisions into consideration. In particular, in the AWG–PSC mode we schedule a data packet on the AWG only if the AWG TR is available. If the AWG TR is busy (or the AWG channels are already occupied), we try to schedule the packet on the PSC. If the PSC TR is busy (or the PSC channels are already occupied), the data packet scheduling fails and the transmitting node retransmits another control packet in the following frame with probability p. In our simulations of the AWG–only mode (PSC–only mode), the data packet scheduling fails if the AWG TR (PSC TR) is busy. Our simulation results in Section 4 indicate that the impact of receiver collision on throughput and delay is negligible. This is consistent with [41] which has shown that the effect of receiver collisions is negligible if the number of nodes N is moderately large, which is typical for metro networks.

- *Non–persistence*: If a control packet fails (in control packet contention or data packet scheduling) we draw a new independent random destination for the corresponding data

packet. Our simulations in Section 4 do not assume non-persistence and demonstrate that the non-persistence assumed in the probabilistic model gives accurate results.

3.2. Control packet contention analysis. A given control slot contains a successfully transmitted control packet if (i) it contains exactly one control packet corresponding to a newly arrived data packet (from one of the idle nodes) and no control packet from the backlogged nodes, or (ii) it contains exactly one control packet from a backlogged node and no control packet corresponding to newly arrived data packets. Let X_i, $i = 1 \ldots M$, denote the number of control packets in slot i. The probability of a given slot containing a successfully transmitted control packet is:

$$P(X_i = 1) = \eta \frac{\sigma}{M} \left(1 - \frac{\sigma}{M}\right)^{\eta-1} \left(1 - \frac{p}{M}\right)^{N-\eta} + (N - \eta) \frac{p}{M} \left(1 - \frac{p}{M}\right)^{N-\eta-1} \left(1 - \frac{\sigma}{M}\right)^{\eta} := \kappa, \quad (4.1)$$

where we assume for simplicity that the number of control packets corresponding to newly arrived data packets is independent of the number of control packets corresponding to backlogged data packets.

The expected number of successfully transmitted control packets in each frame is $\sum_{i=1}^{M} P(X_i = 1)$, which has a binomial distribution $BIN(M, \kappa)$. Hence the expected number of successful control packets per frame is $M \cdot \kappa$.

3.3. AWG–PSC mode data packet scheduling. We assume that the data packet from each of the nodes is destined to any other node with equal probability. There are an equal number of nodes attached to each of the combiners and the splitters of a $D \times D$ AWG. Thus, the probability that a control slot contains a successfully transmitted control packet for data transmission between a given input–output port pair is κ/D^2. For notational convenience, let $\rho := \kappa/D^2$.

In the AWG–PSC mode, the throughput of the network is the combined throughput of the AWG and the PSC. Nodes with successfully transmitted control packets are first scheduled using the wavelengths on the AWG. Let Z_A denote the expected throughput on the AWG (in packets per frame). With R FSRs serving each input–output port pair per half–frame, D input ports and D output ports, the expected number of packets transmitted per frame over the AWG is:

$$Z_A = D^2 \cdot \sum_{i=1}^{2 \cdot R} i \binom{M}{i} \rho^i (1-\rho)^{M-i} + 2 \cdot R \cdot D^2 \cdot \sum_{j=2R+1}^{M} \binom{M}{j} \rho^j (1-\rho)^{M-j}. \tag{4.2}$$

If all of the FSRs for a given input–output pair are scheduled, then the next packet is scheduled on a PSC channel. Let Z_P denote the expected throughput over the PSC channels (in packets per frame). Let $q_{ij}[n]$ denote the probability that there are $n = 0, 1, \ldots, (M-2R)$ overflow packets from AWG input port i, $i = 1, \ldots, D$ to output port j, $j = 1, \ldots, D$. Recall that the control packets are uniformly distributed over the input–output port pairs. Thus, the overflows from all of the input–output port pairs have the same distribution. So we can drop the subscript ij. If the number of packets destined from an input port to an output port is R or less, then there is no overflow to the PSC. If the number of packets for the given input–output port pair is $R+n$ with $n \geq 1$, then there are n overflow packets. Hence,

$$q[n] = \begin{cases} \sum_{i=0}^{2R} \binom{M}{i} \rho^i (1-\rho)^{M-i} & \text{for } n = 0, \\ \binom{M}{n+2R} \rho^{n+2R} (1-\rho)^{M-n-2R} & \text{for } n = 1, \ldots, M-2R. \end{cases} \tag{4.3}$$

Let $Q[m], m = 1, \ldots, (M-2R) \cdot D^2$, denote the probability that there are a total of m overflow packets. To simplify the evaluation of $Q[m]$, we assume that the individual overflows are mutually independent. With this assumption, which as our verifying simulations (see Section 4) indicate gives accurate results, the distribution of the combined arrivals at

the PSC $Q[m]$ is obtained by convolving the individual $q_{ij}[n]$'s, i.e.,

$$Q[m] = q_{11}[n] * q_{12}[n] * \cdots * q_{1D}[n] * \cdots * q_{DD}[n]. \tag{4.4}$$

With $Q[m]$, we obtain the expected PSC throughput as approximately

$$Z_P = \sum_{i=1}^{\Lambda} i \cdot Q[i] + \Lambda \cdot \sum_{j=\Lambda+1}^{(M-2R)\cdot D^2} Q[j]. \tag{4.5}$$

The combined throughput from both AWG and PSC channels is the sum of Z_A and Z_P. To complete the throughput analysis, we note that in equilibrium the throughput is equal to the expected number of newly generated packets, i.e.,

$$Z_A + Z_P = \sigma \cdot E[\eta]. \tag{4.6}$$

For solving this equilibrium equation we make the approximation that the number of idle nodes η has only small variations around its expected value $E[\eta]$, i.e, $\eta \approx E[\eta]$, which as our verifying simulations in Section 4 indicate gives accurate results.

By now substituting (4.2) and (4.5) into (4.6), we obtain

$$D^2 \cdot \sum_{i=1}^{2R} i \binom{M}{i} \left(\frac{\kappa}{D^2}\right)^i \left(1 - \frac{\kappa}{D^2}\right)^{M-i} + 2 \cdot R \cdot D^2 \cdot \sum_{j=2R+1}^{M} \binom{M}{j} \left(\frac{\kappa}{D^2}\right)^j \left(1 - \frac{\kappa}{D^2}\right)^{M-j} + \sum_{i=1}^{\Lambda} i \cdot Q[i] + \Lambda \cdot \sum_{j=\Lambda+1}^{(M-2R)\cdot D^2} Q[j] = \sigma \cdot \eta, \tag{4.7}$$

where κ is given by (4.1) and $Q[\cdot]$ is given by (4.4). We solve (4.7) numerically for η, which can be done efficiently using for instance the bisection method. With the obtained η we calculate κ (and ρ), and then Z_A and Z_P.

3.4. Delay. The average delay in the AWG||PSC network is defined as the average time (in number of frames) from the generation of the control packet corresponding to a data packet until the transmission of the data packet commences. Since in the AWG–PSC

mode the throughput of the network in terms of packets per frame is equal to $Z_A + Z_P$, the number of frames needed to transmit a packet is equal to $1/(Z_A + Z_P)$. Given that there are $N - \eta$ nodes in backlog and assuming that the propagation delay is smaller than the frame length, the average delay in number of frames is

$$Delay = \frac{N - \eta}{Z_P + Z_A}. \tag{4.8}$$

(Propagation delays larger than one frame are considered in the analysis in Appendix I.)

3.5. PSC-only Mode. In the PSC-only mode, the channels are shared by all of the nodes. We assume a scheduling window length of one frame. If a control packet is successfully transmitted, but the corresponding data packet can not be transmitted due to lack of transmission resources, the node has to retransmit the control packet. The maximum number of packets transmitted per frame is equal to the number of channels Λ. The probability of a control slot containing a successfully transmitted control packet is given in (4.1). Hence, the expected number of successfully scheduled transmissions per frame Z_{PM} is

$$Z_{PM} = \sum_{i=1}^{\Lambda} i \binom{M}{i} \kappa^i (1-\kappa)^{M-i} + \Lambda \cdot \sum_{j=\Lambda+1}^{M} \binom{M}{j} \kappa^j (1-\kappa)^{M-j}, \tag{4.9}$$

and in equilibrium the throughput is equal to the expected number of new packet arrivals, i.e.,

$$Z_{PM} = \sigma \cdot E[\eta]. \tag{4.10}$$

Z_{PM}, η, and κ are obtained by simultaneously solving equations (4.1), (4.9), and (4.10).

Analogous to (4.8), the average delay is $(N - E[\eta])/Z_{PM}$ frames.

3.6. AWG-only Mode. In the AWG–only mode we consider two scenarios. In the first scenario, we set the length of the scheduling window to one frame. Recall that under this condition, there is no spatial wavelength reuse. In the second scenario we set the length of the scheduling window to D frames, i.e., one cycle. In this scenario there is full wavelength reuse.

Since transmissions in the AWG–only mode are organized into cycles, we define σ_A as the probability of an idle node having generated a new packet by the beginning of its transmission cycle. Given that an idle node generates a new packet with probability σ at the beginning of a frame, we have $\sigma_A = 1-(1-\sigma)^D$. Similarly, we define p_A as the probability that a backlogged node re–transmits a control packet at the beginning of a cycle, where $p_A = 1-(1-p)^D$. For a $D \times D$ AWG, N/D nodes are allowed to transmit control packets in a given frame. Thus the probability of a given control slot containing a successfully transmitted control packet is

$$\kappa_A = \frac{\eta}{D}\left(\frac{\sigma_A}{M}\right)\left(1-\frac{\sigma_A}{M}\right)^{\frac{\eta}{D-1}}\left(1-\frac{p_A}{M}\right)^{\frac{N-\eta}{D}} + \frac{N-\eta}{D}\left(\frac{p_A}{M}\right)\left(1-\frac{p_A}{M}\right)^{\frac{N-\eta}{D-1}}\left(1-\frac{\sigma_A}{M}\right)^{\frac{\eta}{D}} \quad (4.11)$$

The average throughput over the AWG in packets per *frame* is equal to the average number of packets transmitted from one given input port to the D output ports in one *cycle*. We assume that a control packet is destined to an output port with equal probability. The probability of a control slot containing a successfully transmitted control packet destined to a given output port is κ_A/D. The AWG accommodates up to R packets per input–output port pair per frame, since the R utilized FSRs provide R parallel wavelength channels between each input–output port pair. Without wavelength reuse, i.e., with a scheduling window of one frame, the nodes at a given input port can utilize the R wavelength channels that connect the considered input port to a given output port only during the latter half of one frame out of the D frames in a cycle. Hence, the expected number of successfully

scheduled packets Z_{AM} per frame is

$$Z_{AM} = D \cdot \sum_{i=1}^{R} i \binom{M}{i} \left(\frac{\kappa_A}{D}\right)^i \left(1 - \frac{\kappa_A}{D}\right)^{M-i} + R \cdot D \cdot \sum_{j=R+1}^{M} \binom{M}{j} \left(\frac{\kappa_A}{D}\right)^j \left(1 - \frac{\kappa_A}{D}\right)^{M-j} \quad (4.12)$$

We solve for η numerically using (4.1), (4.12) and the equilibrium condition $Z_{AM} = \sigma_A \cdot E[\eta]/D$. With the obtained η we calculate κ_A and then Z_{AM}. Z_{AM}, η, and κ_A are solved simultaneously using equations (4.11), (4.12),

In the second scenario, i.e., with full wavelength reuse, successful control packets destined for a given output port not scheduled in the current frame are scheduled in the following frame, up to D frames. So the AWG accommodates up to $R \cdot D$ $(= \Lambda)$ packets per input–output port pair per cycle. Hence, with wavelength reuse, the expected number of successfully scheduled packets Z_{RE} per frame is

$$Z_{RE} = D \cdot \sum_{i=1}^{R \cdot D} i \binom{M}{i} \left(\frac{\kappa_A}{D}\right)^i \left(1 - \frac{\kappa_A}{D}\right)^{M-i} + R \cdot D^2 \cdot \sum_{j=R \cdot D+1}^{M} \binom{M}{j} \left(\frac{\kappa_A}{D}\right)^j \left(1 - \frac{\kappa_A}{D}\right)^{M-j} \quad (4.13)$$

Z_{RE}, η, and κ_A are obtained by simultaneously solving equations (4.1), (4.13) and the equilibrium condition $Z_{RE} = \sigma_A \cdot E[\eta]/D$.

With the obtained η we calculate κ_A and then Z_{RE}.

We note that the maximum number of packets that the AWG can accommodate in the AWG–only mode with full wavelength reuse per frame can be increased from $D \cdot \Lambda$ to $D \cdot \Lambda + \Lambda$ by employing spreading techniques for the control packet transmissions. With spreading of the control packet transmissions, the nodes at a given AWG input port can send data packets in parallel with their control packets during the first half of the frame as studied in [38]. We also remark that with an additional LED attached to the PSC, the nodes could send data packets in parallel with (spreaded) control packets over the PSC when the AWG‖PSC network runs in the AWG–PSC mode. This would increase the number of packets that the AWG‖PSC network can accommodate in the AWG–PSC mode per frame

by Λ. In order not to obstruct the key ideas of the AWG‖PSC network, we do not consider the spreading of control information in this paper.

In the scenario without wavelength reuse, there are two delay components. The first component is the delay resulting from the control packet contention and the scheduling process. This component equals the number of backlogged nodes divided by the throughput. The second component is the waiting period in the transmission cycle. All of the idle nodes generate a new packet with probability σ at the beginning a frame. But the nodes transmit control packets once every D frames. Hence, the expected waiting period from the generation of a new data packet to the transmission of the corresponding control packet is the mean of a truncated geometric distribution, i.e.,

$$I_{del} = \frac{\sum_{i=0}^{D}(D-i)\cdot\sigma\cdot(1-\sigma)^i}{1-(1-\sigma)^D}. \tag{4.14}$$

Combining the two components, the total mean delay (in number of frames) is

$$Delay_{AM} = \frac{N - E[\eta]}{Z_{AM}} + I_{del}. \tag{4.15}$$

In the scenario with wavelength reuse, there are three delay components. The first two components are the same as for the scenario without wavelength reuse. The third delay component occurs in the case when the number of scheduled packet is larger than $D \cdot R$. In this case, the packets scheduled in the future frames experience a delay of $(Z_{RE} - D \cdot R)^+/(2 \cdot D \cdot R)$ frames, where $(Z_{RE} - D \cdot R)^+ = \max(0,\ Z_{RE} - D \cdot R)$. To see this note that if $Z_{RE} > D \cdot R$, the packets not scheduled in the current frame have to wait an average $(Z_{RE} - D \cdot R)/(2 \cdot D \cdot R)$ frames for transmission. Combining the three components, the total mean delay (in number of frames) is

$$Delay_{RE} = \frac{N - E[\eta]}{Z_{RE}} + I_{del} + \frac{(Z_{RE} - D \cdot R)^+}{2 \cdot D \cdot R}. \tag{4.16}$$

Table 1. Network parameters and their default values

N	number of nodes in network	200
D	degree (number of ports) of AWG	4
R	number of utilized FSRs	2
Λ	$(= D \cdot R)$, number of wavelengths (transceiver tuning range)	8
p	packet re–transmission probability $(= M/N)$	0.85
F	number of slots per frame	340
M	number of control slots per frame	170
σ	packet generation probability (traffic load)	

4. Numerical and Simulation Results

In this section, we examine the throughput–delay performance of the AWG‖PSC network in the three operating modes: (*i*) AWG–PSC mode, (*ii*) PSC–only mode, and (*iii*) AWG–only mode, by varying system parameters around a set of default values, which are summarized in Table 2. (We set $p = M/N$ as this setting gives typically a large probability κ of success in the control packet contention. Note from (4.1) that κ is maximized for $p = (M - \eta\sigma)/(N - \eta - 1)$.) We provide numerical results obtained from our probabilistic analysis (marked (A) in the plots) as well as from simulations of the network (marked with (S) in the plots). Each simulation was run for 10^6 frames including a warm–up phase of 10^5 frames; the 99% confidence intervals thus obtained were always less than 1% of the corresponding sample mean. Throughout the simulations, we used the σ values 0.01, 0.05, 0.10, 0.15, 0.2, 0.4, 0.6, 0.8, and 1.0. We note that in contrast to our probabilistic analysis, our simulations do take receiver collisions into consideration. Also, in the simulations a given node does not transmit to itself. In addition, in the simulations, we do not assume non–persistence, i.e., the destination of a data packet is not renewed when the corresponding

control packet is unsuccessful.

Fig. 7 compares the throughput–delay performance of the network for different AWG degrees $D = 2, 4$, and 8 (with the number of used FSRs fixed at $R = 2$, thus the corresponding Λ values are 4, 8, and 16). For small σ, the throughput–delay performance for the three

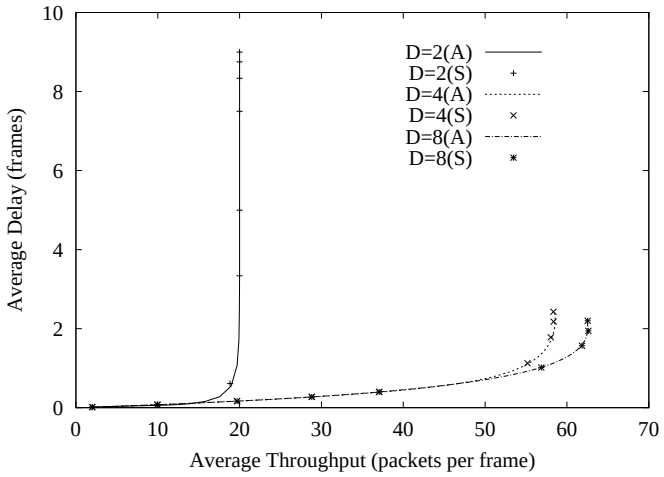

Figure 7. Throughput–delay performance for AWG degree $D = 2$, 4, and 8. ($R = 2$, fixed).

D values are about the same. For large σ, the throughput for $D = 2$ peaks at 20 packets per frame and the delay shoots up to very large values. A network constructed using $D = 8$ achieves higher throughput at lower delays compared to the $D = 4$ network at high traffic levels. Recall that the wavelength reuse property of the AWG allows each wavelength to be simultaneously used at all of the input ports, thus providing $D \cdot \Lambda$ channels. Furthermore, each AWG FSR at each port accommodates 2 data packet transmissions per frame. Thus the maximum combined throughput of AWG and PSC is $2 \cdot D \cdot \Lambda + \Lambda$ data packets per frame. For $D = 2$, the maximum throughput is 20 packets per frame as indicated in the graph. The maximum throughput for $D = 4$ and $D = 8$ are 72 and 272 packets per frame, respectively. For these two cases, the throughput is primarily limited by the number of

successful control packets (per frame); whereas the data packet scheduling is the primary bottleneck for $D = 2$.

Fig. 8 compares the throughput–delay performance of the network for different numbers of used FSRs $R = 1, 2$, and 4 (with the AWG degree fixed at $D = 4$, thus the corresponding Λ values are 4, 8, and 16). The throughput for $R = 1$ peaks at 32 packets per

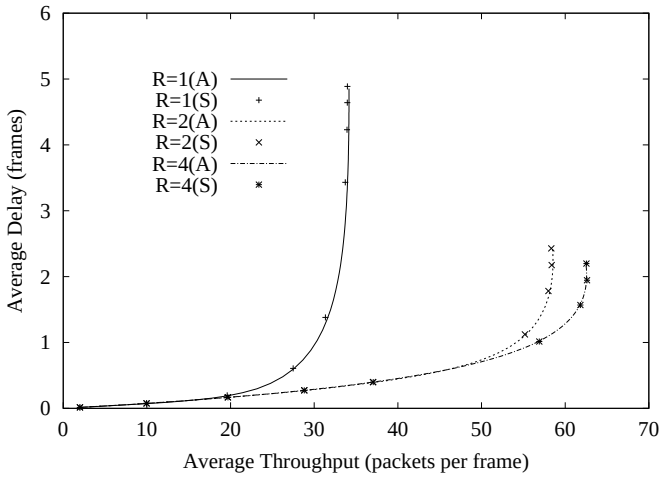

Figure 8. Throughput–delay performance for $R = 1$, 2, and 4 used FSRs. ($D = 4$, fixed)

frame and the delay grows to large values, while the throughput and delay for $R = 2$ and $R = 4$ are approximately the same. Increasing R increases the number of channels for each input–output port pair on the AWG, thus increasing the number of channels in the network. For $R = 1$, the maximum throughput is $2 \cdot D \cdot \Lambda + \Lambda = 36$ packets per frame. The throughput is primarily limited by the scheduling capacity of the network. For $R = 2$ and $R = 4$ the maximum throughputs are 72 and 144 packets per frame, respectively. For these two cases, the throughput is primarily limited by the number of control packets that are successful in the control packet contention. The conclusion is that increasing the number of channels for each input–output port pair does not yield measurable improvements in throughput or

delay when there are not enough successful control packets.

In Fig. 9, we fix the number of wavelengths in the network ($\Lambda = 8$) and examine the throughput–delay performance for different combinations of D and R with $D \cdot R = 8$. We

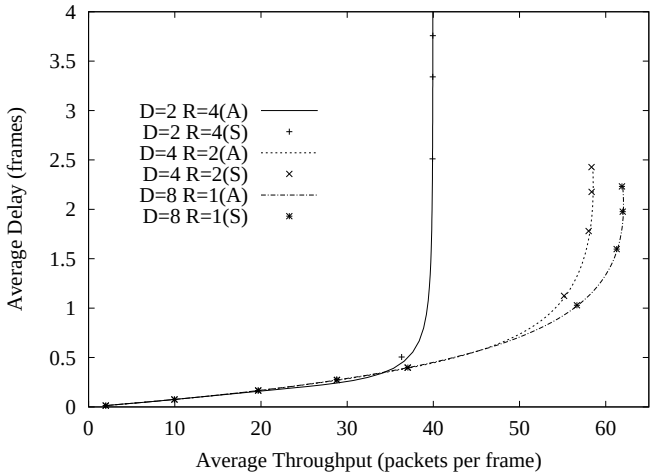

Figure 9. Throughput–delay performance for fixed tuning range $\Lambda = R \cdot D = 8$ wavelengths.

examine the cases: ($D = 2$, $R = 4$), ($D = 4$, $R = 2$), and ($D = 8$, $R = 1$). We observe that ($D = 2$, $R = 4$) has the shortest delay up to a throughput of 21 packets per frame, and a maximum throughput of 40 packets per frame. The delays for ($D = 4$, $R = 2$) and ($D = 8$, $R = 1$) are approximately the same up to a throughput of 50 data packets per frame. At higher traffic levels, the ($D = 8$, $R = 1$) network achieves higher throughput at lower delays compared to the ($D = 4$, $R = 2$) network due to the larger number of channels in the ($D = 8$, $R = 1$) network. The combination ($D = 2$, $R = 4$) achieves the shortest delay at small σ due to higher channel utilization from the larger number of FSRs. The throughput for ($D = 2$, $R = 4$) is bounded by the number of channels $2 \cdot D \cdot \Lambda + \Lambda = 40$.

Fig. 10 compares the throughput–delay performance of the network in the four modes: AWG–only mode without wavelength reuse (i.e., a scheduling window of one frame), AWG–only mode with wavelength reuse (i.e., a scheduling window of one cycle), PSC–only

mode, and AWG–PSC mode. The PSC–only mode has a maximum throughput of 8 data

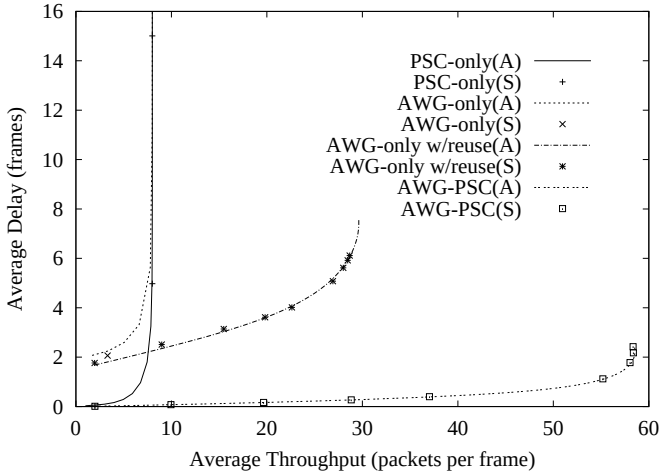

Figure 10. Throughput–delay performance comparison for three modes of operation.

packets per frame. This is expected because the maximum number of channels in a PSC–network is equal to the number of available wavelengths, $\Lambda = 8$. The AWG–only mode with wavelength reuse achieves throughputs up to roughly 30 packets per frame. This is primarily due to the the larger number of $D \cdot \Lambda = 32$ available wavelength channels with spatial wavelength reuse. The delay for the AWG–only mode is larger than for both the PSC–only mode and the AWG–PSC mode at low traffic. This is due to the cyclic control packet transmission in the AWG–only mode. The AWG–PSC mode achieves the largest throughput and the smallest delays for all levels of traffic.

We also observe that for a given level of delay, the throughput for the AWG‖PSC network is significantly larger than the total throughput obtained by combining the throughput of a stand–alone AWG network with the throughput of a stand–alone PSC network. The AWG‖PSC network in the AWG–PSC mode has a maximum throughput of 59 packets per frame and a delay of no more than 3 frames. For the same level of delay, the throughput of a stand–alone PSC network and a stand–alone AWG network are 8 and 12 packets per

frame, respectively. So by combining the AWG and the PSC in the AWG‖PSC network, we effectively tripled the total combined throughput of two stand–alone networks.

Next, we compare the AWG‖PSC network to its peers of homogeneous two–device networks. Fig. 11 compares the throughput–delay performance of the AWG‖PSC network with a PSC‖PSC network (consisting of two PSCs in parallel) and an AWG‖AWG network (consisting of two AWGs in parallel). The throughput–delay performance of these homogeneous two device networks is analyzed in detail in Appendix I. In brief, in the PSC‖PSC network an idle node generates a new packet with probability σ at the beginning of a frame. In the AWG‖AWG network an idle node generates a new packet with probability $\sigma_A = 1-(1-\sigma)^D$ at the beginning of a cycle. The control packet contention and scheduling processes are similar to those described in Section 2. We observe that the average through-

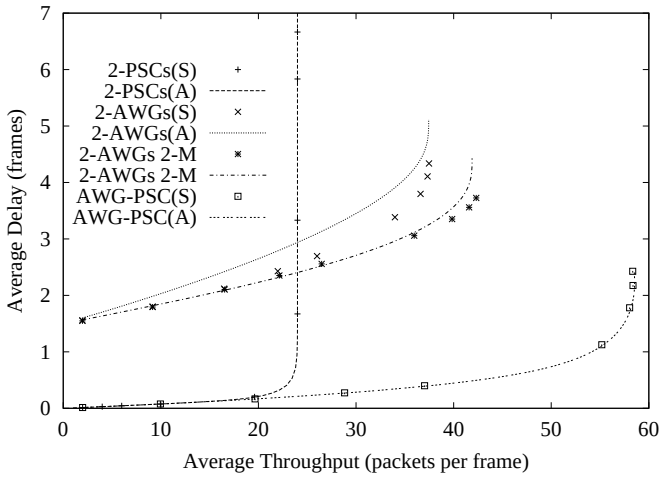

Figure 11. Throughput–delay performance comparison for three networks: PSC‖PSC, AWG‖AWG, and AWG‖PSC

put of the AWG‖PSC network is significantly larger and the delay significantly smaller than for the other two two–device networks. In the PSC‖PSC network, we observe a maximum throughput of 24 packets per frame. We imposed the control packet contention only on

one of the devices. This allows two data slots per frame on the second PSC, which effectively provides three data slots per wavelength on both devices in each frame. With $\Lambda = 8$ wavelengths available, the PSC‖PSC network has a total of 24 data slots per frame. An alternative framing structure is to have control packet contention on both PSCs. This would double the number of contention slots per frame, but there would be only one data slot per frame on each PSC, giving us only 16 data slots per frame. Since the number of wavelength channels is the obvious bottleneck for the PSC‖PSC network, we chose the former framing method to alleviate the bottleneck for data transmission.

For the AWG‖AWG network, we present numerical and simulation results for two framing structures. The first framing structure has control contention only on one of the AWGs. The second framing structure (marked 2–M in the plots) has control packet contention slots and data slots imposed on both devices. We observe that the framing structure with control contention on both AWGs achieves larger throughput and smaller delays compared to the framing structure with contention only over one AWG. The maximum throughput for one control slot contention and two control contentions are 37 packets and 42 packets, respectively. Using one control contention per frame, the maximum number of data slots is $3 \cdot D \cdot \Lambda = 96$. Using two control contentions per frame, the maximum number of data slots is $2 \cdot D \cdot \Lambda = 64$. Although the two control contention framing structure has fewer data slots, it has a larger probability of success for control packet contention, thus resulting in larger throughput and smaller delay. The primary reason that the throughput levels in both of these framing structures are significantly smaller than their data scheduling capacity is the lower traffic as a result of the cyclic control packet transmission structure. For $\sigma = 1$ an idle node in the PSC‖PSC or AWG‖PSC network generates a new packet with probability one at the beginning of a frame, whereas an idle node in the AWG‖AWG

network generates a new packet with the corresponding probability $\sigma_A = 1$ at the beginning of a cycle (consisting of D frames). In other words, the AWG‖AWG network is "fed" with a smaller input traffic rate since each node generates at most one new packet in a cycle. Thus the maximum number of control packets corresponding to new data packet in a 200-node network with a 4×4 AWG is 50 control packets per frame.

To get a better understanding of the relative performance of the AWG‖PSC network with respect to the AWG‖AWG network, we consider an alternative operation of the AWG‖AWG network, which ensures that both networks are "fed" with the same traffic rate. Specifically, we equip each node in the AWG‖AWG network with D packet buffers; one for each of the frames in a cycle. (Each node in the AWG‖PSC continues to have only one packet buffer.) Each node in the AWG‖AWG network generates a new packet with probability σ at the beginning of a frame if the buffer corresponding to that frame is idle. As explained in Section 2.3 the nodes in the AWG‖AWG network can only send control

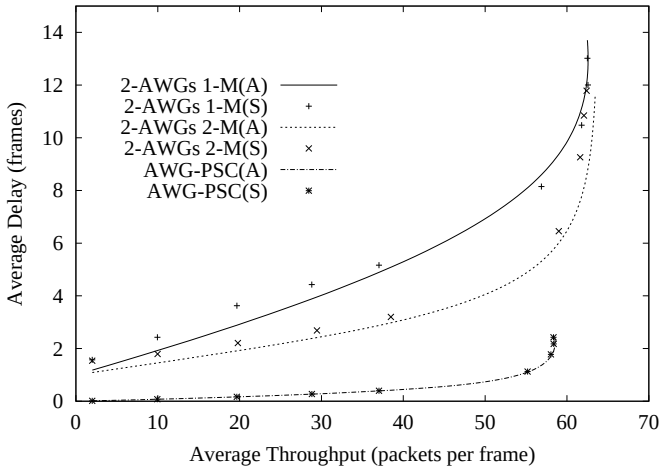

Figure 12. Throughput–delay performance comparison for three networks: D–buffered AWG‖AWG with one control, D–buffered AWG‖AWG with two controls, and AWG‖PSC

packets in the one frame (out of the D frames in the cycle) that is assigned to the node's

combiner. Whereas in the single–buffer operation considered in Section 2.3 and Section 3.6, a node sends at most one control packet in that assigned frame, in the D–buffer operation considered here a node sends up to D control packets—one for each of the packets in its D buffers— in the assigned frame. The control packet contention and data packet scheduling for this D–buffer operation of the AWG‖AWG network and the resulting throughput–delay performance are analyzed in detail in Appendix II.

Fig. 12 compares the throughput–delay performance for the AWG‖PSC network with the throughput–delay performance of the AWG‖AWG network with D–buffer operation, both with control packet contention on one AWG and on two AWGs. We observe that the AWG‖AWG network with D–buffer operation achieves somewhat larger throughput than the AWG‖PSC network. However, the AWG‖PSC network achieves significantly smaller delay throughout. While the comparison in Fig. 12 is fair in that both networks are "fed" with the same traffic rate, the AWG‖AWG network is given the advantage of D packet buffers and a scheduling window of D frames (both resulting in higher complexity), whereas the AWG‖PSC network has a single packet buffer and a scheduling window of one frame. The comparisons in both Fig. 11 and Fig. 12 indicate that the AWG‖PSC network achieves good throughput–delay performance at low complexity.

CHAPTER 5

$FT^{\Lambda} - FR^{\Lambda}$ AWG network

In this chapter we develop and evaluate the $FT^{\Lambda} - FR^{\Lambda}$ AWG network, an AWG based single–hop WDM network with an array of fixed–tuned transmitters and receivers at each network node. The proposed $FT^{\Lambda} - FR^{\Lambda}$ AWG network is practical due to its mature, commercially available building blocks. In addition, the transmitter arrays allow for high speed signaling over the AWG while the receiver arrays relieve the receiver bottleneck arising from multicasting in conjunction with spatial wavelength reuse on the AWG. As we demonstrate through analysis and simulation, the $FT^{\Lambda} - FR^{\Lambda}$ node architecture, aside from being readily deployable, achieves good throughput–delay performance especially for a mix of unicast and multicast traffic.

1. Architecture

Our AWG based network architecture is illustrated in Fig. 13. The AWG has D input ports and D output ports. There are N nodes in the network. At each AWG input port, an $S \times 1$, $S = N/D$, combiner collects transmissions from the transmitters of S attached nodes. At each AWG output port, a $1 \times S$ splitter equally distributes the signal to S individual fibers that are attached to the receivers of the nodes. We use the notation

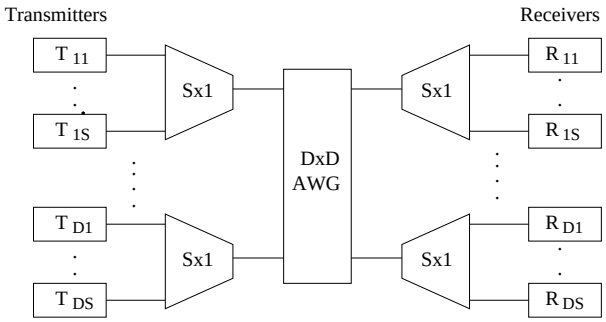

Figure 13. Network architecture

$N_{i,j}$, $i = 1, 2, \ldots, D$, $j = 1, 2, \ldots, S$, to designate the jth node attached to the ith AWG port. In Fig. 13, $T_{i,j}$ and $R_{i,j}$ correspond to the transmitter array and the receiver array of node $N_{i,j}$.

The node architecture is shown in Fig. 14. Each node is equipped with a transmitter

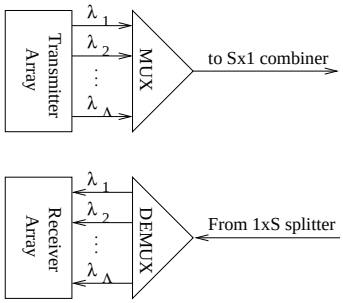

Figure 14. Detailed node architecture

array consisting of Λ fixed tuned transmitters and a receiver array consisting of Λ fixed tuned receivers. The optical multiplexer is used to combine multiple transmissions from the node's transmitter array onto the transmission fiber. The optical demultiplexer is used to separate the signal from the receiving fiber to the receiver array.

We close this overview of the $FT^{\Lambda} - FR^{\Lambda}$ AWG network architecture by noting its

implications on the transmission of unicast and multicast packets. A unicast packet, i.e., a packet that is destined to one destination node, requires one transmission on the wavelength that is routed to the AWG output port that the destination node is attached to.

Now consider a multicast packet, i.e., a packet that is destined to two or more destination nodes. If all destination nodes are attached to the same AWG output port, then only one transmission is required on the wavelength routed to that AWG output port. The splitter locally broadcasts the transmission to all attached nodes, including the intended destination nodes. On the other hand, if the destination nodes of a given multicast packet are attached to different AWG output ports, transmissions on multiple wavelengths routed to the different AWG output ports are required. As discussed in the next section in more detail, these multiple transmissions can be conducted in parallel using multiple transmitters in the source node's transmitter array at the same time.

2. MAC Protocol

In this section we develop a MAC protocol employing pre-transmission coordination together with global scheduling to coordinate the access of the nodes to the shared wavelength channels in the $FT^{\Lambda} - FR^{\Lambda}$ AWG network. This coordination and scheduling are generally recommended strategies for achieving good throughput-delay performance in shared-wavelength single-hop star networks [39]. Time is divided into frames; each frame consists of a control phase and a data phase, as illustrated in Fig. 15. The length of each control packet measured in time is one slot. One control packet is generated for each data packet. The control packet contains the address of the destination node for unicast packets or the multicast group address for multicast packets.

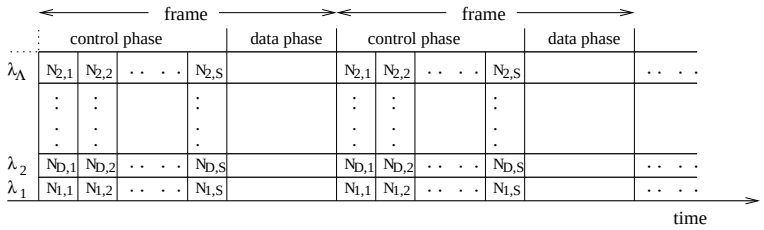

Figure 15. Frame structure and control packet reception schedule for nodes at AWG output port 1 of network with $R = 1$ FSR

We develop two control packet transmission strategies: time-division multiple access (TDMA) and contention similar to slotted Aloha. With either strategy, the periodic wavelength routing property of the AWG requires a transmitting node to use all of the wavelengths covering at least one FSR in order to reach all of the AWG output ports. The spatial wavelength reuse property also allows nodes attached to different ports of the AWG to use the same set of wavelengths without channel collision.

2.1. TDMA control packet transmission. The TDMA sequence for control packet transmission in an AWG network with *one* FSR ($R = 1$) is as follows: In the first slot of the control phase, one node from each input port of the AWG, say the first node $N_{d,1}$ at each port $d = 1, 2, \ldots, D$, transmits its control packet. Each node uses its full array of fixed transmitters for high-speed control packet transmission (in contrast to the lower speed signaling with spreading and spectral slicing employed in the single transceiver network [54, 40]). In the second slot, another node from each AWG input port, say the second node $N_{d,2}$ at each port $d = 1, 2, \ldots, D$, transmits its control packet. This continues until all of the nodes have transmitted their control packets. Fig. 15 shows the corresponding control packet reception schedule by the receiver array of the nodes at AWG output port 1, the reception schedules for the other output ports are analogous. Note that the control

packets do not need to carry the source address, as the source node address can be inferred from the reception schedule. The control phase is S slots long. (Recall that $S = N/D$ and $\Lambda = D \cdot R$. In the considered case $R = 1$ we have $\Lambda = D$ and thus $S = N/\Lambda$.)

In the case of a network with R FSRs, we split the nodes attached to each AWG port into R subgroups. Each subgroup is given a different FSR for the transmission of the control packets. Thus we have R nodes from each input port simultaneously transmitting control packets, each node using all wavelengths in one of the R FSRs. The control packet reception schedule for the nodes at AWG output port 1 of a $R = 2$ FSR network is shown in Fig. 16.

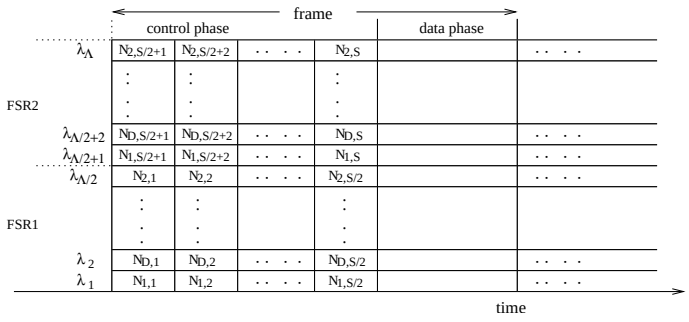

Figure 16. Frame structure and control packet reception schedule for nodes at AWG output port 1 of network with $R = 2$ FSRs

In general, the length of the control phase with the TDMA transmission strategy is S/R slots. Note however that $S = N/D$ and $R = \Lambda/D$ results in a constant control phase length of N/Λ slots, independent of the number of FSRs R. In other words, the length of the control phase depends only on the number of nodes N and the number of transceivers Λ at each node. Consequently, in our performance evaluations in Section 4 we do not need to explicitly include the control phase when considering scenarios with TDMA control packet transmission with fixed N and Λ. When comparing scenarios with different TDMA control

phase lengths N/Λ or control packet contention, we take the different lengths of the control phase into consideration.

2.2. Control Packet Transmission with Contention. With the contention control packet transmission strategy, the control packets are transmitted similar to slotted Aloha. In a network with $R = 1$ FSR, each node sends the control packet uniformly and randomly in one of the slots of the M, $M \leq N/\Lambda$, slot long control phase using its full array of transmitters. In the case of multiple FSRs connecting each input-output port pair,

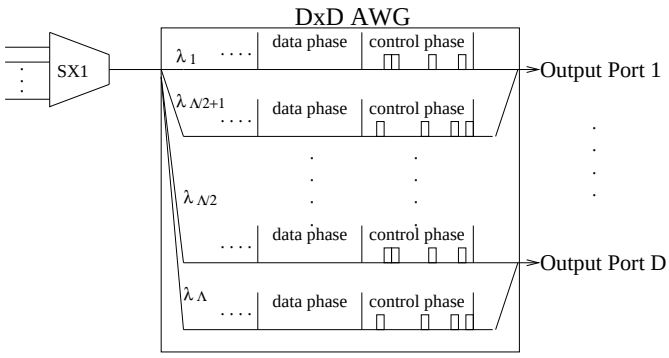

Figure 17. Control packet contention and frame structure for network with $R = 2$ FSRs; the control phase is M slots long

the transmitting node picks from one of the FSRs randomly and uniformly to transmit the control packet in a uniformly randomly chosen slot on all wavelengths in the selected FSR, as illustrated in Fig. 16 for $R = 2$.

A collision occurs when two or more nodes select the same control slot (in the same FSR). Since the transmitter uses all the wavelength of one full FSR and the receiver arrays cover all of the wavelengths, the transmitting node knows the results of control contention after a delay of the one-way end-to-end propagation delay. The nodes with collided control packets retransmit the control packet in the following frame.

Note that for the control packet contention, the control packet need to contain the address of the source node in addition to the addresses of the destination nodes.

We also note that in the $FT^{\Lambda} - FR^{\Lambda}$ AWG network, the R wavelengths (and corresponding receivers) connecting a given AWG input port with a given AWG output port are only shared by the transmissions between nodes attached to these two ports. Thus, the network allows for the development of contention based MAC protocols where control packets are only sent to the AWG output port(s) with attached receivers. Such protocols would have the advantage that typically fewer lasers are required for a control packet transmission compared to our protocol where control packets are transmitted to all output ports using all lasers in one FSR. One drawback of such protocols would be that the sending node does not necessarily receive a copy of a sent control packet. Thus, explicit acknowledgements would be required to verify whether a control packet collision occurred; these acknowledgements would result in increased protocol complexity and delay. Along the same line, the $FT^{\Lambda} - FR^{\Lambda}$ AWG network allows for the development of MAC protocols where the data packets content directly for the R wavelength channels connecting a given AWG input-output port pair without pre-transmission coordination. Such uncoordinated data packet contention however would tend to result in a significant waste of bandwidth due to data packet collisions [39].

2.3. Data Packet Scheduling. Once the control packets of a given control phase are received, all nodes execute the same scheduling algorithm. For a unicast packet, as well as for a multicast packet with all destination nodes attached to one AWG output port, a single packet transmission is scheduled. For a multicast packet with destination nodes at multiple AWG output ports, multiple packet (copy) transmissions are scheduled: one copy is

transmitted to each AWG output port with attached multicast destination nodes. For each unicast and multicast packet (copy) transmission, a wavelength is assigned on a first-come-first-served (FC–FS) basis starting with the lowest FSR in the immediate frame. We adopt the FC–FS scheduling since scheduling algorithms for high-speed WDM networks need to be of low complexity [45]. If the FSRs of the immediate frame are scheduled, then slots in the subsequent frame are assigned, and so on, up to a pre-specified scheduling window. If the data packet corresponding to a control packet can not be scheduled within the scheduling window, the control packet fails. The sending node is aware of the failed control packet as it executes the same scheduling algorithm and retransmits the failed control packet in the next frame. Note that unfairness among the nodes may arise with the FC–FS scheduling if the control packets are transmitted (and received) in the fixed TDMA sequence. To overcome this problem, the received control packets can be randomly resequenced before the scheduling commences. Control packet contention also ensures fairness since the control packets are transmitted in randomly selected slots.

Note that the data packets are buffered in the electronic domain at each source node which can have quite large memory capacity. An arriving packet that finds the node buffer full is dropped and is indicative of congestion. We leave traffic congestion management to the upper layer protocols.

3. Throughput-Delay Analysis Based on Virtual Queue Model

In this section we develop a probabilistic virtual queue based model to evaluate the throughput-delay performance of the $FT^{\Lambda} - FR^{\Lambda}$ AWG network. We assume in this model that the nodal buffers are sufficiently large (infinite in the model) such that only a negligible fraction of the packets is dropped. We demonstrate in Section 5 that a reasonably small

packet buffer at each node is sufficient to achieve packet drop rates of 10^{-2} and less, which in turn implies a correspondingly small modelling error due to the infinite buffer assumption. We also note that throughout we study the network for stable operation, as detailed in Section 3.4.

3.1. Overview of Virtual Queue Network Model. We model each AWG input-output port pair as a "virtual" queue. This queue is virtual because there is *no* electronic buffer or optical memory at the AWG. The queue only exists in the electronic memory domain of each node. These virtual queues are illustrated in Fig. 18. The service capacity for a given virtual queue is the number of FSRs R, with each FSR providing a deterministic service rate of one packet per frame.

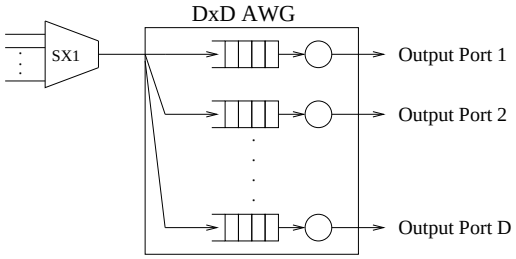

Figure 18. Queuing model: one *virtual* queue for each AWG input-output port pair. Note that there is no physical buffer at the AWG.

We consider the following scenario in our modelling of the $FT^{\Lambda} - FR^{\Lambda}$ AWG network in this section.

- *Bernoulli traffic arrival*: Each node generates a new data packet with probability σ at the beginning of each frame. A given newly generated packet is a unicast packet with probability u and a multicast packet with probability $1 - u$. Let $\sigma_u = \sigma \cdot u$ denote the probability that a new unicast packet is generated in a given frame and

let $\sigma_m = \sigma \cdot (1 - u)$ denote the probability that a new multicast packet is generated in a given frame.

- *Uniform distribution of traffic*: The destination node(s) of a given unicast (multicast) packet are uniformly distributed over all N nodes, including the sending node for mathematical convenience. (Our simulations, which do not allow a node to send to itself, indicate that this simplifying assumption has negligible impact.)

- *Uniform multicast size distribution*: We let Γ, $2 \leq \Gamma \leq N$, represent the maximum number of destination nodes of the multicast packets. The number of destination nodes of a given multicast packet is a random variable γ with $2 \leq \gamma \leq \Gamma$, which is uniformly distributed, i.e., $\gamma \sim U(2, \Gamma)$.

- *Propagation delay*: We initially assume that the propagation delay is negligible. In Section 3.6 we discuss how to incorporate propagation delay in our model.

- *Fixed packet size*: We assume that the data packets are fixed in size. The packet size is such that exactly one data packet fits into the data phase of a given frame.

- *TDMA control packet transmission*: We initially focus on the TDMA control packet transmission. The control packet transmission with TDMA and contention are compared in Section 4.7.

- *Infinite nodal buffers and scheduling window*

To model the multiple transmissions of copies of a multicast packet destined to multiple AWG output ports, we place one packet copy into each corresponding virtual queue. Thus for a multicast packet from a given AWG input port destined to all D AWG

output ports, one packet copy is placed in each of the D virtual queues modelling these D input-output port pairs.

3.2. Definition of Performance Metrics. In our throughput-delay performance evaluation, we consider the following metrics:

- The *multicast throughput* Z_M is defined as the average number of packet transmissions completed per frame in steady state. The transmission of a multicast packet is complete if all copies of the packet have been delivered.

- The *transmitter throughput* Z_T is defined as the average number of packet (copy) transmissions per frame in steady state.

- The *receiver throughput* Z_R is defined as the average number of packets received by their intended destination nodes per frame in steady state. Each intended destination node of a multicast packet copy transmission counts toward the receiver throughput. A given multicast packet copy transmission can result in up to S received packets in case all nodes attached to the splitter are intended destinations.

- The *delay* W_M is the average time in steady state in frames between the following two epochs: (i) the end of the control phase of the frame in which a packet is generated, and (ii) the beginning of the data phase in which the *last* copy of the packet is transmitted.

- The *copy delay* W_{TR} is defined similar to the delay W_M and is the average time between packet generation and the beginning of the transmission of *any given (arbitrary)* copy of the packet.

Note that when only unicast traffic is considered, $Z_M = Z_T = Z_R$ and $W_M = W_{TR}$. Also note that all of these performance metrics are defined with respect to the frame as elementary time unit. This is convenient as for most of our performance studies we consider a network with fixed number of nodes N and fixed number of transceivers Λ per node. For this network, the length of the TDMA control phase N/Λ is constant, which in conjunction with the fixed data phase (data packet size) results in a constant frame length. Toward the end of our performance evaluation, we will study networks with different N and Λ as well as control packet contention and consequently different frame lengths. For those studies we will modify the above definitions and use the slot as elementary time unit. In addition, for all experiments using the slot as time unit, we define the delay as the average period between the packet generation (at the beginning of a frame) and the beginning of the packet transmission, which includes the duration of the control phase.

3.3. Number of Packet Copies. In this section we evaluate the number of packet copy transmissions required to service a given generated packet. Let Δ be a random variable denoting the number of AWG output ports (virtual queues) that lead to destination nodes of a given generated packet. In other words, Δ denotes the number of packet copies that are placed in different virtual queues for a given generated packet. A single packet copy is transmitted if either (i) the generated packet is a unicast packet (which has probability u), or (ii) the generated packet is a multicast packet (which has probability $1 - u$) and all the destination nodes are attached to the same AWG output port. If a multicast has destinations at l, $2 \leq l \leq D$, AWG output ports, then l packet copies are generated and one each is placed in the corresponding virtual queue.

To evaluate the number of packet copies required to service a given generated mul-

ticast packet, we need to find the number of AWG output ports that have at least one destination node of the packet attached. Towards this end, we model the N nodes attached to the D AWG output ports as an urn containing N balls in D different colors, i.e., there are S $(= N/D)$ balls of color $i,\ i = 1, \ldots, D$. Suppose the considered multicast packet has $\gamma,\ 2 \leq \gamma \leq \Gamma$, destinations. To determine the number of packet copy transmissions we draw γ balls (each representing a destination node) from the urn without replacement. (An urn model with replacement which is a simpler, less accurate model of the multicasting is developed in [14]. In Appendix B we examine the differences between the urn models with and without replacement.) We consider the outcome of the drawing without replacement and study formally the following events:

$C_{k_1,\ldots,k_D}$ = "Event that among γ balls drawn without replacement color 1 occurs k_1 times, color 2 occurs k_2 times, ..., color D occurs k_D times with $k_1 + \cdots + k_D = \gamma$".

The probability of this event is given by the polyhypergeometric distribution [61], which can easily be obtained from the hypergeometric distribution [27], as follows

$$P(C_{k_1,\ldots,k_D}) = \frac{\binom{S}{k_1} \cdots \binom{S}{k_D}}{\binom{N}{\gamma}}. \tag{5.1}$$

The family of events

$$C_{k_1,\ldots,k_D}\ (0 \leq k_i \leq \gamma \wedge S; i = 1, \ldots, D; \sum_{i=1}^{D} k_i = \gamma) \tag{5.2}$$

forms a complete system of independent events. Thus

$$P\left(\bigcup_{\substack{0 \leq k_i \leq \gamma \wedge S; 1 \leq i \leq D \\ \sum_{i=1}^{D} k_i = \gamma}} \{C_{k_1,\ldots,k_D}\} \right) = \sum_{\substack{0 \leq k_i \leq \gamma \wedge S; 1 \leq i \leq D \\ \sum_{i=1}^{D} k_i = \gamma}} P(C_{k_1,\ldots,k_D}) = 1. \tag{5.3}$$

Note that we denote $x \wedge y := \min(x, y)$. In our model, the number Δ of required packet copy transmissions corresponds to the number of distinct colors among the γ balls drawn

without replacement. Towards the evaluation of the distribution of Δ, we define the set of color number vectors

$$\begin{aligned} A_\gamma^l &= \{(k_1,\ldots,k_D)\in\{0,\ldots,S\wedge\gamma\}^D | \exists k_{i_1},\ldots,k_{i_l}; i_s\in\{1,\ldots,D\}, 1\le s\le l \\ &\quad \text{with } k_{i_s}\ge 1 \text{ and } \sum_{s=1}^{l} k_{i_s}=\gamma; k_r=0 \text{ for } r\neq i_s, 1\le r\le D\} \end{aligned} \tag{5.4}$$

for $1\le l\le D\wedge\gamma$. Intuitively, this is the set of all color number vectors $(k_1,\ldots,k_D)$ such that there are l distinct colors among the drawn γ balls. The probability that the number Δ of required packet copy transmissions for a given multicast packet with γ destinations is l is then given by

$$P(\Delta=l|\gamma) = \sum_{(k_1,\ldots,k_D)\in A_\gamma^l} P(C_{k_1,\ldots,k_D}). \tag{5.5}$$

Noting that there are $\binom{D}{l}$ ways of choosing l colors out of the D colors (i.e., choosing l destination ports out of all D AWG output ports), we obtain

$$P(\Delta=l|\gamma=n) = \binom{D}{l} \sum_{\substack{1\le k_1,\ldots,k_l\le n\wedge S;\\ \sum_{i=1}^{l} k_i=n}} \frac{\binom{S}{k_1}\cdots\binom{S}{k_l}}{\binom{N}{n}}, \tag{5.6}$$

which can be readily computed via recursion, as detailed in Appendix A.

Note that we have calculated in (5.6) the conditional probability of the event that the number of required packet copies is l given that the generated multicast packet has γ destination nodes, i.e.,

$$P(\Delta=l;\gamma \text{ dest. nodes}) = P(\Delta=l|\gamma \text{ dest. nodes})\cdot P(\gamma \text{ dest. nodes})\cdot P(\text{multicast}) \tag{5.7}$$

with $P(\gamma \text{ dest. nodes}) = \frac{1}{\Gamma-1}$ and $P(\text{multicast}) = 1-u$.

As noted above, a single packet copy is transmitted if either a unicast packet is generated or the generated multicast packet has all $\gamma, 2\le\gamma\le S\wedge\Gamma$, destination nodes

attached to the same AWG output port, i.e.,

$$\begin{aligned} P(\Delta = 1) &= P(\text{“gen. unicast pkt“}) + P(\text{“gen. multicast pkt has all dest. at one port“}) \\ &= u + \frac{(1-u)}{(\Gamma-1)} \sum_{\gamma=2}^{S \wedge \Gamma} P(\Delta = 1|\gamma) \qquad (5.8) \\ &= u + \frac{(1-u)}{(\Gamma-1)} \sum_{\gamma=2}^{S \wedge \Gamma} \frac{D\binom{S}{\gamma}\binom{S}{0}\cdots\binom{S}{0}}{\binom{N}{\gamma}} \\ &= u + \frac{(1-u)D}{\Gamma-1} \sum_{\gamma=2}^{S \wedge \Gamma} \frac{S!(N-\gamma)!}{(S-\gamma)!N!}. \qquad (5.9) \end{aligned}$$

The probability that a given generated packet has destinations at l, $2 \leq l \leq D$, AWG output ports, i.e., requires l packet copy transmissions, is

$$P(\Delta = l) = \frac{(1-u)}{\Gamma-1} \sum_{n=2}^{\Gamma} P(\Delta = l|\gamma = n). \qquad (5.10)$$

We obtain the expected number of required packet copy transmissions as

$$\begin{aligned} E(\Delta) &= P(\Delta = 1) + \sum_{l=2}^{D} lP(\Delta = l) \\ &= u + \frac{(1-u)D}{\Gamma-1} \sum_{\gamma=2}^{S \wedge \Gamma} \frac{S!(N-\gamma)!}{(S-\gamma)!N!} + \sum_{l=2}^{D} \frac{l(1-u)}{\Gamma-1} \left(\sum_{n=2}^{\Gamma} P(\Delta = l|\gamma = n) \right). \qquad (5.11) \end{aligned}$$

3.4. Analysis of Throughput. In this section we calculate the different throughput metrics and establish the stability condition for the network. There are N nodes in the network, each independently generating a new packet at the beginning of a frame with probability σ. Each generated packet requires on average $E[\Delta]$ packet copy transmissions. Thus, the network load in terms of packet copy transmissions per frame is $N \cdot \sigma \cdot E[\Delta]$ in the long run average. Recalling that the AWG provides $D \cdot \Lambda$ wavelength channels, each providing one data phase per frame, we note that the network is stable if $N \cdot \sigma \cdot E[\Delta] < D \cdot \Lambda$.

For stable network operation (and negligible packet drop probabilities), the number of generated packets in a frame is equal to the number of completed packet transmissions

(including all the required packet copy transmissions) in a frame in steady state. Hence, the multicast throughput is given by

$$Z_M = N \cdot \sigma. \tag{5.12}$$

Similarly, we obtain for the transmitter throughput in steady state

$$Z_T = N \cdot \sigma \cdot E[\Delta]. \tag{5.13}$$

The receiver throughput in steady state is given by

$$Z_R = N \cdot \sigma \cdot \left[u + (1-u)\frac{\Gamma+2}{2} \right], \tag{5.14}$$

because a given multicast packet with a maximum multicast size of Γ is received on average by $(\Gamma+2)/2$ nodes.

3.5. Arrivals to Virtual Queue. In this section we analyze the packet (copy) arrival to a given virtual queue representing a given AWG input-output port pair. That is, we study the arrivals to one (arbitrary) of the D virtual queues illustrated in Fig. 18.

There are $S = N/D$ nodes attached to the considered AWG input port. Each of the S nodes generates traffic mutually independently of the other nodes. Recall that a given node generates a new unicast data packet with probability $\sigma_u = \sigma \cdot u$ at the beginning of a given frame. With probability $1/D$ that packet is destined to the considered virtual queue.

Next, recall that a given node generates a new multicast packet with probability $\sigma_m = \sigma \cdot (1-u)$ at the beginning of a frame. The number of destination nodes γ is uniformly distributed over $(2, \Gamma)$ and the individual destination nodes are uniformly distributed over the network nodes, (and consequently AWG output ports and thus virtual queues). Given a multicast packet with γ destination nodes, we need to evaluate the probability that a packet

copy is placed in the considered virtual queue. To evaluate this probability we consider now a fixed virtual queue, say the queue associated AWG output port 1, or equivalently, color 1 in the urn model. The event that the multicast packet has at least one destination at AWG output port 1 corresponds to the events $C_{k_1,\ldots,k_D}$ with $0 < k_1 \leq \gamma \wedge S; 0 \leq k_i \leq \gamma \wedge S$; for $i = 2, \ldots, D$, and $\sum_{i=1}^{D} k_i = \gamma$, in our urn model (5.1). Thus, the probability that a given multicast packet with γ destinations has at least one destination at the considered AWG output port is

$$
\begin{aligned}
& P(\text{“multicast pkt w. } \gamma \text{ dest. has copy to queue 1''}) \\
&= P\left(C_{k_1,\ldots,k_D}(0 < k_1 \leq \gamma \wedge S; 0 \leq k_i \leq \gamma \wedge S; i = 2, \ldots, D; \sum_{i=1}^{D} k_i = \gamma)\right) \\
&= 1 - P\left(C_{k_1,\ldots,k_D}(k_1 = 0; 0 \leq k_i \leq \gamma \wedge S; i = 2, \ldots, D; \sum_{i=2}^{D} k_i = \gamma)\right) \\
&= 1 - \sum_{\substack{0 \leq k_2,\ldots,k_D \leq \gamma \wedge S \\ \sum_{i=2}^{D} k_i = \gamma}} \frac{\binom{S}{k_2} \cdots \binom{S}{k_D}}{\binom{N}{\gamma}} \qquad (5.15) \\
&= 1 - \frac{(N-\gamma)!(N-S)!}{(N-\gamma-S)!N!} \sum_{\substack{0 \leq k_2,\ldots,k_D \leq \gamma \wedge S \\ \sum_{i=2}^{D} k_i = \gamma}} \frac{\binom{S}{k_2} \cdots \binom{S}{k_D}}{\binom{N-S}{\gamma}} \qquad (5.16) \\
&= 1 - \frac{(N-\gamma)!(N-S)!}{(N-\gamma-S)!N!}. \qquad (5.17)
\end{aligned}
$$

Note that we obtained (5.17) by noting that the sum in (5.16) is over a complete set of events. Now considering jointly the possibilities that a generated packet is a unicast packet or a multicast packet, the probability that a given node generates a packet (copy) for the considered queue in a given frame is

$$\sigma_q = \frac{\sigma u}{D} + \frac{\sigma(1-u)}{\Gamma - 1} \sum_{\gamma=2}^{\Gamma} \left(1 - \frac{(N-\gamma)!(N-S)!}{(N-\gamma-S)!N!}\right). \qquad (5.18)$$

Let A be a random variable denoting the number of packet (copy) arrivals to the considered virtual queue in a given frame. Let $a_i = P[A = i]$, $i = 0, 1, \ldots, S$, denote the

distribution of A. Clearly with S independent nodes generating traffic for the considered queue,

$$a_i = \binom{S}{i} \cdot \sigma_q^i \cdot (1-\sigma_q)^{(S-i)}, \tag{5.19}$$

for $0 \leq i \leq S$ and $a_i = 0$ for $i > S$. We remark that the average number of packet copies generated by the S nodes attached to a given AWG input port in a frame equals the average number of packet copies arriving to the D virtual queues connecting the input port to the D AWG output ports in a frame, i.e., $S \cdot \sigma \cdot E[\Delta] = S \cdot \sigma_q \cdot D$, which gives a convenient alternative expression for $E[\Delta]$.

3.6. Queuing Analysis of Virtual Queue. In this section we conduct a queueing analysis of the virtual queue to determine the expected queue length and subsequently the different delay metrics. We begin our formulation by first noting that the arrival process is independent from the state of the queue. Second, we note that the arrival process in frame $t+1$ denoted by A_{t+1} is independent of the arrival process A_t in the prior frame t. Let X_t denote the number of packet (copies) in the queue at the beginning of a given frame t before the new packets are generated for the frame. We impose a maximum virtual queue occupancy J for calculation convenience and set it so large that boundary effects are negligible, i.e., the occupancy J is not reached for stable operation. In each frame up to R packets are served, i.e., $X_{t+1} = \min[(X_t + A_t - R)^+,\ J]$, where $(x)^+ = \max(0, x)$. Thus, $(X_t)_{t\geq 0}$ is a Markov chain with state space $\mathcal{E} := \{0, 1, \ldots, J\}$ and the following transition matrix $\mathbf{P} = (p(x,y))_{x,y\in\mathcal{E}}$ with

$$p(x,0) = \begin{cases} \sum_{i=0}^{R-x} a_i \text{ , for } x \leq R \\ \qquad 0 \text{ , for } R < x \leq J \end{cases} \tag{5.20}$$

and

$$p(x,y) = \begin{cases} a_{R+y-x} \text{ , for } x \leq R+y \\ 0 \text{ , for } x > R+y \end{cases} \tag{5.21}$$

for $0 < y \leq J-1$ and

$$p(x,J) = P(A \geq R+J-x) = \sum_{i=R+J-x}^{N} a_i. \tag{5.22}$$

From (5.20)–(5.22) it follows that $\mathbf{P}$ is an aperiodic and irreducible transition matrix, hence the Markov chain has an unique stationary probability distribution $\pi = [\pi_0, \pi_1, \ldots, \pi_J]$ on $\mathcal{E}$ with $\pi = \pi\mathbf{P}$.

The expected queue length $E[X]$ is given by

$$E[X] = \sum_{j=1}^{J} j \cdot \pi_j. \tag{5.23}$$

We apply Little's theorem to find the mean copy delay

$$W_{TR} = \frac{E[X]}{S \cdot \sigma_q}. \tag{5.24}$$

To analyze the mean delay W_M we need to consider the longest among the Δ virtual queues that a packet copy is placed in for a given generated packet. This analysis is complicated by the fact that multicasts with multiple packet copies destined to multiple queues in parallel tend to introduce correlations among the D virtual queues associated with a given AWG input port. Whereby, the larger the number of packet copies Δ, the stronger the correlation. If $\Delta = D$ with a high probability then the D virtual queues behave essentially identically.

For the analytical evaluation of W_M we need to note that the queueing model developed in this section considers a given virtual queue in isolation, i.e., independently of the other $D-1$ queues associated with the considered AWG input port. To evaluate W_M

based on the developed queueing model we employ the following heuristic. If Δ is below a threshold $\kappa \cdot D$ ($< D$), then we evaluate the longest queue with the order statistics of Δ independent virtual queues. If Δ is above the threshold $\kappa \cdot D$, then we approximate the longest queue by the length of one given independent virtual queue.

More formally, let $\hat{X}$ be a random variable denoting the number of packet copies in the longest queue that a given multicast feeds into in a given frame in steady state. Let $X_{[\delta]}$ be a random variable denoting the longest among $\Delta = \delta$ (independent) queues in steady state. From order statistics we obtain that approximately

$$P(X_{[\delta]} = j) = \delta \cdot \left[\sum_{l=1}^{j} \pi_l \right]^{\delta - 1} \cdot \pi_j. \tag{5.25}$$

Hence, approximately

$$\begin{aligned} E[\hat{X}] &= \sum_{\delta=1}^{\kappa \cdot D} \left(\sum_{j=1}^{J} j \cdot P(X_{[\delta]} = j) \right) \cdot P(\Delta = \delta) \\ &\quad + E[X] \cdot \sum_{\delta=\kappa \cdot D+1}^{D} P(\Delta = \delta), \end{aligned} \tag{5.26}$$

where we assume that $\kappa \cdot D$ is an integer. Applying Little's theorem, we obtain the approximate mean multicast delay

$$W_M = \frac{E[\hat{X}]}{S \cdot \sigma_q}. \tag{5.27}$$

So far we have assumed that the propagation delay in the network is negligible. We now outline how to incorporate propagation delay into our model. We assume that all nodes are equidistant from the central AWG (which can be achieved with fiber delay lines). We let τ denote the one-way end-to-end propagation delay in frames. We assume that the delay incurred for computing the data packet schedule is negligible (if significant, this delay could also be accounted for by τ). In the network with TDMA control packet transmission and infinite scheduling window considered in this section, each packet incurs a delay of τ from

Table 2. Network parameters and their default values

N	# of nodes in network	200
D	degree (# of ports) of AWG	1,2,4,8
R	number of utilized FSRs	1,2,4,8
Λ	$= D \cdot R$, # of wavelengths = # of transceivers in node	8
σ	packet generation probability	
u	fraction of unicast traffic	1, 0, 0.8
$1-u$	fraction of multicast traffic	0, 1, 0.2
Γ	max # of dest. of multicast pkt	5, 15, 50, 200

its generation until the receipt of the corresponding control packet by all nodes and the successful scheduling of the data packet (copies). During this delay period the data packet needs to be stored in the node (which we account for in the node buffer dimensioning in Section 5) and can not yet be serviced. The data packet (copies) then incur the delays W_M (W_{TR}) calculated above from the time the transmission schedule has been computed until the last (any arbitrary) packet copy commences its transmission. A given data packet copy incurs a transmission delay equal to the duration of the data phase (which we may roughly approximate by one frame) and a propagation delay τ for the propagation to the destination node. Thus, we need to add $2 \cdot \tau + 1$ frames to the queueing delays W_M and W_{TR} calculated above in order to account for the propagation delay.

4. Throughput-Delay Performance Results

In this section we numerically study the throughput-delay performance of the $FT^{\Lambda} - FR^{\Lambda}$ AWG network for unicast traffic, multicast traffic, as well as a mix of unicast and multicast traffic. Initially, we fix the number of network nodes at $N = 200$ and the number of used wavelengths (transceivers at each node) at $\Lambda = 8$. The network parameters are summarized in Table 2. We assume that the propagation delay is negligible. We plot the

numerical results from the probabilistic analysis (A), as well as simulation results (S). Each simulation was warmed up for 10^5 frames and terminated when the 99% confidence intervals of all performance metrics are less than 1% of the corresponding sample means.

4.1. Unicast Traffic. In Fig. 19 we plot the delay as a function of the throughput for different network configurations with $D \cdot R = \Lambda$ for unicast traffic ($u = 1$). In all these cases, the network has $\Lambda = 8$ wavelengths and $\Lambda = 8$ transceivers at each node. Note that the configuration $(D = 1, R = 8)$ is equivalent to a PSC based network. We observe that the $(D = 8, R = 1)$ network has the largest throughput of up to 64 packets per frame. Due to spatial wavelength reuse the total number of channels for the $(D = 8, R = 1)$ network is $D \cdot \Lambda = 64$. The maximum throughputs for the other three network configurations $(D = 4, R = 2)$, $(D = 2, R = 4)$, and $(D = 1, R = 8)$ are 32, 16, and 8 packets per frame, respectively.

4.2. Multicast Traffic. In Figures 20 and 21 we plot the throughput and delay for multicast traffic ($u = 0$) for the $(D = 8, R = 1)$ and $(D = 1, R = 8)$ networks for different maximum multicast group sizes Γ. We observe that as Γ increases, both network configurations converge to (i) a maximum multicast throughput of 8 packets/frame, and (ii) the maximum receiver throughput of 800 packets/frame. To understand these dynamics consider the transmission of broadcast packets that are destined to all $N = 200$ receivers in both networks. Clearly, in the PSC equivalent $(D = 1, R = 8)$ network at most eight packet transmissions can take place simultaneously, each reaching all 200 receivers. In the $(D = 8, R = 1)$ network, the broadcast of one packet requires the transmission of eight packet copies, one to each AWG output port, and reaching $N/D = 25$ receivers. Thus in both networks the multicast throughput, i.e., the number of completed multicasts per

frame, is 8 packets/frame and the receiver throughput is 1600 packets/frame. Note that in this broadcast scenario the transmitter throughput is 8 packets/frame in the $(D = 1, R = 8)$ network and 64 packets/frame in the $(D = 8, R = 1)$ network.

Now with multicast traffic with a maximum multicast group size of $\Gamma = 200$, a multicast packet has on average 100 destination nodes. The probability that at least one of these destination node is attached to each AWG output port is $P(\Delta = D|\gamma = 100) = 0.98$. Thus it is very likely that D copies of the multicast packet need to be transmitted. In general, when multicasting over the $FT^{\Lambda} - FR^{\Lambda}$ AWG network, there are two effects at work. On one hand, a large AWG degree D increases the spatial wavelength reuse as all Λ wavelengths can be reused at each AWG port. On the other hand, as the multicast group size increases it becomes (for uniformly distributed destination nodes) increasing likely that at least one destination node is located at each AWG output port. The increase in spatial wavelength reuse in the network configuration with larger D is thus compensated by the increase in the number of required packet copy transmissions when the multicast group size is large. There is a net effect gain in the throughput performance whenever the number of required copy transmissions is smaller than the spatial reuse factor D, i.e., when the multicast group size is relatively small or when the destination nodes tend to be co-located at a small number of AWG output ports. Indeed, as we see from Fig. 21, for a maximum multicast group size of $\Gamma = 5$ and a copy delay of 2 frames, the $(D = 8, R = 1)$ network achieves roughly twice the receiver throughput of the $(D = 1, R = 8)$ network.

Note that these multicast dynamics with transceiver arrays are fundamentally different from the dynamics with a single tunable transceiver at each node. In the single transceiver network [40, 35], large multicasts are very difficult to schedule as it becomes increasingly unlikely to find the receivers of all destination nodes to be free at the same

time, resulting in the so-called receiver bottleneck. Hence it is advantageous to partition multicast groups into several smaller subgroups and transmit copies to each subgroup. The increased number of copy transmissions may lead to a channel bottleneck on the PSC which can be relieved by the increased number of wavelength channels obtained from spatial wavelength reuse on the AWG. The increased number of transmissions on these larger number of channels in turn can exacerbate the receiver bottleneck with single transceiver nodes [40, 35].

Returning to multicasting with transceivers arrays, which overcome the receiver bottleneck, we observe from Figures 20 and 21 that the $(D = 8, R = 1)$ network gives larger delays than the $(D = 1, R = 8)$ network for large multicast group sizes. This is because the multiple packet copy transmissions required for large multicast group sizes in the $(D = 8, R = 1)$ network are more difficult to schedule than the single packet transmission in the $(D = 1, R = 8)$ network.

In summary, we find that the $FT^{\Lambda} - FR^{\Lambda}$ AWG network has significantly improved throughput performance compared with an equivalent PSC network for small multicast groups or co-located multicast destinations. For large multicast groups with uniformly distributed destinations the PSC network achieves smaller delays.

4.3. Mix of Unicast and Multicast Traffic. In this section we consider mixes of unicast and multicast traffic, which are likely to arise in metropolitan area networks. Throughout this section we fix the maximum multicast size at $\Gamma = 200$. In Fig. 22 and Fig. 23 we plot the throughput-delay performance of the $FT^{\Lambda} - FR^{\Lambda}$ AWG network for 80% unicast traffic and 20% multicast traffic for different network configurations. For this traffic mix scenario, we consider both the Bernoulli traffic generation described in Section 3.1 as well as self-similar traffic generation. In particular, we generate self-similar

packet traffic with a Hurst parameter of 0.75, by aggregating ON/OFF processes with Pareto distributed on-duration and geometrically distributed off-duration [50]. We observe that with increasing AWG degree D the network achieves significantly larger multicast throughputs while the delay is increased only very slightly (at lower throughput levels). The throughput levels of the $(D = 8, R = 1)$ configuration are approximately three times larger than for the PSC equivalent $(D = 1, R = 8)$ configuration.

This performance improvement is due to the increased spatial wavelength reuse with increased D, which is only to a small degree compensated for by the increased number of multicast packet copy transmission for that typical mixed traffic scenario. In the PSC based network $(D = 1)$ each packet transmission occupies one of the Λ wavelength channels irrespective of whether the packet is a unicast or a multicast packet. In the AWG based network $(D \geq 2)$, each of the Λ wavelength channels can be reused at each AWG port, i.e., D times, and additional copy transmissions are only required when the destination nodes of a given packet are attached to multiple AWG output ports. Thus, a larger AWG degree is overall beneficial when a significant portion of the traffic is unicast traffic.

We also observe from Fig. **??** that for self-similar traffic, the packet delays are somewhat larger compared to the delays for Bernoulli traffic. This is because with self-similar traffic generation, the packets arrive typically in bursts, which result in larger backlogs and longer queuing delays for the packets making up the tail end of a burst. (The impact of the self-similar traffic on the buffer requirements is studied in Section 5.) Nevertheless, the overall performance trends, i.e., generally larger throughput and slightly increased delay at low throughput levels for larger D, are very similar both for Bernoulli and self-similar traffic. Hence, we focus on Bernoulli traffic for the remainder of this section.

In Figures 24 and 25, we plot the receiver throughput-delay performance for 60%

Table 3. Throughput (in packets/frame) and delay (in frames) for $(D = 1, R = 8)$ network for mixed traffic $(u = 0.8)$ with $\Gamma = 200$

σ	Z_M	Z_T	Z_R	W_M	W_{TR}
0.01	2.0	2.0	41.8.0	0.0	0.0
0.02	4.0	4.0	83.7	0.01	0.01
0.035	7.0	7.0	146.5	0.33	0.33
0.040	0.79	8.0	167.5	45.4	45.4

and 90% unicast traffic. Note that the multicast throughput of the $(D = 1, R = 8)$ network is limited to at most 8 packets/frame. We observe that the gap in performance between the PSC based network $(D = 1, R = 8)$ and the AWG based network with $D = 8$ widens as the fraction of unicast traffic increases. For 90% unicast traffic the $(D = 8, R = 1)$ network achieves about 4.5 times the throughput of the $(D = 1, R = 8)$ network; although the receiver throughput level is overall reduced for the larger portion of unicast traffic. Again we observe that the increase in throughput comes at the expense of only a minor increase in delay (nicely visible in Fig. 25 for the $u = 0.6$ scenario in the throughput range from $100 - 280$ packets/frame).

We observe that the accuracy of our probabilistic analysis is overall quite good. The discrepancies between the analytical and simulation results for the delay W_M for larger D are primarily due to the heuristic approximation (5.26) of the occupancy distribution of the longest queue, for which we set $\kappa = 0.75$ throughout this paper.

Tables 3 and III show the detailed throughput-delay performance metrics obtained from simulation for the scenario with 80% unicast and 20% multicast traffic for the $(D = 1, R = 8)$ and $(D = 8, R = 1)$ network configurations. The stability limit (capacity) for the $(D = 8, R = 1)$ network is $D \cdot \Lambda = 64$ packets per frame. We observe from Table III that for a packet generation probability σ of 0.08 and less, corresponding to a transmitter throughput Z_T of 37.4 or less, or equivalently less than 58% of the capacity, the delays are

Table 4. Throughput (in packets/frame) and delay (in frames) for $(D = 8, R = 1)$ network for mixed traffic $(u = 0.8)$ with $\Gamma = 200$

σ	Z_M	Z_T	Z_R	W_M	W_{TR}
0.01	2.0	4.6	38.6	0.06	0.04
0.02	4.0	9.4	79.7	0.12	0.08
0.04	8.0	18.8	160.4	0.28	0.21
0.08	15.9	37.4	319.8	0.86	0.67
0.125	25.0	58.7	501.9	6.38	5.28
0.135	27.0	63.9	549.0	150.5	127.3

Table 5. Multicast throughput Z_M (in packets/frame) for delay W_M of 4 frames

(D, R)	$(u = 1)$	$(u = 0)$ $\Gamma = 5$	$(u = 0)$ $\Gamma = 15$	$(u = 0)$ $\Gamma = 200$	$(u = 0.8)$ $\Gamma = 200$
$(1, 8)$	7.9	7.9	7.9	7.9	7.9
$(2, 4)$	15.8	9.4	8.1	7.6	12.6
$(4, 2)$	31.2	12.3	8.4	7.4	19.7
$(8, 1)$	60.4	18.7	9.2	7.1	26.9

very small. As the load increases to 90% of the capacity and higher, the delays become quite large. We also observe from Table III that for the $(D = 8, R = 1)$ network the average copy delay W_{TR} is for lower loads typically 75% or less of the corresponding delay W_M for completing the transmission of all packet copies.

In Tables 5 and 6 we summarize the results of the network performance for the various AWG configurations for different traffic conditions. The data entries are extrapolated from our simulation results. In Table 5 we fix the delay at 4 frames and record the maximum multicast throughput. In Table 6 we fix the copy delay at 4 frames and record the

Table 6. Receiver throughput Z_R (in packets/frame) for copy delay W_{TR} of 4 frames

(D, R)	$(u = 1)$	$(u = 0)$ $\Gamma = 5$	$(u = 0)$ $\Gamma = 15$	$(u = 0)$ $\Gamma = 200$	$(u = 0.8)$ $\Gamma = 200$
$(1, 8)$	8	27	62	785	160
$(2, 4)$	16	30	66	762	245
$(4, 2)$	31	41	73	750	397
$(8, 1)$	60	64	97	730	490

maximum receiver throughput. We observe that both in terms of multicast throughput and receiver throughput, the $(D = 8, R = 1)$ network outperforms the networks with small D. In general, the performance of the network improves as D becomes larger. This demonstrates the advantages of the spatial wavelength reuse of the AWG. The performance gap narrows for multicast-only traffic as the average number of destination nodes increases, and for the $u = 0$, $\Gamma = 200$ scenario the $(D = 1, R = 8)$ network gives the largest throughputs. However, for mixed unicast and multicast traffic, both the multicast throughput and the receiver throughput improves significantly as D increases. Both the multicast throughput and the receiver throughput for the $(D = 8, R = 1)$ configuration are over 3 times that of the $(D = 1, R = 8)$ PSC network.

4.4. Impact of Number of Transceivers. In this section we study the throughput-delay performance of the $FT^{\Lambda} - FR^{\Lambda}$ AWG network for different numbers of transceivers Λ in each node. Throughout this section we fix the number of network nodes at $N = 200$ and the number of used FSRs at $R = 1$, hence $D = \Lambda$. Recall from Section 1 that the length of the control phase is N/Λ slots, each carrying one control packet. For our numerical evaluations in this and the following sections we consider a control packet length of 2 bytes and a data packet length of 1500 bytes. Thus the length of the control phase varies between 200 slots for (the degenerate case of) $\Lambda = 1$ and 25 slots for $\Lambda = 8$. The corresponding frame length varies between 950 slots and 775 slots. In Fig. 26 we plot the throughput-delay performance for the different Λ $(= D)$. The delay is given in slots and the throughput is given in steady state, i.e., normalized by the ratio of data phase to total frame length. We observe that the throughput for a fixed tolerable delay approximately triples as the number of nodal transceivers Λ doubles. There are two main effects at work here.

On the one hand, the doubled number Λ of transceivers and the doubled wavelength reuse (governed by $D = \Lambda$) together quadruple the network capacity $D \cdot \Lambda$ (maximum number of data packet (copy) transmissions per frame). On the other hand, the increased number of required packet copy transmissions (for the larger D) results in increased delay. Overall, we observe that large throughputs are achieved for small numbers of transceivers Λ due to the extensive wavelength reuse on the AWG.

4.5. Comparison between TT–TR AWG Network and $FT^{\Lambda} - FR^{\Lambda}$ AWG Network. In this section we compare the throughput-delay performance of the $FT^{\Lambda}-FR^{\Lambda}$ AWG network with the TT–TR AWG network employing one tunable transceiver at each node. Specifically, we consider (i) a TT–TR AWG network where the control packets are transmitted with an LED (as in [54, 40]) over the AWG, and (ii) a TT–TR–FT–FR AWG network where the control packets are transmitted over a PSC with a separate FT–FR at each node and the wavelengths on the AWG are available for data transmission all the time. TDMA control packet transmission is employed in all networks. We employ greedy data packet scheduling in the TT–TR AWG networks, which schedules a data packet for transmission to an AWG output port if at least one of the intended receivers at the port is free. This may result in multiple transmissions of a given multicast packet to a given AWG output port. This greedy policy is a reasonable benchmark for our comparisons as it tends to alleviate the receiver bottleneck at the expense of an increased burden on the transmitters, which as we demonstrate in the next section is a reasonable strategy.

In Figure 27, we plot the throughput-delay performances of the two types of TT–TR AWG networks for different (D, R) combinations and compare with the $(D = 1, R = 8)$ $FT^{\Lambda} - FR^{\Lambda}$ AWG network, which gives the worst throughput-delay performance of all

(D, R) combinations for the $FT^{\Lambda} - FR^{\Lambda}$ AWG network. We observe that all configurations of the TT–TR–FT–FR AWG network, which represents the best possible performance of a TT–TR AWG network in that all control is conducted in parallel over the PSC, has significantly lower performance than the worst performing $FT^{\Lambda} - FR^{\Lambda}$ AWG network configuration. The large delays for the TT–TR AWG network are due to the LED control packet transmission which is conducted in cycles of length D frames [54, 40].

4.6. Transceiver Utilization. In this section we study the utilization of the transmitters and receivers in the $FT^{\Lambda} - FR^{\Lambda}$ and TT–TR AWG networks We define the transmitter utilization U_T as the average fraction of time that any given transmitter is busy transmitting data packets in steady state. For the $FT^{\Lambda} - FR^{\Lambda}$ AWG network, clearly $U_T = Z_T/(N \cdot \Lambda) = \sigma \cdot E[\Delta]/\Lambda$. We define the receiver utilization U_R as the average fraction of time that any given receiver is busy receiving data packets in steady state. For the $FT^{\Lambda} - FR^{\Lambda}$ AWG network, clearly $U_R = Z_R/(N \cdot \Lambda) = \sigma \cdot [u + (1-u) \cdot (\Gamma + 2)/2]/\Lambda$.

For the TT–TR AWG network, the transmitter utilization is difficult to compute because with the employed greedy scheduling algorithm, a packet copy destined to multiple receivers attached to the same splitter can be transmitted multiple times depending on receiver availability. The receiver utilization for the TT–TR AWG network is approximately equal to the average number of destinations per packet multiplied by the packet throughput, i.e., $U_R = \sigma \cdot [u + (1-u) \cdot (\Gamma + 2)/2]$.

In Table 7 we compare the average transceiver utilization of the TT–TR–FT–FR AWG and the $FT^{\Lambda} - FR^{\Lambda}$ AWG networks for traffic loads resulting in an average delay of 10,000 slots. We observe that for the considered traffic mix with 80% unicast traffic and 20% multicast traffic, the utilization of the fixed tuned transmitters in the $FT^{\Lambda} - FR^{\Lambda}$

Table 7. Transceiver utilization comparison for mixed traffic ($u = 0.8$) with $\Gamma = 200$ for delay of 10,000 slots

AWG Network	Z_T	U_T	Z_R	U_R
FT–FR–TT–TR ($D = 8, R = 1$)	22	0.11	126	0.63
FT^{Λ} ($D = \Lambda = 1$)	0.7	0.003	15	0.08
FT^{Λ} ($D = \Lambda = 2$)	3.6	0.01	60	0.15
FT^{Λ} ($D = \Lambda = 4$)	15	0.02	191	0.24
FT^{Λ} ($D = \Lambda = 8$)	61	0.04	535	0.33

AWG network is below 4% for all considered configurations. On the other hand, the fixed tuned receivers are fairly well utilized, especially for the configurations with larger D. This suggests to study $TT^i - FR^{\Lambda}$ AWG networks, with $1 \leq i < \Lambda$ in future work. This is further indicated by the utilization of approximately 11 % of the tunable transmitter in the TT–TR–FT–FR AWG network. The tunable receiver in the TT–TR–FT–FR AWG network is heavily utilized, which illustrates the receiver bottleneck in TT–TR AWG networks and also indicates that an array of fixed tuned receivers is a good choice for an AWG based metro network carrying mixed traffic.

4.7. Control Packet Transmission: TDMA vs. Contention. In this section we examine the impact of the TDMA and contention based control packet transmission strategies described in Sections 2.1 and 2.2. We consider the $FT^{\Lambda} - FR^{\Lambda}$ AWG network with $D = 4$, $R = 2$, and $\Lambda = 8$ for a mix of 80% unicast ($u = 0.8$) and 20% multicast traffic with $\Gamma = 200$. The length of the data phase is fixed at 1500 bytes or equivalently 750 slots throughout. In Fig. 28 and Fig. 29 we compare the throughput-delay obtained from simulation for (i) TDMA control packet transmission with a control phase with N/Λ slots, and (ii) control packet transmission with contention with a control phase with $M =$ 5 and 10 slots. We observe from Fig. 28 that for $N = 200$ nodes, TDMA control packet transmission gives better throughput delay performance than control packet contention.

This is because the effect of the slightly shorter control phase with contention is outweighed by the delay introduced due to control packet collisions and subsequent retransmissions. Note that each retransmission introduces an additional delay of one frame, whereby in the considered scenario the frame is significantly longer than the control phase.

Control packet transmission with contention is advantageous when the length N/Λ of the TDMA control phase makes up a significant portion of the frame length, i.e., when either the number of nodes N is large or the data packets are short. We illustrate this effect by scaling up the number of nodes to $N = 2000$ in Fig. 29. We observe that in this scenario, control packet contention with a data phase consisting of $M = 10$ slots gives consistently better throughput-delay performance than TDMA control packet transmission. This is because in this scenario, the effect of the significantly shorter control phase with contention outweighs the effect of occasional control packet collisions and retransmissions. If the number of control slots is too small, then the control packet contention becomes increasingly a bottleneck as the traffic load increases, as illustrated in Fig. 29 for $M = 5$.

5. Node Buffer Dimensioning

In this section we address the problem of dimensioning the buffer in a node. Note that the analysis in Section 3 considered virtual queues, whereby a virtual queue buffers the packet (copies) originating from the nodes attached to a given AWG input port and destined to nodes at a given AWG output port. We introduced the virtual queue as a modelling concept to make the above analysis tractable. In a real network the packets are buffered in node buffers. The dimensioning of these node buffers is important for network dimensioning and resource allocation. The probabilistic modelling of the nodal buffer occupancy is a complex problem due to the sharing of the wavelengths connecting a

given AWG input-output port pair among the nodes connected to the input port and the multiple packet copies required to serve a multicast packet and is left for future work.

We conduct simulations of the $FT^{\Lambda} - FR^{\Lambda}$ network with the buffering at the nodes to determine the packet drop probability P_{loss} which we define as the probability that a newly generated packet finds the nodal buffer full and is dropped. We denote L for the buffer capacity in number of data packets at each node, whereby only one copy of each data packet is stored irrespective of the number of packet copy transmissions required to serve the packet. We consider the network with the default parameters given in Table 2 for a mix of 80% unicast traffic ($u = 0.8$) and 20% multicast traffic with $\Gamma = 200$. We consider both a network with a negligible propagation delay $\tau = 0$ and a network with a propagation delay of $\tau = 94$ frames, which corresponds to a typical scenario with a distance of 48.6 km between a node and the central AWG, a propagation speed of $2 \cdot 10^8$ m/sec, a frame length of 1550 bytes, and an OC48 transmission rate of 2.4 Gbps. Also, we consider both Bernoulli (denoted `ber`) and self-similar (denoted `ssim`) traffic generation. In Fig. 30 and Fig. 31 we plot the packet drop probability P_{loss} at a node as a function of the probability σ that a node generates a new packet in a frame.

We observe that for Bernoulli traffic and a negligible propagation delay, relatively small node buffers with a capacity of 5 or 10 data packets are sufficient to achieve small loss probabilities on the order of 10^{-2} or less for traffic loads close to the stability limit of the networks. Recall from Section 3.4 that the stability limit for the network is $\sigma < D \cdot \Lambda/(N \cdot E[\Delta])$, which is $\sigma < 0.1$ for the considered $D = 4$, $R = 2$ network and $\sigma < 0.14$ for the considered $D = 8$, $R = 1$ network. For a propagation delay of $\tau = 94$ frames, correspondingly larger buffers are needed to store the data packets for which the control packets are propagating through the network. For self-similar traffic and a propagation delay

of $\tau = 94$ frames, yet larger buffers are required to ensure small packet drop probabilities. We observe from Fig. **??**(b), however, that a buffer capable of holding 500 data packets (= 750 kbyte for the considered 1500 byte data packets) is sufficient to ensure loss probabilities below $10^{-3.5}$ for a long run mean packet generation probability of $\sigma = 0.1$ (which for the considered network parameters corresponds to a long run average traffic generation rate of 232 Mbps of the bursty self-similar traffic with Hurst parameter $H = 0.75$).

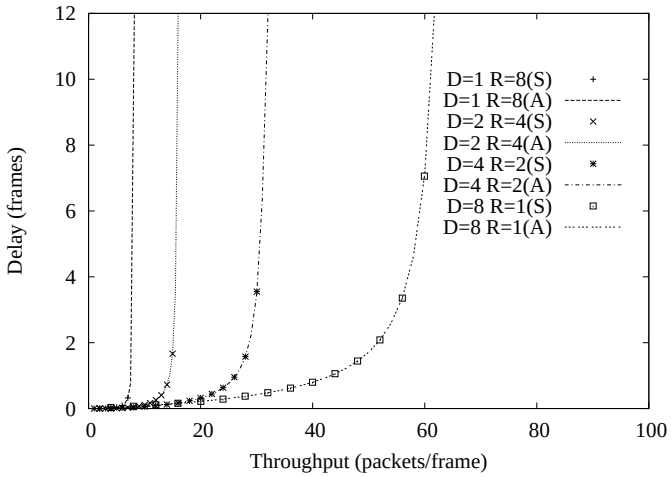

Figure 19. Delay W_M as a function of throughput Z_M for unicast traffic ($u = 1$).

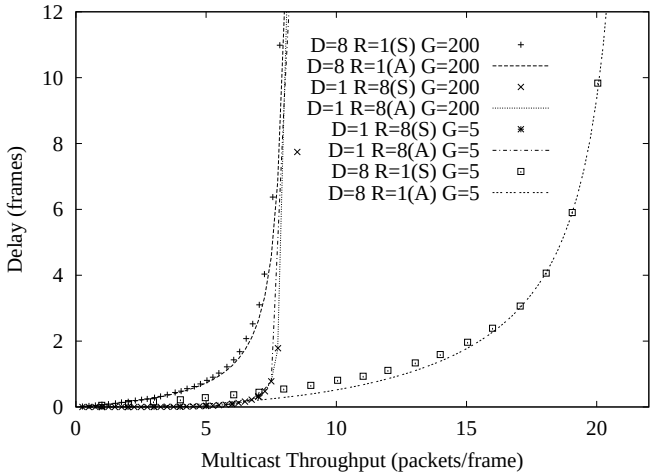

Figure 20. Delay W_M as a function of multicast throughput Z_M for multicast traffic ($1-u = 1$) with $\Gamma = 5$ and $\Gamma = 200$.

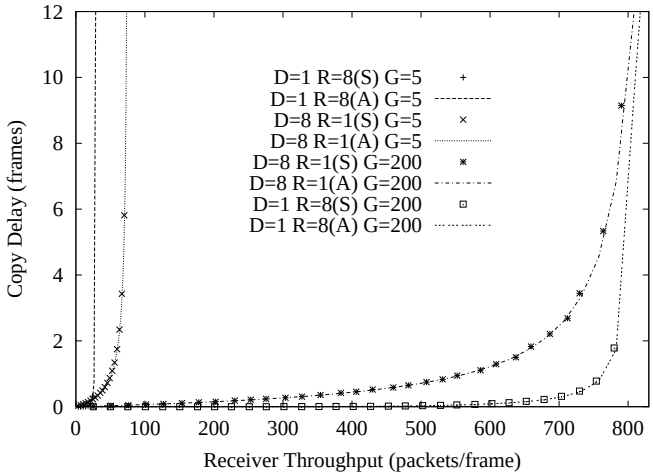

Figure 21. Copy delay W_{TR} as a function of receiver throughput Z_R for multicast traffic ($u = 0$) with $\Gamma = 5$ and $\Gamma = 200$.

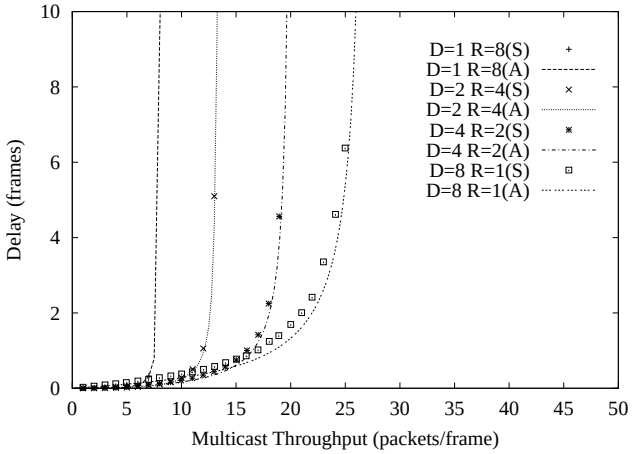

Figure 22. Delay W_M as a function of multicast throughput Z_M for mix of 80% unicast ($u = 0.8$) and 20% multicast traffic with $\Gamma = 200$ with Bernoulli traffic.

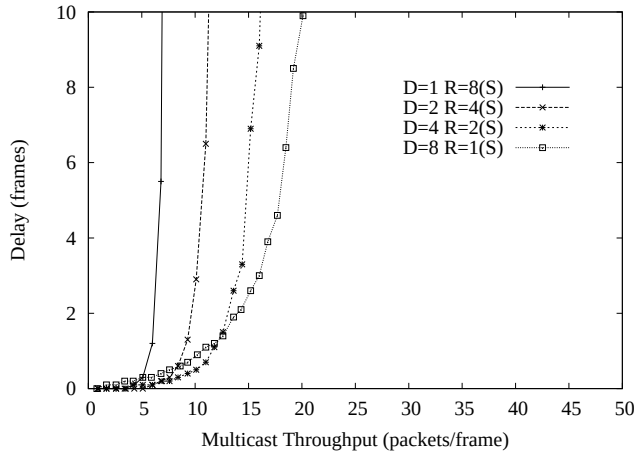

Figure 23. Delay W_M as a function of multicast throughput Z_M for mix of 80% unicast ($u = 0.8$) and 20% multicast traffic with $\Gamma = 200$ with self-similar traffic.

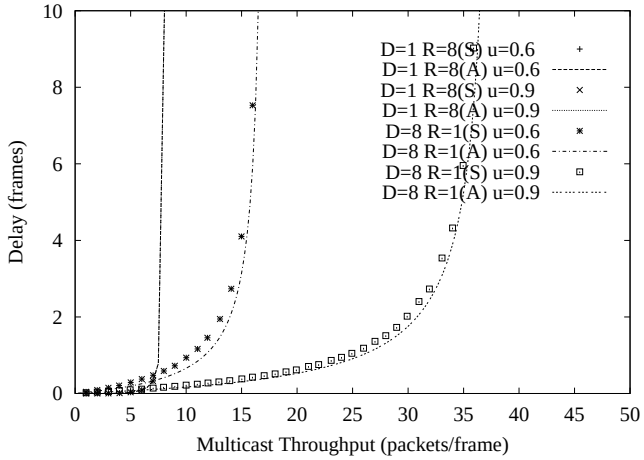

Figure 24. Delay W_M as a function of multicast throughput Z_M for mixed traffic $u = 0.9$ and $u = 0.6$.

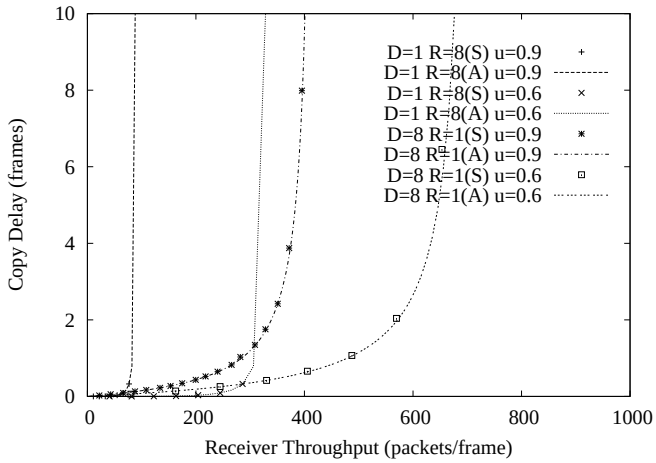

Figure 25. Copy delay W_{TR} as a function of receiver throughput Z_R for mixed traffic $u = 0.9$ and $u = 0.6$.

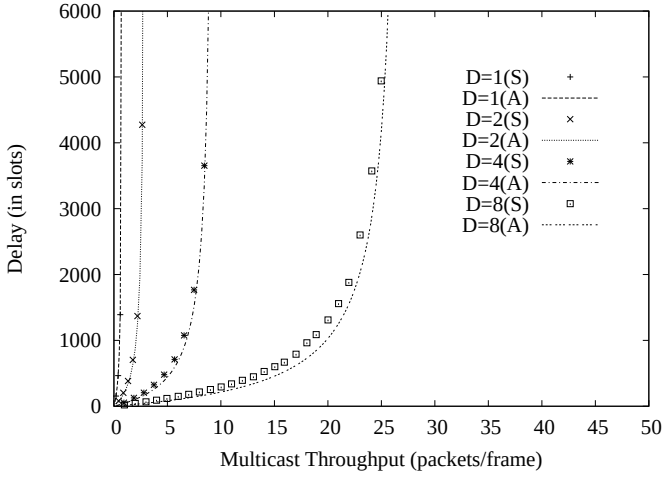

Figure 26. Delay W_M as a function of multicast throughput Z_M for mix of 80% unicast ($u = 0.8$) and 20% multicast traffic with $\Gamma = 200$ for different number of transceivers Λ ($= D$)

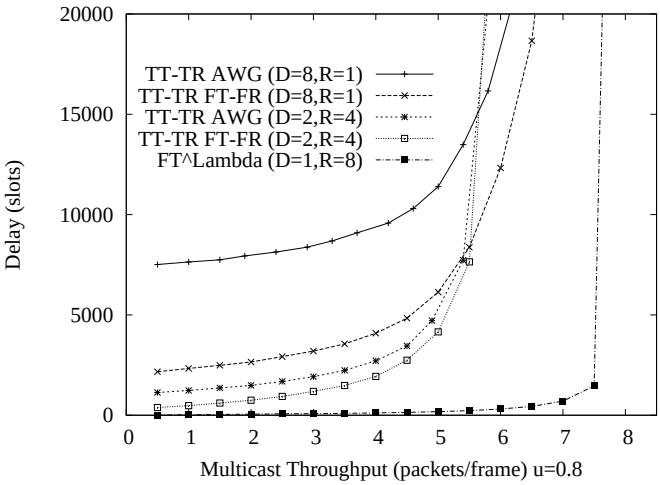

Figure 27. Delay W_M as a function of multicast throughput Z_M for TT–TR AWG, TT–TR–FT–FR AWG, and $FT^{\Lambda} - FR^{\Lambda}$ AWG networks

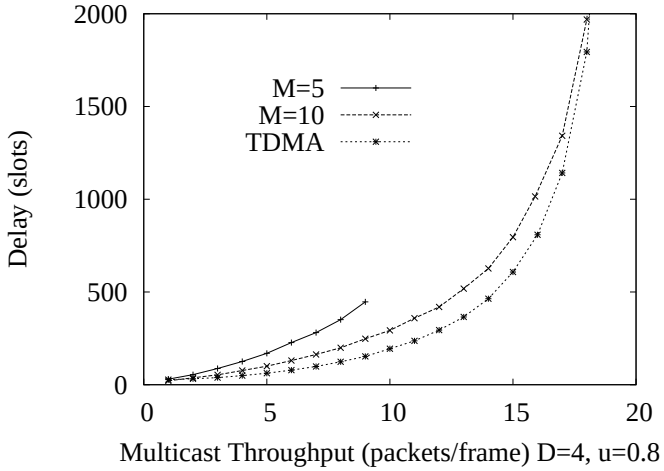

Figure 28. Delay W_M in slots as a function of multicast throughput Z_M for control packet transmission with TDMA (N/Λ slot control phase) and contention (M slot control phase),N = 200

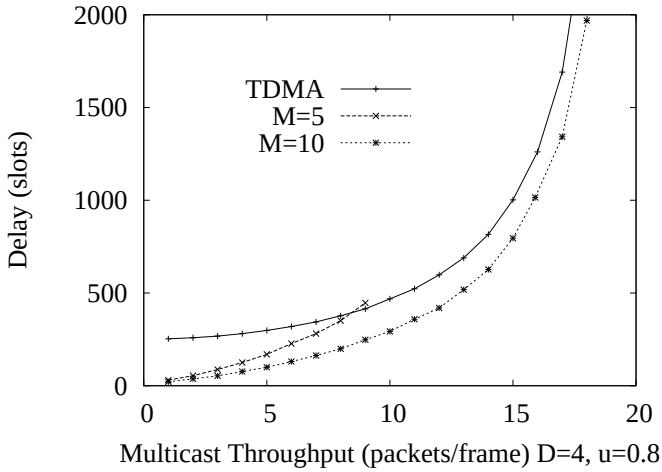

Figure 29. Delay W_M in slots as a function of multicast throughput Z_M for control packet transmission with TDMA (N/Λ slot control phase) and contention (M slot control phase),N = 2000

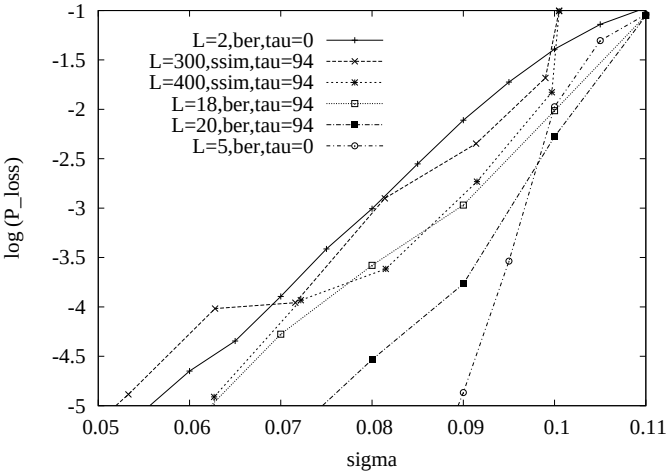

Figure 30. Packet drop probability P_{loss} at node as a function of packet generation probability σ at node for different node buffer capacities L in packets. $D = 4$ AWG ports and R = 2 FSRs

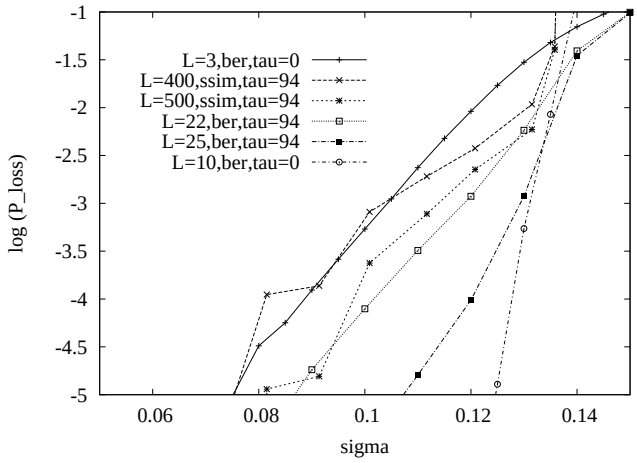

Figure 31. Packet drop probability P_{loss} at node as a function of packet generation probability σ at node for different node buffer capacities L in packets.$D = 8$ AWG ports and $R = 1$ FSRs.

CHAPTER 6

Summary

We note that the devices we selected to overcome the signaling challenges of the AWG can also be exploited to greatly enhance the performance of the network. In the AWG||PSC network, we used the PSC as the signaling device. When both AWG and PSC are functional, the AWG||PSC network uniquely combines the respective strengths of the two devices. This network architecture also addresses the problem of the single point of failure in single–hop WDM networks. By means of analysis and verifying simulations we find that the throughput of the AWG||PSC network is significantly larger than the total throughput obtained by combining the throughput of a stand–alone AWG network with the throughput of a stand–alone PSC network. We also find that the AWG||PSC network gives over a wide operating range a better throughput–delay performance than a network consisting of either two load sharing PSCs in parallel or two load sharing AWGs in parallel.

Another method to overcome the signaling over the AWG is through the use of transceiver arrays. In the $FT^{\Lambda} - FR^{\Lambda}$ AWG network, we proposed an AWG based single–hop metro WDM network with a fixed–tuned transceiver based node architecture. Our analytical and simulation results indicate that the $FT^{\Lambda} - FR^{\Lambda}$ AWG network efficiently supports a typical mix of unicast and multicast traffic. For such a traffic mix the $FT^{\Lambda} - FR^{\Lambda}$ AWG network with an 8×8 AWG achieves about three times the throughput of an equivalent

PSC based network. In summary, we have demonstrated that the wavelength reuse property of the AWG makes it a promising device for building wavelength efficient single–hop WDM networks.

REFERENCES

[1] D. Banerjee, J. Frank, and B. Mukherjee. Passive optical network architecture based on waveguide grating routers. *IEEE Journal on Selected Areas in Communications*, 16(7):1040–1050, September 1998.

[2] P. Bernasconi, C. Doerr, C. Dragone, and M. Cappuzzo *et al.* Large $n \times n$ waveguide grating routers. *IEEE/OSA Journal of Lightwave Technology*, 18(7):985–991, July 2000.

[3] A. Bianco, E. Leonardi, M. Mellia, and F. Neri. Network controller design for SONATA — a large–scale all–optical passive network. *IEEE Journ. on Sel. Areas in Comm.*, 18(10):2017–2028, October 2000.

[4] N. P. Caponio, A. M. Hill, F. Neri, and R. Sabella. Single–layer optical platform based on WDM/TDM multiple access for large–scale 'switchless' networks. *European Trans. on Telecomm.*, 11(1):73–82, Jan./Feb. 2000.

[5] M. Cerisola, T.K. Fong, R.T. Hofmeister, and *et al.* Novel distributed slot synchronization technique for optical WDM packet networks. In *OFC 1996 Technical Digest, paper FD1*, pages 314–315, 1996.

[6] C. J. Chae, H. Park, and Y. Park. Hybrid optical star coupler suitable for wavelength reuse. *IEEE Photonic Technology Letters*, 10(2):279–281, February 1998.

[7] E.Y. Chan, Q N. Le, M.W. Beranek, Y. Huang, D.G. Koshinz, and H.E. Hager. A 12–channel multimode fiber–optic 1.0625–gb/s fiber channel receiver based on COTS devices and MCM-L/COB/BGA packaging. *IEEE Photonics Technology Letters*, 12:1549–1551, November 2000.

[8] M-S. Chen, N.R. Dono, and R. Ramaswami. A media–access protocol for packet–switched wavelength division multiaccess metroplitan area networks. *IEEE Journal on Selected Areas in Communication*, 8(6):1048–1057, August 1990.

[9] R. Chipalkatti, Z. Zhang, and A. S. Acampora. Protocols for optical star–coupler network using WDM: Performance and complexity study. *IEEE Journal on Selected Areas in Communications*, 11(4):579–589, May 1993.

[10] K.R. Desai and K. Ghose. An evaluation of communication protocols for star–coupled mutidemnsional WDM networks for multiprocessors. In *Proc. of Int. Conf. on Massive Parallel Processing Using Optical Interconnections '95*, pages 42–48, San Antonio, TX, October 1995.

[11] P. W. Dowd. Random access protocols for high–speed interprocessor communication based on an optical passive star = topology. *IEEE/OSA J. of Lightwave Technol.*, 9(6):799–808, June 1991.

[12] C. Dragone. Optimum design of a planar array of tapered waveguides. *J. Opt. Soc. Amer.*, 7:2081–2093, November 1990.

[13] C. Dragone, C.A. Edwards, and R.C. Kistler. Integrated optics N×N multiplexer on silicon. *IEEE Photon. Techno. Lett.*, 3(10):896–899, October 1991.

[14] C. Fan, M. Reisslein, and S. Adams. The $FT^{\Lambda}-FR^{\Lambda}$ AWG network: A practical single-

hop metro WDM network for efficient uni- and multicasting. In *IEEE INFOCOM '04*, Hong Kong, March 2004.

[15] A. Ganz and Y. Gao. Time–wavelength assignment algorithms for high performance WDM star based systems. *IEEE Trans. on Commun.*, 42(2/3/4):1827–1836, Feb./March/April 1994.

[16] K. Ghose, R.K. Horsell, and N.K. Singhvi. Hybrid multiprocessing using WDM optical fiber interconnections. In *Proc. of the First Int. Workshop. on Massive Parallel Processing Using Optical Interconnections '94*, pages 182–196, Cancun, Mexico, April 1994.

[17] B. Glance, I. P. Kaminow, and R. W. Wilson. Applications of the integrated waveguide grating router. *IEEE/OSA J. Lightwave Technol.*, 12(6):957–962, June 1994.

[18] M.S. Goodman, H. Kobrinshki, M.P. Vecchi, R.M. Bulley, and J.J. Gimlett. The LAMBDANET multiwavelength network: architecture, application, and demonstration. *IEEE J. Select. Areas in Commun.*, 8:995–1004, August 1990.

[19] I. M. I. Habbab, M. Kavehrad, and C.-E. W. Sundberg. Protocols for very high–speed optical fiber local area networks using a passive star topology. *IEEE/OSA J. of Lightwave Technol.*, LT–5(12):1782–1794, April 1987.

[20] A.M. Hamad and A.E. Kamal. A survey of multicasting protocols for broadcast–and-select single–hop networks. *IEEE Network*, 16:36–48, 2002.

[21] Y. Hibino. An array of photonic filtering advantages. *Circuits and Devices*, pages 23–27, November 2000.

[22] A.M. Hill, M. Brierley, R.M. Percival, R. Wyatt, D. Pitcher, K.M. Ibrahim Pati, I. Hall, and J.-P. Laude. Multi–star wavelength–router network and its protection strategy. *IEEE J. Select. Areas Commun.*, 16(7):1134–1145, September 1998.

[23] A.M. Hill, S. Carter, J. Armitage, and M. Shabeer *et al.* A scalable and switchless optical network structure, employing a single 32×32 free–space grating multiplexer. *IEEE Photonics Technology Letters*, 8(4):569–571, April 1996.

[24] R.T. Hofmeister, C.L. Lu, M.C. Ho, and *et al.* Distributed slot synchronization (DSS): a network–wide slot synchronization technique for packet–switched optical networks. *IEEE/OSA Journal of Lightwave Technology*, 16(12):2109–2116, December 1998.

[25] M. Ibsen, S. Alam, M.N. Zervas, A.B. Grudinin, and D.N. Payne. 8- and 16–channel all-fiber DFB laser WDM transmitters with integrated pump redundancy. *IEEE Photonics Technology Letters*, 11:1114–1116, September 1999.

[26] M.W. Janoska and T.D. Todd. Coupled reservation protocols for hierarchical single–hop photonic networks. *IEE Proceedings Communications*, 144(4):247–255, August 1997.

[27] N.L. Johnson and S. Kotz, editors. *Urn models and their applications.* Wiley, 1977.

[28] J.P. Jue and B. Mukherjee. The advantages of partitioning multicast transmissions in a single–hop WDM network. In *Proc. of ICC '97*, pages 427–431, Montreal, Canada, June 1997.

[29] D.K. Jung, S.K. Shin, C.H. Lee, and Y.C. Chung. Wavelength–division–multiplexed passive optical network based on spectrum–slicing techniques. *IEEE Photonics Technology Letters*, 10(9):1334–1336, September 1998.

[30] B. Kannan, S. Fotedar, and M. Gerla. A two level optical star WDM metropolitan area network. In *Proc. of IEEE Globecom '94 Communications*, San Francisco, CA, November 1994.

[31] M. J. Karol and B. Glance. A collision–avoidance WDM optical star network. *Computer Networks and ISDN Systems*, 26:931–943, March 1994.

[32] K. Kato, A. Okada, Y. Sakai, and K. Noguchi *et al.* 10–Tbps full–mesh WDM network based on a cyclic–frequency arrayed–waveguide grating router. In *Proc. of ECOC '00*, volume 1, pages 105–107, Munich, Germany, September 2000.

[33] T. Kitamura, M. Iizuka, M. Sakuta, Y. Nishino, and I. Sasase. A new partition scheduling alrogithm by prioritizing the transmission of multicast pakcets with less destination address overlap in WMD single–hop networks. In *Proc. GLOBECOM '01*, pages 1469–1473, San Antonio, TX, November 2001.

[34] K. Li. Probabilistic analysis of cyclic packet transmission scheduling in WDM optical networks. In *Proceedings of IEEE International Conference on Parallel Processing*, pages 531–538, March 2000.

[35] H. C. Lin and C. H. Wan. A hybrid multicast scheduling algorithm for single–hop WDM networks. *IEEE/OSA Journal of Lightwave Technology*, 19(11):1654–1664, November 2001.

[36] J.J. Liu, B. Reily, P.H. Shen, N. Das, P. Newman, W. Chang, and G. Simonis. Ultralow–threshold sapphire substrate–bonded top–emitting 850–nm VCSEL array. *IEEE Photonics Technology Letters*, 14:1234–1236, September 2002.

[37] J. Lu and L. Kleinrock. A wavelength division multiple access protocol for high–speed

local area networks with a passive star topology. *Performance Evaluation*, 16(1–3):223–239, November 1992.

[38] M. Maier, M. Reisslein, and A. Wolisz. High performance switchless WDM network using multiple free spectral ranges of an arrayed–waveguide grating. In *Proc. of SPIE Terabit Optical Networking: Architecture, Control, and Management Issues*, pages 101–112, Boston, MA, November 2000. Paper won the *Best Paper Award* of the conference.

[39] M. Maier, M. Reisslein, and A. Wolisz. Towards efficient packet switching metro WDM networks. *Optical Networks Magazine*, 3(6):44–62, November/December 2002.

[40] M. Maier, M. Scheutzow, and M. Reisslein. The arrayed–waveguide grating based single–hop WDM network: an architecture for efficient multicasting. *IEEE Journal on Selected Areas in Communications*, 21(9):1414–1432, November 2003.

[41] M. Maier, M. Scheutzow, M. Reisslein, and A. Wolisz. Wavelength reuse for efficient transport of variable–size packets in a metro WDM network. In *Proc. of IEEE Infocom '02*, pages 1432–1441, New York, NY, June 2002.

[42] K.A. McGreer. Arrayed waveguide gratings for wavelength routing. *IEEE Communications Magazine*, 36(12):62–68, December 1998.

[43] M.W. McKinnon, G.N. Rouskas, and H.G. Perros. Performance analysis of a photonic single–hop ATM switch architecture, with tunable trasmitters and fixed frequency receivers. *Performance Evaluation*, 33(2):113–136, July 1998.

[44] N. Mehravari. Performance and protocol improvements for very high speed optical fiber local area networks using a passive star topology. *IEEE/OSA J. of Lightwave Technol.*, 8(4):520–530, April 1990.

[45] E. Modiano. Random algorithms for scheduling multicast traffic in WDM broadcast–and–select networks. *IEEE/ACM Transactions on Networking*, 7(3):425–434, June 1999.

[46] B. Mukherjee. WDM–based local lightwave networks part I: Single–hop systems. *IEEE Network Magazine*, 6(3):12–27, May 1992.

[47] B. Mukherjee. WDM–based local lightwave networks part I: Single–hop systems. *IEEE Network*, 6(3):12–27, May 1992.

[48] Y. Ofek and M. Sidi. Design and analysis of hybrid access control to an optical star using WDM. In *Proc. of IEEE Infocom '91*, pages 20–31, Bal Harbour, FL, May 1991.

[49] A. Okada, T. Sakamoto, Y. Sakai, and K. Noguchi *et al.* All–optical packet routing by an out–of–band optical label and wavelength conversion in a full–mesh network based on a cyclic–frequency AWG. In *Proc. of OFC 2001 Technical Digest, paper ThG5*, Anaheim, CA, March 2001.

[50] K. Park and W. Willinger, editors. *Self–similar network traffic and performance evaluation.* John Wiley & Sons Inc., 2000.

[51] F. Ruehl and T. Anderson. Cost–effective metro WDM network architectures. In *OFC 2001 Technical Digest, paper WL1*, Anaheim, CA, March 2001.

[52] Y. Sakai, K.Noguchi, R. Yoshimura, and T. Sakamoto *et al.* Management system for full-mesh WDM AWG-STAR network. In *Proc. of ECOC '01.*, volume 3, pages 264 –265, Amsterdam, Netherlands, September 2001.

[53] A.A.M. Saleh and H. Kogelnik. Reflective single–mode fiber–optic passive star couplers. *IEEE/OSA Journal of Lightwave Technology*, 6(3):392–398, March 1988.

[54] M. Scheutow, M. Maier, M. Reisslein, and A. Wolisz. Wavelength reuse for efficient packet–switched transport in an AWG-based metro WDM network. *IEEE/OSA Journal of Lightwave Technology*, 21(6):1435–1455, 2003.

[55] K. V. Shrikhande, I. M. White, M.S. Rogge, F.-T An, and *et al.* Performance demonstration of a fast–tunable transmitter and burst–mode packet receiver for HORNET. In *Exhibit of OFC 2001*, volume 4, pages ThG–1 ThG–3, Anaheim, CA, March 2001.

[56] K. M. Sivalingam. Design and analysis of a media access protocol for star coupled WDM networks with TT–TR architecture. In K. M. Sivalingam and S. Subramaniam, editors, *Optical WDM Networks — Principles and Practice, Chapter 9*. Kluwer Academic Publishers, 2000.

[57] M. J. Spencer and M.A. Summerfield. WRAP: A medium access control protocol for wavelength–routed passive optical networks. *IEEE Journal of Lightwave Technology*, 18(12):1657–1676, December 2000.

[58] G. N. M. Sudhakar, N. D. Georganas, and M. Kavehrad. Slotted Aloha and Reservation Aloha protocols for very high–speed optical fiber local area networks using passive star topology. *IEEE/OSA J. of Lightwave Technol.*, 9(10):1411–1422, October 1991.

[59] M. Tabiani and M. Kavehrad. Theory of an efficient N×N passive optical star = coupler. *IEEE/OSA Journal of Lightwave Technology*, 9(4):448–445, April 1991.

[60] Y. Tachikawa, Y. Inoue, M. Ishii, and T. Nozawa. Arrayed waveguide grating multiplexer with loop–back optical paths and its applications. *Journal of Lightwave Technology*, 14(6):977–984, June 1996.

[61] L Takacs. *Combinatorial methods in the theory of stochastic processes*. Wiley, 1967.

[62] L.A. Wang and K.C. Lee. A WDM based virtual bus for universal commuincation and computing systems. In *Proc. of ICC '92*, pages 888–894, June 1992.

APPENDIX A

NODAL TRANSCEIVER BACK–UP

In this section, we describe the second level of back–up, the transceiver back–up. Although nodal transceiver back–up in single–hop networks is not as critical as in multi–hop networks where the node has to forward packets from other nodes in the network, the proposed MAC protocol takes advantage of the node architecture to enable transceiver back–up.

In the proposed single–hop architecture, we define six states, illustrated in Fig. 32, where the node with malfunctioning transceivers can still communicate. However, not all

Node State	PSC		AWG	
	TT	TR	TT	TR
0	u	u	u	u
1	d	u	u	u
2	u	d	u	u
3	u	u	d	u
4	u	u	u	d
5	d	d	u	u
6	u	u	d	d

u: up, functional
d: down, non–functional

Figure 32. Node status based on transceiver functional status

nodes in any one of the six states can communicate with one another. For example, a node with a malfunctioning PSC TT can not transmit to a node with a malfunctioning AWG TR. The node with malfunctioning PSC TT must transmit using its AWG TT. But if the receiving node's AWG TR is malfunctioning, there is no way to setup a communication path. Conversely, a node with a malfunctioning AWG TT can not transmit to a node with a malfunctioning PSC TR. The communication matrix for the 6 states is depicted in Fig. 33. (For multi–hop networks, a node with any combination of one or more operating

Originating Node State		Destination Node State						
		0	1	2	3	4	5	6
0	control	P	P	A	P	P	A	P
	data	P/A	P/A	A	P/A	P	A	P
1	control	A	A	A	A	X	A	X
	data	A	A	A	A	X	A	X
2	control	P	P	A	P	P	A	P
	data	P/A	P/A	A	P/A	P	A	P
3	control	P	P	X	P	P	X	P
	data	P	P	X	P	P	X	P
4	control	P	P	A	P	P	A	P
	data	P/A	P/A	A	P/A	P	A	P
5	control	A	A	A	A	X	A	X
	data	A	A	A	A	X	A	X
6	control	P	P	X	P	P	X	P
	data	P	P	X	P	P	X	P

P: PSC only
A: AWG only
P/A:PSC or AWG
X: No coomunication

Figure 33. Transmission matrix based on node transceiver functional status

transmitter and one or more operating receiver can communicate with other nodes on the network).

We define a *universal mode* for maintaining communication to nodes with down PSC TT's and/or TR's. In the universal mode, both the AWG frames and the PSC frames are divided into a control phase and a data phase. A node with a data packet transmits a control packet during the control phase of the frame on either the PSC or the AWG based on its and the receiving node's transceiver status. For example, if a node wants to send a data packet to a node with a malfunctioning PSC TR, it transmits a control packet on the AWG during its turn in the AWG control packet transmission cycle. If the control packet is successfully transmitted, then the scheduling algorithm assigns a wavelength on the AWG.

To enable transceiver back–up, every node must know all other nodes' transceiver function status. To accomplish this, the MAC protocol executes the following: If a malfunction occurs on a node's AWG TR and/or AWG TT, the node signals to all of the nodes in

the network its status using its PSC TT during the control phase. Once this information is successfully transmitted to all of the nodes, the scheduling algorithm is updated such that future successfully transmitted control packets from and/or destined to the affected nodes are assigned wavelengths on the PSC.

If a malfunction occurs on a node's PSC TR and/or PSC TT, the signaling to the rest of the nodes becomes more complicated. There are several scenarios for signaling based on the component failure.

In the first scenario, a node with a malfunctioning PSC TR signals the network by transmitting a universal request packet using its functioning PSC TT during the control phase. The successfully transmitted packet is processed by all of the nodes on the network. Since the node with the malfunctioning PSC TR can not find out the result of its request packet, a pre-designated node will send an acknowledgment response to the AWG TR of the malfunctioning node on a pre-designated channel to inform the node that the network is in universal mode. If the malfunction node does not receive an acknowledgment on its AWG TR after a delay of a few round-trips then it knows that the request packet was unsuccessful and sends another one immediately.

In the second scenario, a node with a malfunctioning PSC TT signals the network by transmitting a request packet using its functioning AWG TT. First, the malfunctioning node listens to the PSC transmission and waits for an idle node and transmits the request packet to the idle node's AWG TR. After the idle node processes the request, it transmits a request packet during the PSC control phase on behalf of the malfunctioning node and identifies the malfunctioning node. After the request packet is successfully transmitted, the network switches to universal mode.

In the third scenario, a node with both PSC TT and PSC TR malfunctioning or

with a cut on the PSC fiber, broadcasts a request packet using its AWG LED. Since the malfunctioning node can not receive the control information that is exchanged over the PSC, it does not know about the ongoing transmissions on the AWG channels. Thus, the broadcast of the request packet may collide with ongoing data packet transmissions. In addition, the AWG TRs of the other nodes may be tuned to a different FSR and thus miss the broadcast request. In a typical network operating scenario, however, there is a reasonable chance that the broadcast request is successfully received by one (or more) of the other nodes. These other node(s) forward the request on the PSC channel used for control during the PSC control phase. A pre–designated node will then send an acknowledgment response on a pre–designated channel to the AWG TR of the malfunctioning node. If the malfunctioning node does not receive this acknowledgment response within a few round-trip times, it re–broadcasts its requested packet on its AWG LED.

APPENDIX B

THROUGHPUT–DELAY ANALYSIS FOR PSC‖PSC NETWORK AND AWG‖AWG NETWORK

We analyze the throughput–delay performance of the PSC‖PSC network and the AWG‖AWG network. We make the following traffic assumptions for these two homogeneous networks:

- A node selects one of the two devices with equal probability for transmission.
- Each node can have at most one data packet in the buffer to ensure a fair comparison with the AWG‖PSC network.

1. PSC‖PSC Network

For the PSC‖PSC network with control packet contention over one PSC, the control packet contention analysis is the same as in Section 3.2. Because there are three data slots per frame for each wavelength: one data slot per frame for the PSC with contention phase, two data slots per frame for the PSC dedicated to data, the throughput equation for the PSC‖PSC network is:

$$Z_{2PM} = \sum_{i=1}^{3\cdot\Lambda} i \binom{M}{i} \kappa^i (1-\kappa)^{M-i} + 3 \cdot \Lambda \cdot \sum_{j=3\Lambda+1}^{M} \binom{M}{j} \kappa^j (1-\kappa)^{M-j}, \tag{B.1}$$

The equilibrium condition for the PSC‖PSC network is $Z_{2PM} = \sigma \cdot E[\eta]$, which is used to solve numerically for the unknown η. The average delay (in frames) is $(N - E[\eta])/Z_{2PM}$.

2. AWG‖AWG Network

For the AWG‖AWG network, we consider two scenarios (i) with control contention over only one AWG, and (ii) control contention over both AWGs. In the case of control contention over one AWG, the contention analysis is the same as in Section 3.6. The

throughput is modified to reflect the additional 2 data slots per FSR per frame for the AWG dedicated to data transmission:

$$Z_{1M} = D \cdot \sum_{i=1}^{3\Lambda} i \binom{M}{i} \left(\frac{\kappa_A}{D}\right)^i \left(1 - \frac{\kappa_A}{D}\right)^{M-i} + 3 \cdot R \cdot D^2 \cdot \sum_{j=3\Lambda+1}^{M} \binom{M}{j} \left(\frac{\kappa_A}{D}\right)^j \left(1 - \frac{\kappa_A}{D}\right)^{M-j} \quad \text{(B.2)}$$

The equilibrium condition is $Z_{1M} = \sigma_A \cdot E[\eta]/D$, which is again used to solve numerically for η.

In the scenario of control contention over both AWGs, we assume that a node selects one of the two devices with equal probability for transmission. We define σ_{2A} as the probability that a given idle node generates a new packet by the beginning of its transmission cycle and sends this control packet to a given AWG. Clearly, $\sigma_{2A} = 1-(1-\sigma/2)^D$. Similarly, we define p_{2A} as the probability that a given backlogged node re–transmits a control packet over a given AWG at the beginning of a given cycle. Clearly, $p_{2A} = 1 - (1 - p/2)^D$. The probability that a given control slot on a given AWG contains a successfully transmitted control packet is

$$\kappa_{2A} = \frac{\eta}{D} \left(\frac{\sigma_{2A}}{M}\right) \left(1 - \frac{\sigma_{2A}}{M}\right)^{\frac{\eta}{D-1}} \left(1 - \frac{p_{2A}}{M}\right)^{\frac{N-\eta}{D}} + \frac{N-\eta}{D} \left(\frac{p_{2A}}{M}\right) \left(1 - \frac{p_{2A}}{M}\right)^{\frac{N-\eta}{D-1}} \left(1 - \frac{\sigma_{2A}}{M}\right)^{\frac{\eta}{D}} \quad \text{(B.3)}$$

This κ_{2A} is used to evaluate the average throughput over a given AWG, which — for a scheduling window of one cycle — is given by:

$$Z_{2M} = D \cdot \sum_{i=1}^{\Lambda} i \binom{M}{i} \left(\frac{\kappa_{2A}}{D}\right)^i \left(1 - \frac{\kappa_{2A}}{D}\right)^{M-i} + \cdot R \cdot D^2 \cdot \sum_{j=\Lambda+1}^{M} \binom{M}{j} \left(\frac{\kappa_{2A}}{D}\right)^j \left(1 - \frac{\kappa_{2A}}{D}\right)^{M-j} \quad \text{(B.4)}$$

The equilibrium condition is $Z_{2M} = \sigma_{2A} \cdot E[\eta]/D$, which is again used to solve numerically for η. The average throughput of the AWG‖AWG network (in packets per frame) is then given as $2 \cdot Z_{2M}$ and the average delay in the network (in frames) is $(N - E[\eta])/(2 \cdot Z_{2M}) + I_{del} + (Z_{2M} - D \cdot R)^+/(2 \cdot D \cdot R)$.

APPENDIX C

THROUGHPUT–DELAY ANALYSIS FOR THE AWG‖AWG NETWORK WITH D–BUFFER OPERATION

Here we analyze the throughput–delay performance of the AWG‖AWG network with D–buffer operation and full wavelength reuse (i.e., a scheduling window of one cycle). In the D–buffer operation, an idle buffer corresponding to a given frame (out of the D frames in the cycle) generates a new packet with probability σ at the beginning of that corresponding frame. In the frame assigned to the node for control packet transmission, control packets are sent for all packets that have been newly generated in the past D frames. In addition, control packets are sent for each backlogged (packet) buffer with probability p. Let η_D denote the total number of idle buffers in the network. Note that the are $D \cdot N - \eta_D$ backlogged buffers in the network. Also note that each frame is assigned N/D nodes for control packet transmission. Thus, in equilibrium, there are $\eta_D/D = \eta$ newly generated packets contenting in a given frame. In addition, there are $(D \cdot N - \eta_D)/D = N - \eta$ backlogged buffers contending in a given frame. Thus the probability of a control slot containing a successfully (without collision) transmitted control packet is κ given in (4.1). The throughput of the AWG‖AWG network in D–buffer operation with control packet contention on one AWG is thus obtained by replacing κ_A by κ in (B.2) and σ_A by σ in corresponding equilibrium condition.

The throughput of the AWG‖AWG network in D–buffer operation with control packet contention on two AWGs is obtained by replacing κ_{2A} by

$$\eta \left(\frac{\sigma}{2M}\right) \left(1 - \frac{\sigma}{2M}\right)^{\eta-1} \left(1 - \frac{p}{2M}\right)^{(N-\eta)} + (N-\eta) \left(\frac{p}{2M}\right) \left(1 - \frac{p}{2M}\right)^{(N-\eta)} \left(1 - \frac{\sigma}{2M}\right)^{\eta} \quad \text{(C.1)}$$

in (B.4) and σ_{2A} by σ in the corresponding equilibrium condition.

1. Analysis of Impact of Propagation Delay

Recall that the analysis in Section 3 assumed that the one–way end–to–end propagation delay in the network is less than one frame. In this appendix, we develop a more general analytical model which accommodates larger propagation delays. This more general model allows us to accurately characterize the performance of the AWG‖PSC network for the larger propagation delays in realistic networking scenarios.

For our analysis, we assume that all nodes are equidistant from the central AWG‖PSC. (This can be achieved in a straightforward manner by employing standard low–loss fiber delay lines.) Let τ denote the one–way end–to–end (from a given node to the central AWG‖PSC and on to an arbitrary node) propagation delay in integer multiples of frames (as defined in Section V). We furthermore assume that each node has a buffer that holds $\tau + 1$ packets.

In a typical scenario with a distance of 50 km from each node to the central AWG‖PSC and a propagation speed of $2 \cdot 10^8$ m/sec, the one–way end–to–end propagation delay is 0.5 msec. With an OC48 transmission rate of 2.4 Gbps and a frame size of 1,596 bytes (corresponding to a maximum size Ethernet frame) the propagation delay is $\tau = 94$ frames. (Buffering the corresponding 94 packets requires at most 150 kbytes of buffer in the electronic domain.) Note that if we had considered a frame size corresponding to the maximum size of a Sonet frame of 1,600 kbytes, the propagation delay would only be a fraction of one frame, which is accommodated by the analysis in Section 3.

We now proceed with the analysis for a propagation delay of multiple frames. The basic time unit in our analysis is the slot, i.e., the transmission time of a control packet, as defined in Section IV. Note that a propagation delay of τ frames is equivalent to a delay of

$\tau \cdot F$ slots. For our analysis, we introduce the concept of time–sequenced buffering.

2. Time–sequenced Buffering at Nodes

We view a given node's buffer capable of holding $\tau + 1$ packets as consisting of $\tau + 1$ *buffer slots*, as illustrated in Fig.34. Each buffer slot can hold one packet. In each frame, one of the buffer slots is the *active* buffer slot. The active buffer slot behaves exactly in the same way as the single-packet buffer considered in Section 3, i.e., if idle, it generates a new packet with probability σ and sends a control packet. If backlogged it sends a control packet with probability p. We stress that although our frame structure allows two sequential data transmissions per frame, there is only one data packet in a buffer slot corresponding to each frame.

The other τ buffer slots are *inactive*. The inactive buffer slots do not generate any new packets nor do they send any packets into the network. The purpose of the inactive buffer slots is to hold the data packets that correspond to the control packets that are propagating in the network.

A given buffer slot that is active in a given frame is inactive in the following τ frames (allowing each of the τ other buffer slots to be active for one frame), and then becomes again active $\tau + 1$ frames later.

Suppose a buffer slot is active in a given frame and in one of the M control slots in this frame sends out a control packet. This control packet arrives back at the node by the time the node becomes again active at the start of the $(\tau + 1)$th frame (i.e., after "sitting out" for τ frames). If the control packet is successful in control packet contention and data packet scheduling, the corresponding data packet is sent out in this $(\tau + 1)$th frame.

Also if the control packet is successful, a new data packet is generated with probability σ at the beginning of this $(\tau+1)$th frame. If a new data packet is generated, the corresponding control packet is sent in one of the M control slots of the $(\tau+1)$th frame. Note that we have tacitly assumed here that the nodal processing takes no more than $F-M$ slots. If the processing delay is larger, it can be accommodated in a straightforward manner by adding more buffer slots.

For an illustration of the concept of time–sequenced buffering, consider the buffer slots of a given node depicted in Fig. 34. Suppose buffer slot 1 is empty prior to time $t=0$, and generates a new packet, designated by D(1), at $t=0$. The control packet corresponding to D(1), designated by C(1), is sent in one of the M control slots of the frame that is sent between $t=0$ and $t=F$ (slots). By the time $t=F$, this frame is completely "on the fiber", as illustrated in the second snapshot in Fig. 34. (Note that this frame contains no data packets, as we assumed that buffer slot 1 was empty before $t=0$.) At $t=F$, buffer slot 1 becomes inactive, while buffer slot 2 becomes active. Suppose the node generates a new data packet D(2) at $t=F$. At $t=2F$ the frame with the control packet C(2) is completely on the fiber and buffer slot 3 becomes active, and so on.

At time $t=\tau F$ the frame containing C(1) starts to arrive back at the node. By time $t=\tau F+M$, the control packet is completely received and its processing commences. With an assumed processing delay of less than $F-M$ slots, the processing is completed by $t=(\tau+1)F$, which is exactly when buffer slot 1 becomes again active. Suppose C(1) was successful and the corresponding D(1) is scheduled on the AWG. Also suppose a new data packet $D(\tau+2)$ is generated at $t=(\tau+1)F$. By $t=(\tau+2)F$, the frame containing $D(1)$ and $C(\tau+2)$ is completely on the fiber, and buffer slot 2 becomes active, and so on.

3. Network Analysis

The key insight to the analysis of the network with time–sequenced buffering at the nodes is that in steady state it suffices to consider only the active buffer slot at each of the N network nodes. Specifically, at each instance in time, each node has exactly one active buffer slot. This active buffer slot is either idle or backlogged (similar to the way a node is either idle or backlogged in the analysis of Section 3). A buffer slot is considered idle if (i) it contains no data packet, or (ii) it successfully transmitted a control packet the last time it was active and the corresponding data packet has been successfully scheduled (although this data packet may still be in the buffer slot.)

An active buffer slot is considered backlogged if it contains a data packet whose corresponding control packet failed in the control packet contention or data packet scheduling. Let η denote the number of idle nodes (active buffer slots). Clearly, the number of backlogged nodes (active buffer slots) is $N - \eta$.

Now note that the control packet contention with time–sequenced buffer in a given frame is analogous to the control packet contention with the single–packet buffer considered in Section 3. In a given frame, each of the η idle active buffer slots generates a new data packet and sends a control packet with probability σ. Each of the $N - \eta$ backlogged active buffer slots retransmits a control packet with probability p. Thus the expected number of successful control packets per frame $M \cdot \kappa$, as given in Section 3.2.

Next note that the time–sequenced buffering does not interfere with the data packet scheduling as described in Section 2 and analyzed in Section 3. Thus, the throughput results derived for the different operating nodes in Section 3 apply without any modification to the time–sequenced buffer scenario.

Finally, note that the delays for the different operating modes as derived in Section 3 are scaled by the propagation delay of τ frames when considering the time–sequenced buffer scenario. Specifically, for the AWG–PSC mode, there is a delay component of τ frames for the initial control packet. In addition, there is a delay component due to control packet retransmissions (if control packet contention or data packet scheduling failed.) This second delay component is the expected number of backlogged nodes $N - E[\eta]$ divided by the expected throughput $Z_A + Z_P$ (similar to the case analyzed in Section 3.4), but is now scaled by the propagation delay τ. Thus, the average delay is

$$Delay = \tau \cdot \left(1 + \frac{N - \eta}{Z_P + Z_A}\right)$$

in frames, where we make again the reasonable approximation $E[\eta] \approx \eta$.

In analogous fashion, the average delay for the PSC–only mode is

$$Delay = \tau \cdot \left(1 + \frac{N - \eta}{Z_P}\right)$$

frames.

As discussed in Section 3.6, in the AWG–only mode with wavelength reuse, there are two additional delay components, cyclic control transmission delay I_{del} and scheduling delay if the data packet is not immediately transmitted. These two delay components are not affected by the propagation delay. Thus, the average delay for the AWG–only mode with spatial wavelength reuse is

$$Delay_{RE} = \tau \cdot \left(1 + \frac{N - \eta}{Z_{AM}}\right) + I_{del} + \frac{(Z_{RE} - D \cdot R)^+}{2 \cdot D \cdot R}$$

frames.

4. Numerical and Simulation Results

In this section, we examine the throughput–delay performance of the 2–device networks, AWG‖PSC, AWG‖AWG, and PSC‖PSC with time sequenced buffering. For the AWG‖AWG network we consider both single buffer and D–buffer operation. For the D–buffer operation we combine the time–sequenced buffering introduced in this appendix with the D packet buffers analyzed in Appendix II, for a total of $D \cdot (\tau + 1)$ packet buffers at each node of the AWG‖AWG network with D–buffer operation. (Each node has only $\tau + 1$ packet buffers in the other considered networks.) Throughput we consider the AWG‖AWG network with control packet contention on both AWGs and a scheduling window of D frames (the PSC‖PSC and AWG‖AWG networks have a scheduling window of one frame.) The numerical and simulation results are presented for one–way end–to–end propagation delays of $\tau = 4$ frames, $\tau = 16$ frames, and $\tau = 96$ frames in Fig. 35, Fig. 36, and Fig. 37, respectively. We observe that the throughputs for all of the networks are independent of the τ values and are the same. The throughput for the three networks are also the same as the throughput for a propagation delay of less than one frame, see Fig. 11. Thus, the time–sequenced buffering allows us to effectively utilize the full transmission capacity of the networks even for large propagation delays. Also it allows us to apply the probabilistic analytical model developed in Section. 3.

We observe that the AWG‖PSC network has smaller delay compared to the AWG‖AWG network for small τ. As the propagation delay τ increases the gap in delay between the AWG‖PSC network and the AWG‖AWG network becomes smaller. For small τ, the relatively larger delay for the AWG‖AWG network is due to the cyclic control packet transmission. As τ increases the delay due to the cyclic control packet transmission

becomes less and less dominant. We also observe that the single–buffer AWG‖PSC network gives larger throughput than the single–buffer AWG‖AWG network. The throughput of the D–buffer AWG‖AWG network is somewhat larger (at the expense of more complexity) than the throughput of the single–buffer AWG‖PSC network. Overall, the results indicate that the low-complexity AWG‖PSC network gives favorable throughput–delay performance for realistic propagation delays.

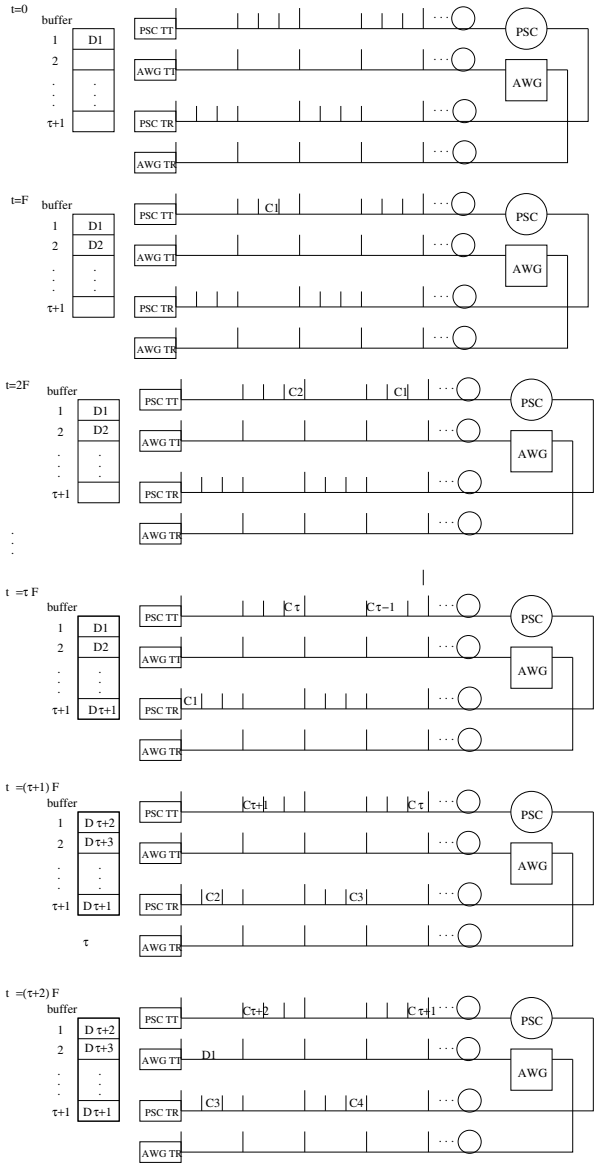

Figure 34. Illustration of time–sequenced buffering.

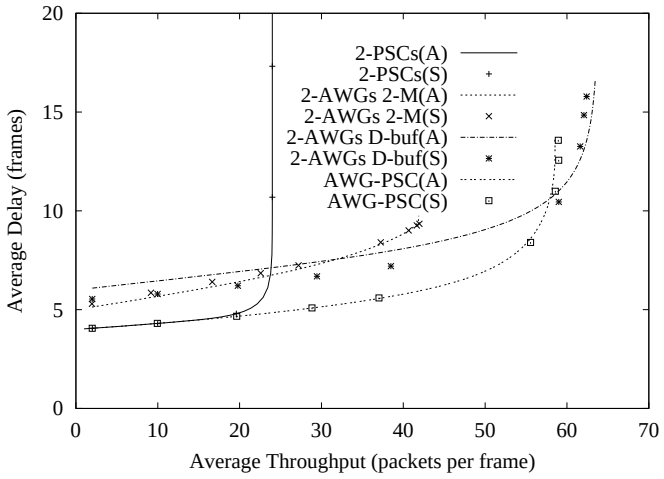

Figure 35. Throughput–delay performance comparison for two–device networks for a propagation delay of $\tau = 4$ frames ($N = 200$, fixed).

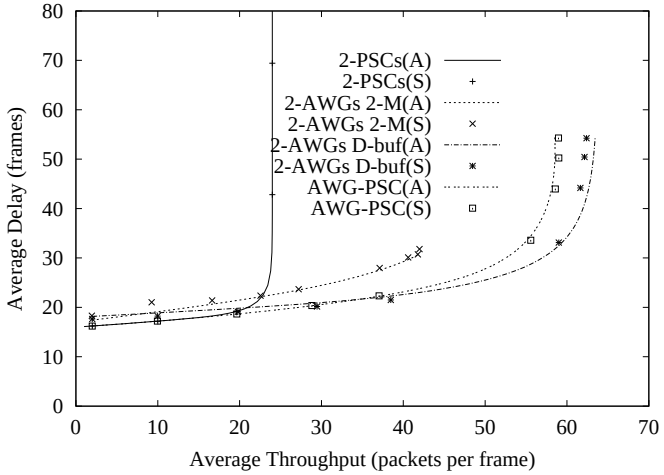

Figure 36. Throughput–delay performance comparison for two–device networks for a propagation delay of $\tau = 16$ frames ($N = 200$, fixed).

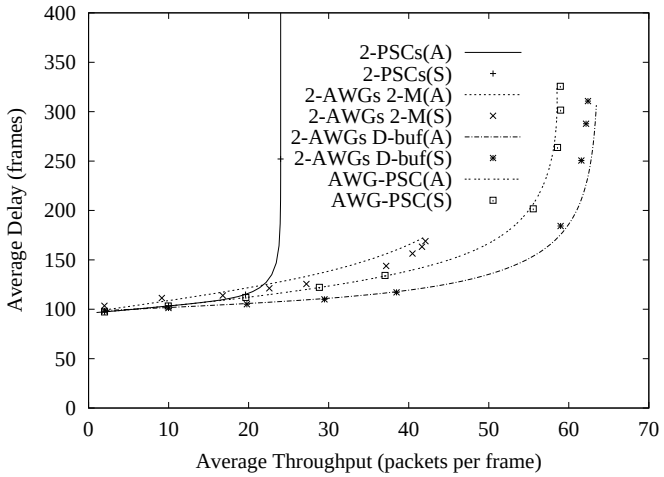

Figure 37. Throughput–delay performance comparison for two–device networks for a propagation delay of $\tau = 96$ frames ($N = 200$, fixed).

APPENDIX D

EVALUATION OF $P(\Delta = L|\gamma = N)$

In this appendix, we detail how to evaluate $P(\Delta = l|\gamma = n)$ given by (5.6). For $l = 2$, (5.6) takes the form

$$P(\Delta = 2|\gamma = n) = \binom{D}{2} \sum_{\substack{1\le k_1,k_2\le n\wedge S \\ k_1+k_2=n}} \frac{\binom{S}{k_1}\binom{S}{k_2}}{\binom{N}{n}} \tag{D.1}$$

$$= \binom{D}{2} \sum_{k_1=\max(1,n-S)}^{\min(n-1,S)} \frac{\binom{S}{k_1}\binom{S}{n-k_1}}{\binom{N}{n}}. \tag{D.2}$$

We define Q to represent the sum in (D.2), i.e.,

$$Q(\Delta = 2|\gamma = n) = \sum_{k_1=\max(1,n-S)}^{\min(n-1,S)} \binom{S}{k_1}\binom{S}{n-k_1}. \tag{D.3}$$

We note that when l increases by one in (5.6), we are adding one more term $\binom{S}{k_l}$. Thus

$$Q(\Delta = 3|\gamma = n) = \sum_{\substack{1\le k_1,k_2,k_3\le n\wedge S \\ k_1+k_2+k_3=n}} \binom{S}{k_1}\binom{S}{k_2}\binom{S}{k_3} \tag{D.4}$$

$$= \sum_{k_3=\max(1,n-2S)}^{\min(n-2,S)} Q(\Delta = 2|\gamma = n-k_3)\binom{S}{k_3}. \tag{D.5}$$

In general,

$$Q(\Delta = l|\gamma = n) = \sum_{k_l=\max(1,n-(l-1)S)}^{\min(n-l+1,S)} Q(\Delta = l-1|\gamma = n-k_l)\binom{S}{k_l}. \tag{D.6}$$

With the $Q(\Delta = l|\gamma = n)$, we can easily compute

$$P(\Delta = l|\gamma = n) = \binom{D}{l} \cdot \frac{Q(\Delta = l|\gamma = n)}{\binom{N}{n}}. \tag{D.7}$$

Next we compare the urn model with replacement for the multicasting developed in [14] with the urn model without replacement developed in this paper. The urn model with replacement is simpler as it does not keep track of the balls that have already been drawn. Instead, when a ball (node) is drawn, the color (AWG output port) of the ball is noted, and the ball is put back into the urn. Then the next ball is drawn, and so on. This urn model with replacement makes a modelling error in that it allows a given node to

Table 8. Probability distribution and expected value of number of AWG output ports with multicast destinations for $N = 20$ node network with $D = 4$ and $S = 5$ for multicast traffic ($u = 0.0$) with $\Gamma = 10$

	$P(\Delta = 1)$	$P(\Delta = 2)$	$P(\Delta = 3)$	$P(\Delta = 4)$	$E[\Delta]$
Urn with Repl.	0.037	0.222	0.371	0.371	3.075
Urn w/o Repl.	0.028	0.189	0.310	0.473	3.228
Simulation	0.027	0.190	0.309	0.473	3.228

be drawn multiple times as a destination of a given multicast. In contrast, the urn model without replacement allows each node to be counted only once as a destination of a given multicast. To illustrate these effects, consider a network with $D = 2$ AWG input ports and $D = 2$ output ports, $N = 2$ nodes and $S = 1$ nodes attached to each AWG output port for multicast traffic ($u = 0$) destined to two nodes ($\gamma = 2$). Clearly, in this scenario, each packet is destined to both AWG output ports, i.e., $P(\Delta = 2|\gamma = 2) = 1$, as correctly modelled by the urn model without replacement. With the urn model with replacement, on the other hand, we obtain $P(\Delta = 1|\gamma = 2) = 0.5$ and $P(\Delta = 2|\gamma = 2) = 0.5$. To see this, note that with probability 0.5 the ball selected in the second drawing is identical to the ball selected in the first drawing, with probability 0.5 the other ball is selected.

The modelling error of the urn model with replacement decreases as the probability of drawing the same ball multiple times decreases, which decreases as the number of balls (nodes in the network) increases. To illustrate the effect of the decreasing modelling error, we compare in Table 8 the probability distribution and expected value of the number of AWG output ports with attached destination nodes obtained from the urn model with replacement, the urn model without replacement, and simulations for a network with $N = 20$ nodes with $D = 4$ and $S = 5$ for multicast traffic $u = 0$ with a maximum of $\Gamma = 10$ destination nodes. We observe from the table that the urn model with replacement gives

Table 9. Probability distribution and expected value for number of AWG output ports with multicast destinations for $N = 200$ node network with $D = 8$ and $S = 25$ for mix of 80% unicast traffic ($u = 0.8$) and 20 % multicast traffic with $\Gamma = 200$

	$P(\Delta = 1)$	$P(\Delta = 2)$	$P(\Delta = 3)$	$P(\Delta = 4)$	$P(\Delta = 5)$	$P(\Delta = 6)$	$P(\Delta = 7)$	$P(\Delta = 8)$	$E[\Delta]$
Replacement	0.800	0.001	0.002	0.002	0.003	0.004	0.008	0.180	2.351
Refined	0.800	0.001	0.002	0.002	0.003	0.004	0.007	0.181	2.353
Simulation	0.800	0.001	0.001	0.002	0.003	0.004	0.007	0.183	2.364

too large values for the probabilities that the destinations are attached to a small number of AWG output ports and too small values for the probability that the multicast destinations are attached to a large number of AWG output ports. The urn model with replacement gives thus overall too small values for the expected number of AWG output ports with multicast destinations $E[\Delta]$. For the considered $N = 20$ node network, the urn model with replacement underestimates $E[\Delta]$ by almost 5%, which results in a correspondingly large underestimation of the transmitter throughput Z_T (5.13), the probability of generating a packet copy for a virtual queue σ_q (5.18), and the delays. We also observe from the table that the results obtained with the urn model with replacement closely match the simulation results.

In Table 9, we consider a $N = 200$ node network with $D = 8$ and $S = 25$ for a mix of 80% unicast traffic ($u = 0.8$) and 20% multicast traffic with $\Gamma = 200$. We observe from the table that the results from both urn models match the simulation results very closely. This is due to (i) the large fraction of unicast traffic for which the modelling error of selecting the same ball multiple times does not arise, and (ii) the large number of network nodes, which results in a small probability of selecting the same ball multiple times in the urn model with replacement.