



ELSEVIER

Available at  
www.ComputerScienceWeb.com  
POWERED BY SCIENCE @ DIRECT®

Computer Networks 41 (2003) 407–433

COMPUTER  
NETWORKS

www.elsevier.com/locate/comnet

# A hybrid MAC protocol for a metro WDM network using multiple free spectral ranges of an arrayed-waveguide grating

Martin Maier<sup>a</sup>, Martin Reisslein<sup>b,\*</sup>, Adam Wolisz<sup>a</sup>

<sup>a</sup> Telecommunication Networks Group, Technical University of Berlin, 10587 Berlin, Germany

<sup>b</sup> Telecommunications Research Center, Department of Electrical Engineering, Goldwater Center, MC7206, Arizona State University, P.O. Box 877206, Tempe, AZ 85287-7206, USA

Received 28 August 2001; received in revised form 19 September 2002; accepted 21 October 2002

Responsible Editor: M. Hamdi

## Abstract

In this paper, we report on a novel quality-of-service supporting reservation-based medium access control (MAC) protocol for a reliable, scalable, and cost-effective switchless wavelength division multiplexing network. The network is completely passive and is based on an arrayed-waveguide grating (AWG). Each node at the network periphery is equipped with one single fast tunable transceiver for data and one low-cost broadband light source which is spectrally sliced for broadcasting control information. Direct sequence spread spectrum techniques are used to enable simultaneous transmission of data and control. All wavelengths are used for data transmission and signaling is done in-band. Each node has global knowledge and schedules variable-size data packets on a deterministic first-come-first-served and first-fit basis guaranteeing fairness and completely avoiding channel and receiver collisions. The proposed protocol provides both packet and circuit switching and supports multicasting. The network efficiency is significantly increased by spatially reusing wavelengths and exploiting multiple free spectral ranges (FSRs) of the AWG. The analysis accounts for propagation delay and protocol processing time. Results show that using three FSRs instead of one significantly decreases the mean delay and improves the mean throughput by up to 88%, resulting in a normalized mean throughput of approximately 78%. The analytical results are verified by simulation.

© 2002 Elsevier Science B.V. All rights reserved.

**Keywords:** Arrayed-waveguide grating; Medium access control; Multiple free spectral ranges; Packet/circuit switching; Reservation; Scalability; Scheduling; Spectrum slicing; Spread spectrum; Wavelength division multiplexing

## 1. Introduction

Traffic in future optical networks will be predominantly IP based. The ubiquitous IP/ATM/SONET/WDM layer structure is likely to be replaced with a significantly less complex protocol stack IP/WDM for the following reasons. IP traffic

\* Corresponding author. Tel.: +1-480-965-8593; fax: +1-480-965-8325.

E-mail addresses: [maier@ee.tu-berlin.de](mailto:maier@ee.tu-berlin.de) (M. Maier), [reisslein@asu.edu](mailto:reisslein@asu.edu) (M. Reisslein), [wolisz@ee.tu-berlin.de](mailto:wolisz@ee.tu-berlin.de) (A. Wolisz).

URL: <http://www.eas.asu.edu/~mre>.

has some distinctive properties such as self-similarity, asymmetry, and server-based congestions (hot spots) [1]. SONET/SDH can carry this type of traffic only very inefficiently since it is designed for synchronous and symmetric traffic. ATM also suffers from inefficiencies which stem from a large cell tax. Beside large bandwidth and low attenuation one of the main drivers for optics is significant cost reduction [2]. By deploying optics, complex and quite costly ATM switches and SONET/SDH multiplexers can be bypassed. Another issue is the fact that both ATM and SONET/SDH provide resilience supporting features [3]. This redundant functionality leads to inefficiencies and requires rather complex layer interworking schemes.

In IP-over-WDM networks single or aggregated IP datagrams are directly transmitted over fibers using wavelength division multiplexing (WDM). To alleviate the electro-optic bottleneck and provide transparency, several photonic switching techniques have been proposed such as photonic slot routing, optical packet switching (OPS), optical label switching, and (labeled) optical burst switching ((L)OBS) [4–6]. Those approaches reduce the requirements on electronic processing devices by keeping the payload in the optical domain. Note that no optical random access memory is feasible at present. Instead, simple fiber delay lines with a fixed delay are used. While those delay lines allow for only suboptimal OPS, OBS does not need any buffers at all, resulting in bufferless networks. Network costs and complexity can be further reduced by using passive *wavelength-sensitive* devices such as an arrayed-waveguide grating (AWG) [7–10]. A heuristic algorithm for lightpath establishment in a *multihop* mesh network consisting of AWG-based network nodes was presented in [11]. A network consisting of multiple stages of AWGs is studied in [12] for connecting subscribers to a central office. AWGs have routing characteristics that depend on the wavelength. In addition, all wavelengths can be used at every AWG input port simultaneously as opposed to the passive star coupler (PSC), resulting in a significantly improved throughput-delay performance of the network [13]. Transmitters can reach different destinations by simply changing the wavelength. Thus, the switching functionality is naturally

moved toward the network periphery. With tunable transceivers we are able to realize *single-hop* WDM networks [14]. Single-hop networks have some very desirable properties such as minimum mean hop distance (unity), high channel utilization since no capacity is lost due to data forwarding, inherent transparency, and low nodal protocol processing requirements because nodes have to process only data packets which are addressed to them. They are well suited for realizing access or metro networks which will become increasingly important in the near future due to widely deployed bandwidth-consuming technologies such as digital subscriber line and cable modem [15]. Access networks are an important WDM market segment [16]. Recently, AWG-based metro WDM networks have attracted much attention in the literature [17–19].

In this paper, we develop a novel MAC protocol for an innovative AWG-based metro WDM network. The proposed network significantly differs from the networks presented in the existing literature: (1) Resources (time slots, wavelength channels) are allocated in a *distributed* fashion as opposed to the centralized resource management (used for instance in [8,19]). (2) Due to the novel node architecture, data and control can be transmitted *simultaneously* requiring neither an additional transceiver at each node nor one or more separate control channels. (3) The proposed MAC protocol makes use not only of WDM, TDM, and SDM, but also of *spreading techniques* and CDMA resulting in an improved network performance. (4) *Multiple free spectral ranges* (FSRs) are used to increase the degree of concurrency leading to a significantly increased network efficiency. We note that networks exploiting multiple FSRs of an AWG have received only little attention so far [20]. To our knowledge this is the first paper to develop and analyze a random MAC protocol for a network exploiting multiple FSRs of an AWG.

The remainder of the paper is organized as follows. In Section 2, we briefly outline the underlying principles of the network and MAC protocol. In Section 3, we discuss the network and node architecture. Section 4 explains the MAC protocol which is subsequently modeled in Section 5. The network throughput-delay performance is ana-

lyzed in Section 6. Numerical results are presented in Section 7. Section 8 concludes the paper.

## 2. Underlying principles

The AWG is a passive and polarization independent wavelength-routing device [21]. Due to its unique characteristics the AWG can be used to realize a wide range of WDM components such as add-drop multiplexer [22], discretely tunable filter and equalizer [23,24], optical crossconnect (OXC) [25], high-speed packet switch [26], multifrequency laser [27], packet synchronizer [28], broadband dispersion and dispersion slope compensator [29,30], or simple multiplexer/demultiplexer. We first discuss the following salient functions of an AWG: (1) Periodic wavelength routing, (2) spectral slicing of a broadband signal, and (3) spatial wavelength reuse. Without loss of generality we consider a  $2 \times 2$  AWG.

### 2.1. Periodic wavelength routing

In Fig. 1 we show a scenario where six wavelengths are launched into the upper AWG input port. The AWG routes every second wavelength to the same output port. This period of the wavelength response is called FSR. In our example, there are three FSRs, each containing two wavelengths. Generally, the FSR of a  $D \times D$  AWG,  $D \in \mathbb{N}$ , consists of  $D$  wavelengths, i.e., the physical degree of an AWG is identical to the number of wavelengths per FSR.

There is one important point to keep in mind. Each FSR provides one wavelength for communication between a given AWG input port and an arbitrary AWG output port. Hence, using  $R$  FSRs,  $R \in \mathbb{N}$ , allows for  $R$  simultaneous transmissions between each AWG input/output port pair.

### 2.2. Spectrum slicing

Spectral slicing is a means to realize broadcasting in WDM networks that are based on wavelength-sensitive devices [31]. Fig. 2 depicts the same scenario as shown in Fig. 1; in addition, a broadband signal (e.g., a light emitting diode (LED) signal) is fed into the upper AWG input port. In our example, the broadband signal spans all six wavelengths (channels). The AWG slices the broadband spectrum such that in each FSR one slice is routed to either AWG output port. Using  $R$  FSRs, there are  $R$  slices at each AWG output port. All those slices carry the same information. Hence, receivers attached to the AWG output ports are free to choose one of the  $R$  slices in order to retrieve the information. Note that while being tuned to any of those slices the receiver can also monitor wavelengths that originate from the same AWG input port as the broadband signal. However, wavelengths and broadband signal can be used simultaneously at the same AWG input port only if they can be distinguished at the receiver. This problem will be addressed in the next section. While the wavelengths serve as data channels a spectrally sliced broadband light source is well suited for broadcasting control information.

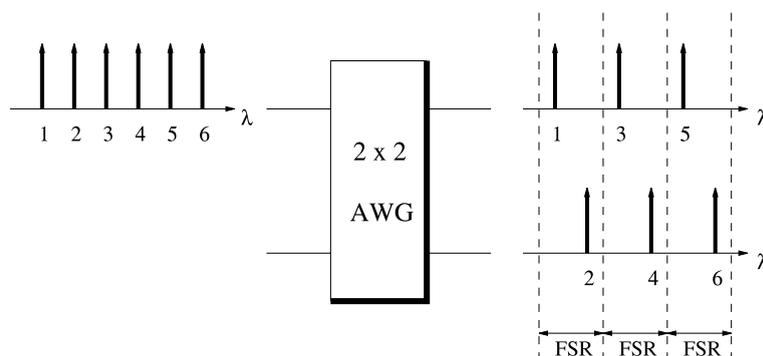


Fig. 1. Periodic wavelength routing of an AWG.

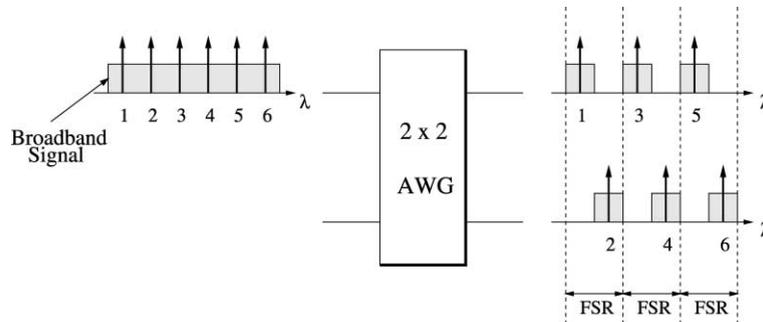


Fig. 2. Spectral slicing of a broadband signal.

There are three important points to keep in mind. First, spectrum slicing is an elegant way to realize broadcasting. Secondly, at every AWG output port each FSR contains one slice of the original broadband signal carrying the same information. Finally, while listening to a slice, the receiver is able to simultaneously obtain data on wavelengths that originate from the same AWG input port as the broadband signal.

### 2.3. Spatial wavelength reuse

Fig. 3 illustrates that all six wavelengths and an additional broadband signal can be applied at both AWG input ports simultaneously without resulting in collisions at the AWG output ports. Note that spectrally overlapping wavelengths and slices always emanate from the same AWG input port.

As a consequence, while listening to a slice a node can receive data only on wavelengths that

originate from the same AWG input port as the broadband signal. There are two important points to keep in mind. First, each wavelength and broadband signal can be applied on all AWG input ports simultaneously. The AWG routes wavelengths such that no channel collisions occur at the AWG output ports. Thus, with a  $D \times D$  AWG each wavelength can be spatially reused  $D$  times. Secondly, listening to a slice restricts the receiver to wavelengths that originate from the same AWG input port as the slice.

## 3. Architecture

### 3.1. Network and node architecture

Fig. 4 schematically shows the proposed AWG based single-hop network. There are  $N$  nodes, each attached to the network via two fibers. Every node

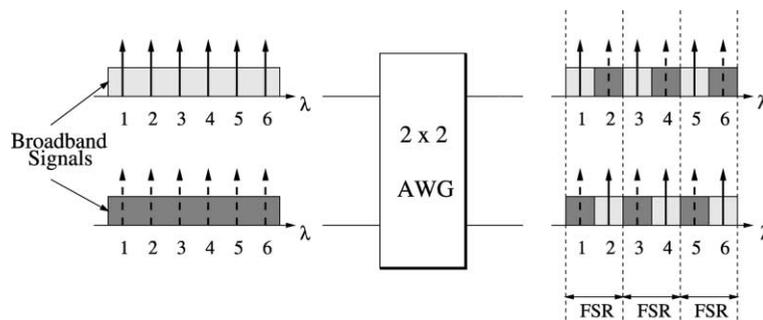


Fig. 3. Spatial wavelength reuse.

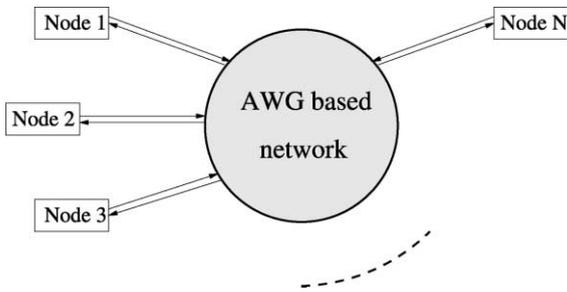


Fig. 4. Network architecture.

uses one fiber for transmission and the other fiber for reception.

The network and node architecture is depicted in more detail in Fig. 5. The network is based on a  $D \times D$  AWG. At each AWG input port a wavelength-insensitive  $S \times 1$  combiner is attached. Similarly, at each AWG output port signals are distributed by a wavelength-insensitive  $1 \times S$  splitter. Since at each AWG port all wavelengths can be simultaneously used for transmission, combiners and splitters are required to provide additional ports for attaching nodes. Those nodes can transmit on different wavelengths simultaneously, increasing the degree of concurrency. Each combiner collects the signals of  $S$  nodes while

each splitter distributes the signals to  $S$  receivers. (Note that the routing characteristics of an AWG require that both combiner and splitter be wavelength insensitive in order to guarantee full connectivity.) Each node is composed of a transmitting and receiving part. The transmitting part of a node is attached to one of the combiner ports. The receiving part of the same node is attached to the opposite splitter port. Thus, opposite combiners and splitters have the same physical degree, i.e., number of ports. However, the combiners (splitters) do not necessarily have to have the same degree  $S$ . Since the network is intended to be scalable, additional nodes can be attached to some combiners (splitters). As a consequence, combiners (splitters) can have different physical degrees. In this paper we consider the case in which all splitters and combiners have the same degree  $S$ . This ensures that the splitting loss is the same for all nodes.

The network connects  $N$  nodes, with  $N = D \cdot S$ . For a given number of nodes  $N$  there are several possible architectures with different values of  $D$  and  $S$ . For instance, eight nodes can be connected via a  $2 \times 2$  AWG with two  $4 \times 1$  combiners and two  $1 \times 4$  splitters or via a  $4 \times 4$  AWG with four  $2 \times 1$  combiners and four  $1 \times 2$  splitters. There are also cases, such as  $N = 7$ , where one or more ports

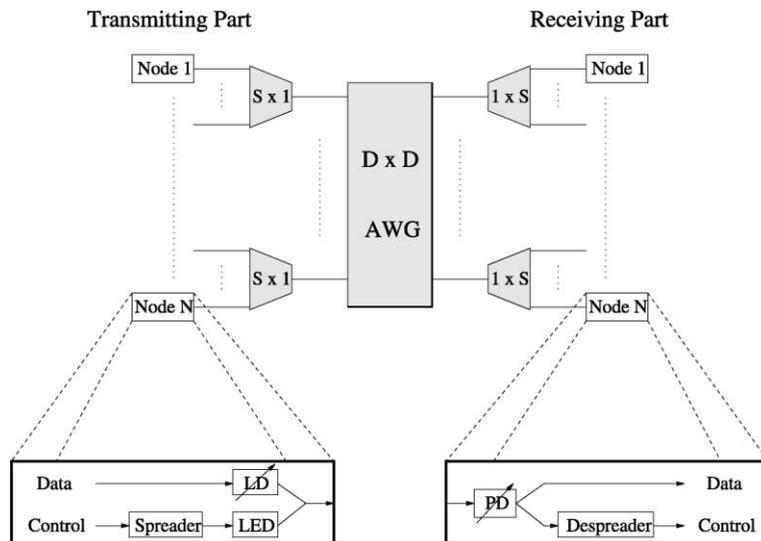


Fig. 5. Detailed network and node architecture.

are left unused. The choice of  $D$  and  $S$  trades off spatial wavelength reuse and receiver throughput. As we have seen in the previous section an AWG enables spatial wavelength reuse. Spatial wavelength reuse increases the degree of concurrency resulting in an improved throughput-delay performance. Therefore, from the spectrum reuse point of view it is reasonable to choose a large  $D$  for a given  $N$ . On the other hand, small values of  $D$  imply that many receivers are attached to the same splitter, i.e.,  $S$  becomes large. This has the advantage that each packet can be received by more nodes leading to an increased receiver throughput. The receiver throughput is defined as the mean number of busy receivers at steady state. An increased receiver throughput allows for efficient multicasting since multicast packets have to be transmitted fewer times. (We note, however, that a detailed investigation of multicasting is beyond the scope of this paper.)

Let us now take a look at the node structure. Each node contains a laser diode (LD) and a photodiode (PD) for data transmission and reception, respectively. Owing to the wavelength routing characteristics of the AWG, *both* transmitter and receiver have to be tunable over at least  $D$  wavelengths in order to provide full connectivity. In passive star coupler based single-hop WDM networks, on the other hand, it might be sufficient to have either a tunable transmitter or a tunable receiver [14]. (Though the node structure with a single transceiver is rather simple, costs could be reduced by replacing the tunable receiver with an array of fixed tuned receivers. However, it was shown in [32] that as the load increases the average number of wavelengths in use increases too. Thus, employing a tunable receiver instead of multiple fixed receivers is better because the wavelengths those fixed receivers are tuned to may all be in use (by other nodes) and a given node could not receive any packet even though not all of its receivers were busy. This situation is likely to arise when the number of channels is smaller than the number of nodes.) In addition, each node uses a broadband light source, say, a LED, for broadcasting control packets. The broadband LED signal (10–100 nm) is spectrally sliced such that all receivers are able to obtain the control information. No additional re-

ceiver is required if the signaling is done in-band, i.e., LED and LD signals overlap spectrally. However, data and control information have to be distinguishable at the receiver. This can be achieved by means of direct sequence spread spectrum techniques. The control information is spreaded before externally modulating the LED [33]. Accordingly, at the receiving part the control information is retrieved by despreding a part of the incoming signal. By using multiple spreading codes, several nodes are able to transmit multiple control packets at the same time, leading to code division multiple access (CDMA). CDMA not only allows simultaneous transmission of control packets but also provides some form of security. Only receivers which have the correct code are able to retrieve the information of a given control packet [34].

### 3.2. Physical limitations

In this section, we discuss the relevant physical factors that pose constraints on implementations of the above-mentioned network. Let us start with the side effects of CDMA. CDMA ensures a high level of concurrency, i.e., several simultaneous transmissions are possible by using different codes. Since in the proposed architecture each node has to process the control signals of all nodes the protocol computational overhead can become a serious bottleneck that affects the network scalability. In order to accommodate a large number of nodes and make the entire network scalable it is important to keep the computational complexity at each node small [35]. Therefore, we make use of one single code, just to enable the simultaneous transmission of data and control signals. This code is applied by all nodes, as will be explained in the next section. In addition, deploying a single code reduces the crosstalk penalty.

In the previous section we have seen that the AWG allows spatial wavelength reuse. Ideally, each wavelength is routed to a different output port without channel collision and crosstalk. However, real devices suffer from leakage [36]. As a consequence, each wavelength is routed not only to the intended AWG output port but is received in part at the other output ports as well. This in-

trachannel crosstalk has the same wavelength as the proper signal and cannot be removed by a demultiplexer at the destination, resulting in power penalties [37] and limited network scalability [38]. However, recent developments show that AWGs with uniform pass wavelength loss and negligible intrachannel crosstalk can be realized [39]. Note that due to spatial wavelength reuse the wavelength pool, i.e., the number of used wavelengths, is kept small. Thus, tunable transceivers with a limited tuning range can be deployed, e.g., electro-optic transceivers with a tuning range of about 15 nm. Those devices exhibit negligible tuning times in the range of a few nanoseconds. In the future, tunable transceivers not only with negligible tuning times but also with large tuning ranges should be feasible [40]. These transceivers will enable the use of multiple FSRs of an AWG.

Another crucial issue is the small bandwidth-distance product of LEDs. Especially for a large number of nodes the splitting loss due to the combiners and splitters in the proposed architecture puts severe constraints on the power budget. Those constraints can be relaxed by inserting erbium-doped fiber amplifiers (EDFAs) between each combiner/splitter and the corresponding AWG port. Since the physical degree of the AWG is rather small, only a few EDFAs would be required. Alternatively, each LED signal could be preamplified [41], or the LEDs could be replaced by other broadband light sources such as fiber amplifiers [42], Fabry–Perot lasers driven into clipping [43], or semiconductor optical amplifiers whose amplified spontaneous emission (ASE) is crossgain modulated by an additional laser diode [44]. The latter method helps increase the modulation speed up to 10 Gb/s and avoids the insertion loss of external modulators. Furthermore, incoherent sources such as LEDs and ASE sources suffer from spontaneous emission beat noise that places a significant limit on the spectral efficiency of those transmitters. This problem can be avoided by using highly coherent broadband sources such as modelocked lasers and supercontinuum (SC) generators [45]. However, all those solutions are either not very economic or support only small transmission rates. In our architecture the most promising approach for

transmitting spreaded control information appears to be the use of superluminescent diodes (SLDs) which provide a significantly improved power budget [46].

#### 4. MAC protocol

In the proposed network a MAC protocol is required for the following three reasons:

- Normally, the network layer is responsible for packet switching. However, since we consider a single-hop network there are no intermediate nodes and alternative routes to choose from. Consequently, in our architecture the network layer is not present and packet switching has to be handled by the MAC sublayer [35].
- Remember that each node is equipped with a single tunable transmitter and a single tunable receiver for data transmission and reception, respectively. To avoid tuning latencies fast tunable transceivers with a limited tuning range are deployed. As a consequence, there are more nodes than channels. The shared access to the channels has to be controlled by a MAC protocol.
- Due to the routing characteristics of the AWG each transceiver has to be tuned over at least one FSR in order to provide full connectivity (see Section 2.1). Hence, all wavelengths are shared by all nodes again calling for a MAC protocol.

##### 4.1. Basic principles

In the literature many MAC protocols have been proposed for single-hop WDM networks that are based on a PSC. Although the PSC is used to realize broadcast-and-select WDM networks as opposed to the AWG, some learnt lessons are considered generally valid and are therefore applied in our MAC protocol and reviewed briefly.

For single-hop networks with tunable transmitters and tunable receivers, pretransmission coordination by sending a control packet prior to data transmission is advantageous [47]. In the so-called tell-and-go protocols the data packet is transmitted immediately after the corresponding

control packet without waiting for any acknowledgement (ACK) from the intended receiver(s). This approach works well only at low traffic loads, i.e., if the control packet experiences neither channel nor receiver collision and the destination node has enough free resources to receive the data packet. In [48] it was shown that the throughput-delay performance of the network is improved if data packets are transmitted only after successful control packets. Here successful means that the packet has not suffered from channel or receiver collisions. A source node can learn about the success of its control packet by waiting for an ACK sent by the destination node. This conventional approach not only requires additional bandwidth for the ACK but also implies a large delay which is at least one round-trip time between source and destination nodes. The round-trip time is equal to the amount of time a packet takes from a source to a destination and again back to the source. Higher efficiency is achieved if the source node receives the corresponding control packet as well. This is feasible in the considered architecture since control signals are broadcast. In doing so, each node is able to check whether control packets have experienced channel collisions or not. This makes explicit ACKs unnecessary and reduces the delay to half the round-trip time. Note that this approach works well as long as both sender and destination receive an intact copy of the control packet. Since our architecture is passive and all-optical in nature the transmission path between any source–destination pair is not prone to errors. Furthermore, as we will see in the next section, a few redundant bits can be added to each control packet to protect it against single bit errors. In [49] it was shown that a normalized throughput (ratio of throughput and capacity) of up to 100% can be achieved if each node has global knowledge about all other nodes' activities. For this purpose each control packet has to be received by all nodes. As a consequence, receiver collisions of control packets are completely avoided, bandwidth is saved, and, as mentioned above, each node is able to immediately check whether channel collisions of control packets have occurred or not. All successful control packets are put in a distributed queue. Each node executes the same arbitration (scheduling)

algorithm. Thus, all nodes come to the same scheduling conclusions, avoiding channel and receiver collisions of data packets and reducing the signaling overhead. Control packets are not required to have a field that indicates the wavelength of the corresponding data transmission. Instead, wavelengths can be assigned on a first-come-first-served basis reducing the signaling overhead [50]. Finally, in high-speed optical networks propagation delay independent protocols reduce the channel collision probability and in terms of throughput-delay performance, channel collision avoidance is superior to retransmission [51].

#### 4.2. Details

The wavelength assignment at a given AWG input port is schematically shown in Fig. 6. The  $y$ -axis denotes the wavelengths used for transmission and reception. As illustrated,  $R$  adjacent FSRs are exploited at each AWG input port. Each FSR consists of  $D$  contiguous channels, where  $D$  denotes the physical degree of the underlying AWG. Transceivers are tunable over the range of  $R \cdot D$  contiguous wavelengths. To avoid interferences at the receivers during simultaneous transmissions in different FSRs of the AWG, the FSR of the receivers has to differ from the FSR of the AWG. In our case, the FSR of the receivers is equal to  $R \cdot D$  wavelengths. The  $x$ -axis denotes the time. Time is divided into cycles which are repeated periodically. Nodes are assumed to be synchronized to the cycle boundaries (how network-wide synchronization can be realized is discussed in [35]). Each cycle is further subdivided into  $D$  frames.

The frame format of one wavelength is depicted in Fig. 7. A frame contains  $F \in \mathbb{N}$  slots with a slot length equal to the transmission time of a control packet (function and format of a control packet will be explained later). The transceiver tuning time is assumed to be negligible. This is due to the fact that in the considered architecture the physical degree of the AWG is chosen above a certain threshold. This guarantees spatial wavelength reuse that is high enough to significantly reduce the wavelength pool and thereby the required transceiver tuning range. Transceivers with a limited tuning range such as electro-optic transceivers

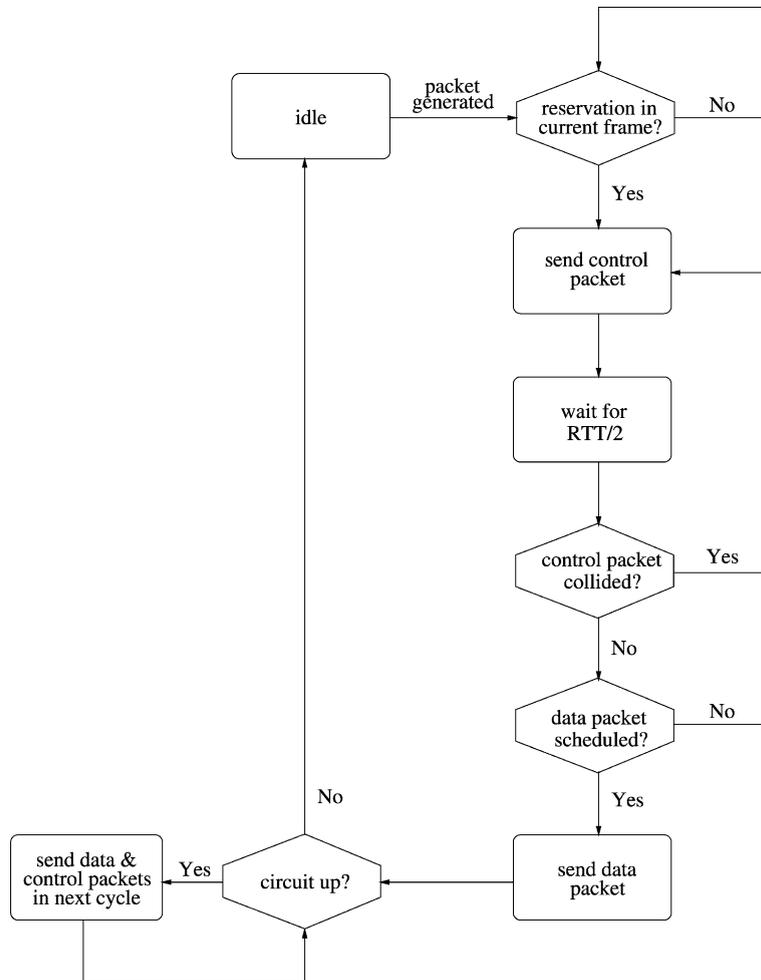


ALOHA is superior to carrier sensing multiple access because the ratio of propagation delay to packet transmission time (the famous  $a$ ) becomes too large [52]. Control packets arrive at the receivers after a propagation delay that is equal to half the round-trip time. The round-trip time is equal to the time a packet takes from a source to a destination and again back to the source. In the last  $(F - M)$  slots of each frame no control packets are sent, allowing receivers to be tuned to any arbitrary wavelength. This freedom enables transmissions between any pair of nodes. During those slots each node processes the received control packets by executing the same scheduling algorithm. The parameter  $M$  trades off two kinds of concurrency. During the first  $M$  slots of each frame, control and data packets can be transmitted simultaneously, but only from nodes which are attached to the same AWG input port. In this time interval, packets originating from other AWG input ports cannot be received. Whereas, during the last  $(F - M)$  slots of each frame all receivers are unlocked and can be tuned to any arbitrary wavelength. As a consequence, during this time interval, data packets from any AWG input port can be received. This allows for spatial wavelength reuse.

The MAC protocol works as follows. First, we consider the transmitting part of a node whose state transition diagram is depicted in Fig. 8. If a node has no data packet in its buffer, the LED and LD remain idle. When a data packet destined to node  $j$ ,  $1 \leq j \leq N$ , arrives at node  $i \neq j$ ,  $1 \leq i \leq N$ , node  $i$ 's LED broadcasts a control packet in one of the  $M$  slots of the frame allocated to the AWG input port that node  $i$  is attached to. The slot is chosen randomly according to a uniform distribution. A control packet consists of four fields, namely, destination address, length and type of the corresponding data packet, and forward error correction (FEC) code. Note that control packets do not have to carry the source address since each source node knows the propagation delay and the slot in which it has sent the corresponding control packet. Since all nodes are assumed to be synchronized to the frame boundaries each node knows from which AWG input port the corresponding control packets originate. As illustrated

in Fig. 7, the data packet can be of variable size  $L$ ,  $1 \leq L \leq F$ , where  $L$  denotes the length in units of slots. The type field contains one bit and is used to enable packet and circuit switching. The FEC is used by the receivers to detect and correct sporadic bit errors in the control packet, which—given the extremely small bit error rates of optical systems in general and the passive nature of our single-hop network—are rather unlikely. Thus, a moderately strong FEC makes the event of uncorrected sporadic bit errors extremely unlikely. Note that uncorrected sporadic bit errors could lead to a situation where a source node  $S$  could receive an intact control packet while destination node  $D$  could obtain a damaged copy of the same control packet which  $D$  would not consider further for scheduling. As a consequence, node  $S$  would send the corresponding data packet and node  $D$  would most likely not listen on the appropriate wavelength resulting in a receiver collision and wasted bandwidth. We also note that a large number of bit errors, which cannot be corrected by the moderately strong FEC is almost surely due to control packet collision and is interpreted as such by the source node (which re-transmits the control packet) and all the other nodes (which ignore the collided control packet). We finally remark that even though the transmission of packets can be made error free with the FEC, malfunction and/or failure of nodes may affect the operation of the distributed scheduling protocol. This issue has to be addressed by higher-layer protocols and is beyond the scope of this paper.

Let us now take a look at the receiving part of a node. Fig. 9 shows the corresponding state transition diagram. Every node collects all control packets by tuning its receiver to one of the corresponding channels during the first  $M$  slots of each frame. Thus, it learns about all other nodes' activities and whether its own control packet was successful or not by using the FEC field. In frame  $k$ ,  $1 \leq k \leq D$ , each receiver collects the control packets which have been sent half a round-trip time ago by nodes that are attached to AWG input port  $k$ . If its control packet has collided node  $i$  retransmits the control packet in the next cycle with probability  $p$  and with probability  $(1 - p)$  it will defer the transmission by one cycle. The node



RTT = Round Trip Time

Fig. 8. MAC protocol: State transition diagram of a node's transmitting part.

retransmits the control packet in this next cycle with probability  $p$ , and so forth. Successful control packets are put in a distributed queue at each node.

All nodes process the successfully received control packets by executing the same arbitration (scheduling) algorithm in the last  $(F - M)$  slots of each frame. Consequently, all nodes come to the same transmission and reception schedule. Note that  $M$  has to be smaller than  $F$  to provide enough

time for executing the algorithm. Since each node has to process the control packets of all nodes the computational complexity at each node puts severe constraints on the network scalability. A simple arbitration algorithm is required to relax those constraints [53]. Therefore, we apply a straightforward greedy algorithm which schedules the data packets on a first-come-first-served and first-fit basis. Note that this algorithm together with the fact that each node randomly selects one

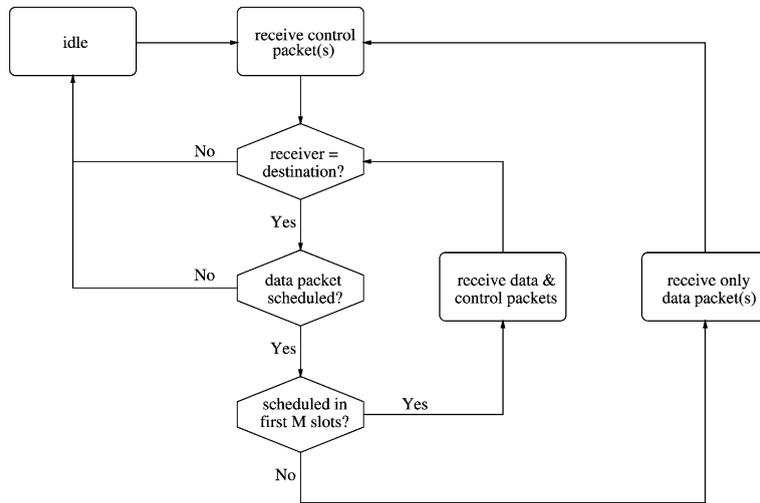


Fig. 9. MAC protocol: State transition diagram of a node's receiving part.

of the  $M$  reservation slots according to a uniform distribution guarantees fairness. After receiving a successful control packet the arbitration algorithm tries to schedule the transmission of the corresponding data packet within the following  $D$  frames. Those  $D$  frames do not necessarily have to coincide with the cycle boundaries. The data packet is sent in the first possible slot(s) using the lowest available wavelength. If there are not enough slots available within the  $D$  frames the data packet is not transmitted and the source node has to retransmit the control packet in the next cycle. Nodes that lose the arbitration are aware of this because all nodes execute the same scheduling algorithm. Note that global knowledge in conjunction with distributed scheduling reduces the delay by avoiding explicit ACKs and can achieve a normalized throughput of up to 100%.

The length of the scheduling window is equal to  $D$  frames (i.e., one cycle) for two reasons. First, by limiting the length of the scheduling window to a small number of frames the computational requirements at each node are kept low. Due to the relatively small scheduling window each node has to maintain and update only small schedule tables. Secondly, the scheduling window of one cycle allows for simultaneous transmission of data packets and control packets. To see this, note that the nodes attached to a given AWG input port send

their control packets every  $D$  frames. A scheduling window of  $D$  frames thus ensures that the nodes can also send data packets in the frame in which they send control packets. This concurrency leads to an improved throughput-delay performance.

Next, we discuss the support for multicasting and circuit switching. Multicasting is realized by the splitters. Each splitter distributes an incoming packet to all attached nodes. By tuning the receivers to the respective wavelength the packet can be obtained by more than one node. The resulting increased receiver throughput has a positive impact on the network performance. (We note that a detailed study of multicast is beyond the scope of this paper.) Circuit switching is realized by using the type and length fields of the control packet. The length field denotes the required number of slots per cycle. By setting the bit in the type field, the source node indicates that the number of slots given in the length field must be reserved in each cycle. After receiving the control packet the circuit is setup by choosing the first possible free slot(s) at the lowest available wavelength. The slot(s) is (are) reserved in the subsequent cycles until the connection is terminated. If there are not enough free resources the control packet is discarded and has to be retransmitted in the next cycle. The termination of a circuit works as follows. Suppose node  $i$ ,  $1 \leq i \leq N$ , has setup a circuit, i.e., node  $i$  is

granted a certain number of slots per cycle. Furthermore, suppose  $j$ ,  $1 \leq j \leq M - 1$ , other nodes attached to the same combiner currently hold circuits. Then, in each cycle node  $i$  repeats the control packet in slot  $j + 1$  of the corresponding reservation window. To terminate the circuit, node  $i$  simply stops repeating the control packet. In doing so, all other nodes notice that the circuit has terminated and the respective slot is freed up for contention. Note that during the holding time of a circuit other circuits can be torn down. As a consequence, the corresponding slot, say  $k$ ,  $1 \leq k \leq j + 1$ , becomes idle. Whenever this happens, all slots which are larger than  $k$  and are used to indicate the existence of circuits are decremented by one. Thus, the first  $j$  slots of the corresponding reservation window indicate the existence of circuits while the remaining  $(M - j)$  slots are free to be used for reservations. A node with a control packet to send chooses one of slots  $j + 1, j + 2, \dots, M$  at random. Consequently, while circuits are setup, not all  $M$  slots are available for reservation, resulting in an increased congestion of the slotted ALOHA channel. However, each node only has to monitor the control channel to find out the end of a circuit and does not have to maintain and update a lifetime variable for each single connection. Again, this reduces the computational burden on each node which is an important factor, especially in high-speed optical networks where channel access control is based on global knowledge.

Finally, we point out that the proposed reservation and circuit setup is able to provide QoS to real-time applications, such as voice, video, and audio. Circuits could provide QoS to individual flows (following the IntServ paradigm) or flow aggregates (following the DiffServ paradigm). Similarly, in the packet switching mode different scheduling algorithms could provide QoS to individual flows or flow aggregates.

#### 4.3. An illustrative example

In this section, we illustrate the following features of the proposed MAC protocol:

- dynamic channel allocation,
- packet switching,

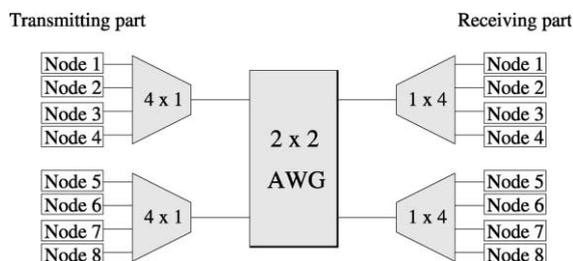


Fig. 10. Network architecture ( $N = 8$ ,  $D = 2$ ,  $S = 4$ ).

- circuit switching,
- channel collision of control packets and its resolution,
- simultaneous reception of control and data packets,
- variable-size data packets,
- using multiple FSRs.

As illustrated in Fig. 10, we consider  $N = 8$  nodes which are connected via a  $2 \times 2$  AWG ( $D = 2$ ) with attached  $4 \times 1$  combiners and  $1 \times 4$  splitters ( $S = 4$ ). Fig. 11 depicts the wavelength assignment. In this figure we assume that only nodes attached to the upper AWG input port have data packets to send. Two FSRs are used ( $R = 2$ ), each consisting of two wavelengths. Each frame consists of  $F = 5$  slots. The reservation window is  $M = 3$  slots long. The upper part shows the transmitters while the lower part indicates the reception of the transmitted packets after a propagation time  $\tau$ . For the sake of simplicity, we assume that the distance between each node and the AWG is equal, i.e., the propagation delay  $\tau$  is the same for all nodes. In our example  $\tau$  is equal to 1 frame.

Let us start considering the left-most slot of the transmitting part. In frame 1 of cycle 1 only nodes 1–4 are permitted to send control packets. In the first slot nodes 1 and 4 simultaneously transmit a control packet resulting in a channel collision. Node 2 randomly selects slot 3. After the propagation delay of one frame the control packets arrive at all nodes. Every node has tuned its receiver such that control packets can be obtained, i.e., nodes 1–4 are tuned to wavelength 1, while nodes 5–8 are tuned to wavelength 2. Note that nodes could also receive the control packets by tuning

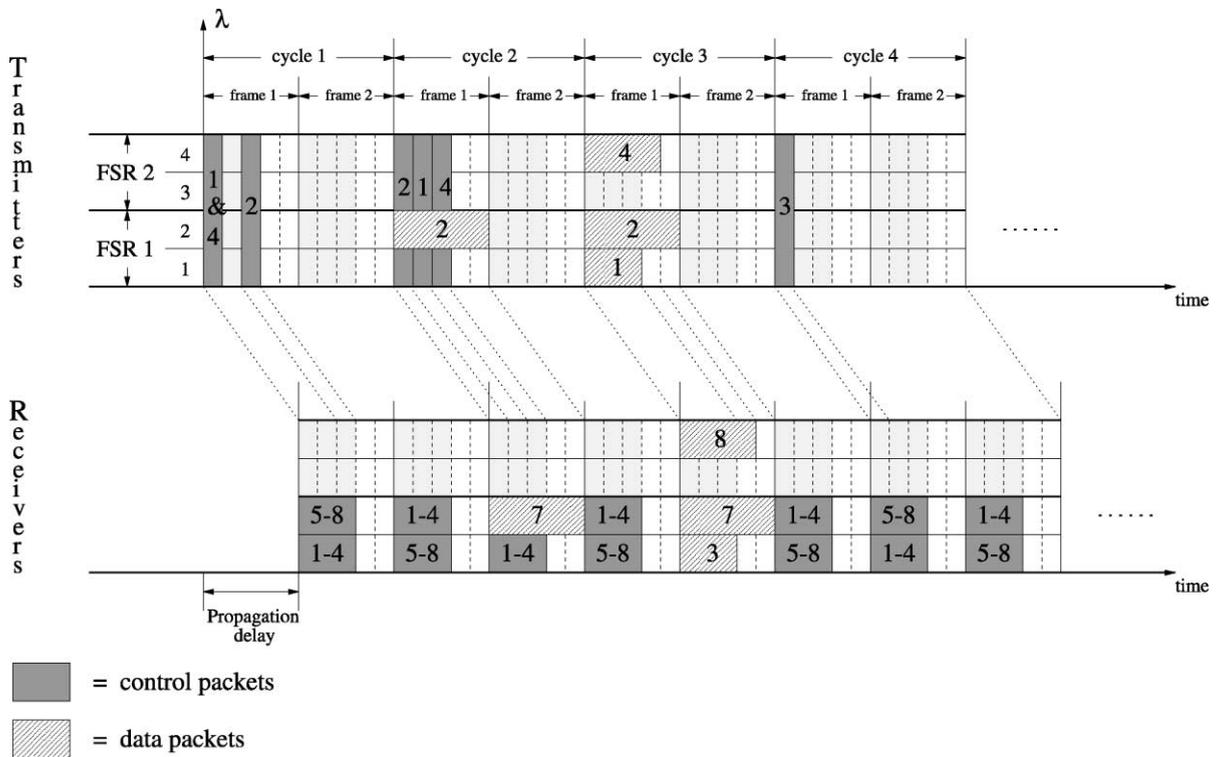


Fig. 11. Wavelength assignment ( $R = 2, F = 5, M = 3, \tau = 1$  frame).

their receivers to the corresponding channels of the other FSR, i.e., nodes 1–4 could tune their receivers to wavelength 3 and nodes 5–8 could tune their receivers to wavelength 4. The control packet of node 2 is received successfully by all nodes. In our example, that control packet is destined to node 7 and requests a circuit with five slots per cycle. The execution of the arbitration algorithm is assumed to take two slots. The data packet is transmitted on the first available channel at the earliest possible time. Accordingly, the data packet is sent on wavelength 2 during frame 1 of cycle 2.

Node 2 repeats the control packet in each cycle until the circuit is torn down. Since in our example there are no other circuits currently setup, node 2 sends the control packet in the first slot of frame 1. As illustrated in Fig. 11, slot 1 becomes idle in cycle 3. This tells all nodes that the circuit between nodes 2 and 7 is terminated and that this slot can be used again by all nodes. Node 3 captures that slot in cycle 4 to announce the transmission of a

single packet or to setup a circuit in the subsequent cycle.

Now, let us go back to the initially collided control packets of nodes 1 and 4. After  $\tau$  frames both nodes learn about the channel collision of their control packets. As a consequence, they retransmit their control packets in the next cycle with probability  $p$  keeping in mind that the first slot of frame 1 is fixed assigned to node 2 for indicating the corresponding circuit. In our example, nodes 1 and 4 successfully retransmit their control packets in frame 1 of cycle 2. First, we take a look at node 1. Node 1 has one single data packet which is three slots long and is destined to node 3. After waiting for three slots, node 1 sends the data packet on wavelength 1. Similarly, node 4 has a data packet for node 8 which is four slots long. This data packet is sent at the beginning of frame 1 of cycle 3. Note that node 4 chooses wavelength 4 since wavelength 2 is already used by node 2. This is an example for using multiple FSRs for com-

munication between a given AWG input–output port pair. In addition, while receiving the data packets nodes 3, 7, and 8 monitor the control channel. Node 8 thereby receives the corresponding slice of the second FSR. Recall that this is possible due to spreading and spectrally slicing the LED signal.

#### 4.4. Discussion

In the proposed network each node is able to acquire global knowledge without requiring an additional transceiver which is fixed tuned to a separate control channel as in [54]. Previously investigated architectures with a single tunable receiver at each node suffer from the problem that while receiving control packets no data packets can be received and vice versa resulting in a decreased throughput-delay performance [55,56]. As we have seen, this problem can be solved by means of direct sequence spread spectrum techniques. This approach can be extended to CDMA by applying multiple spreading codes simultaneously. By using a modified version of slotted ALOHA for reservation our network is scalable as opposed to networks with fixed assigned reservation slots [57]. Note that for the design of the MAC protocol reducing the computational overhead was more important than maximizing the performance. This was done by adopting a simple scheduling algorithm with a rather small scheduling window and sacrificing some reservation slots for indicating the existence of circuits which in turn relieves each node from maintaining and updating state tables. Nevertheless, in Section 7 we will see that a normalized mean throughput of up to 82% can be achieved for realistic parameter values. Further performance improvements are feasible at the cost of higher computational complexity at each node.

### 5. Model

In our analysis we focus on the case of packet switching of fixed-size data packets. For the analysis we make the following typical assumptions:

- A node with an empty buffer generates a data packet with probability  $\sigma$  at the end of a frame.
- Each node has a *single-packet buffer* (a typical assumption for the analysis of MAC protocols for WDM networks [58]). After transmitting a data packet in a given frame the buffer becomes empty at the end of that frame.
- A data packet has a *fixed size* of  $F$  slots, i.e.,  $L = F$ .
- *Uniform unicast* traffic: A data packet is destined to any one of the other  $(N - 1)$  nodes with equal probability.
- The propagation delay  $\tau$  is the same for all nodes and is an integer multiple of one frame, i.e., all nodes are equidistant from the AWG.
- *Nonpersistence*: Random selection of a destination node among the other  $(N - 1)$  nodes is renewed for each attempt of transmitting a control packet. (The nonpersistence assumption is needed to obtain a Markovian model [59].)
- *Delayed first-time transmission*: A node sends out its control packet in a frame with probability  $p$ , for both first-time transmission as well as retransmissions. (This assumption simplifies the calculation of the probability of control packet collisions [58].)

Fig. 12 depicts an approximate model for the MAC protocol. Each node can be in one of the  $(2\tau + 3)$  modes during any frame. Transitions from one mode to another mode occur only at the beginning of a frame. The modes are defined as follows:

- *TH*: Nodes in the *TH* (thinking) mode generate a data packet with probability  $\sigma$  at the end of a frame.
- *B*: Nodes in this mode are backlogged and send a control packet with probability  $p \cdot \beta$  (where  $\beta$  accounts for the cycles in the time structure, as is explained later) at the beginning of the next frame.
- $PQ_1, PQ_2, \dots, PQ_\tau$ : Those modes represent the propagation delay of successfully transmitted control packets. Nodes move from mode  $PQ_i$  to mode  $PQ_{i+1}$ ,  $i = 1, 2, \dots, \tau - 1$ , at the beginning of the next frame with probability 1.

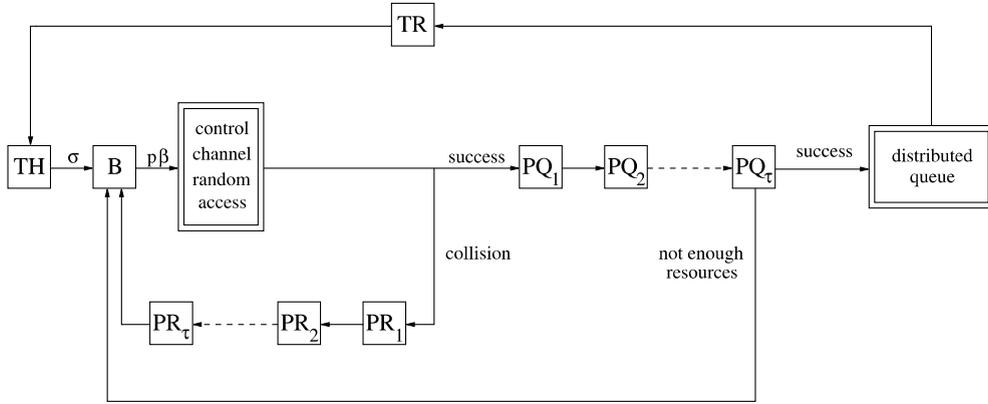


Fig. 12. Model for MAC protocol.

- $PR_1, PR_2, \dots, PR_\tau$ : Those modes are similar to the  $PQ_i$  modes,  $i = 1, 2, \dots, \tau$ . Nodes whose collided control packets have to be retransmitted enter the mode  $B$  after  $\tau$  frames.
- $TR$ : Nodes in the mode  $PQ_\tau$  whose data packets are successfully scheduled are put into a distributed queue. A node leaves the distributed queue and moves to mode  $TR$  (transmission) in that frame which it also uses for sending control packets. After transmitting the data packet the nodes return to the  $TH$  mode. Nodes in mode  $PQ_\tau$  whose data packets are not scheduled due to the lack of free resources (not enough free slots and/or wavelengths) move to mode  $B$ .

The system state in frame  $n$ ,  $n \in \mathbb{Z}$ , is completely described by the following state vector:

$$\mathbf{N}(n) = (N_B(n), N_{PR_1}(n), \dots, N_{PR_\tau}(n), \\ N_{PQ_1}(n), \dots, N_{PQ_\tau}(n), N_{TR}(n))$$

where  $N_X(n)$  denotes the number of nodes in mode  $X$  in frame  $n$ . Note that  $N_{TH}$  is not included in the state vector since it is linearly dependent on the other modes. With the nonpersistence assumption  $\{\mathbf{N}(0), \mathbf{N}(1), \dots, \mathbf{N}(n), \dots\}$

is a discrete-time multi-dimensional Markov chain with finite but quite large state space. The exact analysis of that Markov chain would involve the calculation of the state transition probability matrix which is computationally prohibitive. Therefore, we analyze the system at an equilibrium point

using the equilibrium point analysis (EPA) approach [58,60].

## 6. Analysis

In the EPA method the system is assumed to be always at an equilibrium point, defined as

$$\mathbf{N} = (N_B, N_{PR_1}, \dots, N_{PR_\tau}, N_{PQ_1}, \dots, N_{PQ_\tau}, N_{TR}).$$

At an equilibrium point the expected increase in the number of nodes in each mode per unit time (i.e., frame) is zero. Applying this condition to all the modes, we get a set of so-called equilibrium point equations.

### 6.1. Equilibrium point equations

By writing the equation for each mode, we get  $(2\tau + 3)$  equations. Let  $\delta_X(\mathbf{N})$  be the conditional expectation of the increase in the number of nodes in mode  $X$  in a frame, given that the system is in state  $\mathbf{N}$ . Since  $\delta_{PR_i}(\mathbf{N}) = N_{PR_{i-1}} - N_{PR_i} = 0$ ,  $i = 2, 3, \dots, \tau$ , we can omit the subscript of  $PR$  by letting  $N_{PR} = N_{PR_1} = \dots = N_{PR_\tau}$ .

Similarly, for the modes  $PQ_i$ ,  $i = 1, 2, \dots, \tau$ , we get

$$N_{PQ} = N_{PQ_1} = \dots = N_{PQ_\tau}. \quad (2)$$

For the modes  $TH$  and  $B$  we have the following equations:

$$\begin{aligned} \delta_{TH}(\mathbf{N}) &= N_{TR} - N_{TH}\sigma \\ &= N_{TR} - [N - N_B - \tau(N_{PR} + N_{PQ}) - N_{TR}]\sigma = 0, \end{aligned} \quad (3)$$

$$\delta_B(\mathbf{N}) = [N_{TH}\sigma + (N_{PQ} - N_{TR}) + N_{PR}] - N_B p\beta = 0. \quad (4)$$

To obtain the remaining equilibrium point equations for the modes  $PR_1$ ,  $PQ_1$  and  $TR$  we introduce the quantities  $Y(\mathbf{N})$  and  $Z(\mathbf{N})$ . Let  $Y(\mathbf{N})$  denote the conditional expectation of the number of nodes that move out from mode  $B$  to mode  $PQ_1$ , given that the system is in state  $\mathbf{N}$ .  $Y(\mathbf{N})$  is the average number of control packets transmitted in a frame without collision. With  $Y(\mathbf{N})$  we obtain the following equations for the modes  $PR_1$  and  $PQ_1$ :

$$\delta_{PR_1}(\mathbf{N}) = N_B p\beta - Y(\mathbf{N}) - N_{PR} = 0, \quad (5)$$

$$\delta_{PQ_1}(\mathbf{N}) = Y(\mathbf{N}) - N_{PQ} = 0. \quad (6)$$

Let  $Z(\mathbf{N})$  denote the conditional expectation of the number of nodes that move from mode  $PQ_\tau$  to mode  $TR$ , given that the system is in state  $\mathbf{N}$ .  $Z(\mathbf{N})$  is the average number of nodes that successfully transmit a data packet in a frame. With  $Z(\mathbf{N})$  we obtain the following equation for the mode  $TR$ :

$$\delta_{TR}(\mathbf{N}) = Z(\mathbf{N}) - N_{TR} = 0. \quad (7)$$

Next, we need to solve for the unknown quantities  $\beta$ ,  $Y(\mathbf{N})$  and  $Z(\mathbf{N})$ . Recall that a backlogged node can transmit a control packet only in one frame per cycle that consists of  $D$  frames. Let  $\beta$  denote the probability that the next frame is allocated to the backlogged node. Thus, we get

$$\beta = \frac{1}{D}. \quad (8)$$

The average number of successfully transmitted control packets per frame is given by [58]

$$Y(\mathbf{N}) = \sum_{i=1}^{N_B} i \left(1 - \frac{1}{M}\right)^{i-1} \binom{N_B}{i} (p\beta)^i (1 - p\beta)^{N_B-i}, \quad (9)$$

$$Y(\mathbf{N}) = N_B p\beta \left(1 - \frac{p\beta}{M}\right)^{N_B-1}. \quad (10)$$

The result can be interpreted such that  $p\beta(1 - p\beta/M)^{N_B-1}$  is the probability that a node's control packet is transmitted collisionfree. The average number of nodes that move from mode  $B$  to mode  $PQ_1$  is given by Eq. (10).

Let  $q$  be the probability that a given slot of the first  $M$  slots of a frame contains exactly one control packet that is to be scheduled. Then,

$$q = \frac{Y(\mathbf{N})}{M}. \quad (11)$$

The probability that exactly  $i$  control packets are to be scheduled in a frame is

$$P_i = \binom{M}{i} q^i (1 - q)^{M-i}, \quad i = 0, 1, 2, \dots, M. \quad (12)$$

Each of the  $i$  control packets originates from one of the  $N$  nodes with equal probability  $1/N$ . With  $i$  control packets, in each frame the average number of control packets that belong to nodes attached to the same combiner is equal to  $i \cdot \beta = i/D$ . Recall that due to their fixed size of  $F$  slots, data packets can be sent from those nodes only every  $D$  frames. Data packets cannot be transmitted in other frames since they are larger than  $(F - M)$  slots. Thus, in each frame only nodes attached to the same combiner can send data packets. Control packets emanating from nodes attached to the same combiner aggregate over the interval of  $D$  frames until data transmission takes place. As a consequence, in each frame the average number of control packets to be scheduled is given by  $i/D \cdot D = i$ .

The probability that at least one among those  $i$  control packets is destined to a given node under the assumption that a node does not transmit to itself is equal to [59]

$$\begin{aligned} p_o(i) &= 1 - \left[ \frac{i}{N} \left(1 - \frac{1}{N-1}\right)^{i-1} \right. \\ &\quad \left. + \left(1 - \frac{i}{N}\right) \left(1 - \frac{1}{N-1}\right)^i \right], \end{aligned} \quad (13)$$

$$p_o(i) = 1 - \left(1 - \frac{1}{N-1}\right)^{i-1} \frac{N^2 - 2N + i}{N(N-1)}. \quad (14)$$

Let  $g(i)$  denote the average number of nodes that successfully transmit a data packet in a frame, given that  $i$  control packets are to be scheduled. Given this, the number of data packets destined to nodes that are attached to the same splitter is binomially distributed  $BIN(S, p_o(i))$ . However, no more than  $R$  data packets can be simultaneously transmitted to those nodes. This holds for each of the  $D$  splitters and we finally obtain

$$g(i) = D \left\{ \sum_{k=0}^R k \binom{S}{k} p_o(i)^k [1 - p_o(i)]^{S-k} + R \sum_{k=R+1}^S \binom{S}{k} p_o(i)^k [1 - p_o(i)]^{S-k} \right\}. \quad (15)$$

Note that at most  $S$  nodes can transmit data packets in a frame. Hence,  $g(i)$  is bounded and the number of actually transmitting nodes is equal to  $\min\{g(i), S\}$ .

The conditional expectation of the number of nodes that successfully transmit a data packet in a frame, given that the system is in state  $\mathbf{N}$ , is given by

$$Z(\mathbf{N}) = \sum_{i=0}^M g(i) \cdot P_i, \quad (16)$$

$$Z(\mathbf{N}) = \sum_{i=0}^M D \left\{ \sum_{k=0}^R k \binom{S}{k} p_o(i)^k [1 - p_o(i)]^{S-k} + R \sum_{k=R+1}^S \binom{S}{k} p_o(i)^k [1 - p_o(i)]^{S-k} \right\} \times \binom{M}{i} q^i (1 - q)^{M-i}. \quad (17)$$

Using Eqs. (5)–(8) we can modify Eqs. (3) and (10). Eq. (3) becomes

$$Z(\mathbf{N}) = \frac{\sigma}{1 + \sigma} \left[ N - \left(1 + \frac{\tau p}{D}\right) N_B \right] \quad (18)$$

and Eq. (10) becomes

$$Y(\mathbf{N}) = N_B \frac{p}{D} \left(1 - \frac{p}{D \cdot M}\right)^{N_B - 1}. \quad (19)$$

Eqs. (17)–(19) can be solved simultaneously for the variables  $N_B$ ,  $Y(\mathbf{N})$  and  $Z(\mathbf{N})$ . The system is unstable if there is more than one solution. Otherwise, if only one solution exists, the system is stable.  $N_B$ ,  $Y(\mathbf{N})$  and  $Z(\mathbf{N})$  can then be used to provide the steady-state solution of the entire system.  $N_{PR_i}$ ,  $i = 1, 2, \dots, \tau$ , is given by Eqs. (1) and (5). Similarly,  $N_{PQ_i}$ ,  $i = 1, 2, \dots, \tau$ , is given by Eqs. (2) and (6). According to Eq. (7),  $N_{TR}$  is equal to  $Z(\mathbf{N})$ . And  $N_{TH}$  equals  $N$  minus the sum of the nodes in all other modes.

## 6.2. Performance measures

The performance measures of interest are throughput and delay at an equilibrium point. The throughput  $S(\mathbf{N})$  is defined as the expected number of nodes in the active mode  $TR$ :

$$S(\mathbf{N}) = N_{TR}. \quad (20)$$

The mean packet delay  $D(\mathbf{N})$  is measured from the time the packet is generated at a node until the end of the frame during which it is transmitted. The system shown in Fig. 12 is a closed system, and, by Little's law,  $N/S(\mathbf{N})$  is the average time that a packet experiences from the moment the packet enters mode  $TH$  until the time it returns to mode  $TH$ . Also,  $1/\sigma$  is the average time that a packet stays in mode  $TH$ . Thus, we get the average packet delay as

$$D(\mathbf{N}) = \frac{N}{S(\mathbf{N})} - \frac{1}{\sigma}. \quad (21)$$

Note that  $D(\mathbf{N})$  is measured in number of frames.

## 7. Numerical results

In this section, we investigate the impact of the system parameters on the throughput-delay performance of the network by varying them around the default values. We consider number of nodes  $N$ , retransmission probability  $p$ , physical degree of the AWG  $D$ , propagation delay  $\tau$ , number of FSRs  $R$ , and number of reservation slots per frame  $M$ . Unless stated otherwise, the parameters are set to the following default values:  $N = 240$ ,  $p = 0.5$ ,

$D = 2$ ,  $\tau = 10$  frames,  $R = 3$ , and  $M = 8$ . (Note that  $F$  denotes the frame length in slots; we use the frame length as basic time unit in our performance evaluation.) To verify the accuracy of our analysis we simulated a more realistic system. As opposed to the analysis, in the simulation the first-time transmission of a control packet is not delayed and the destination of a collided control packet is not renewed each time it is retransmitted. Each simulation was run for  $10^6$  cycles including a warm-up phase of  $10^5$  cycles. We used the method of batch means to obtain confidence intervals for the mean throughput and the mean delay. For all simulation results the 98% confidence interval was less than 1% of the sample mean.

Fig. 13 depicts the mean throughput (mean number of transmitting nodes) vs. the mean arrival rate  $\sigma$  (packet/frame) for different populations  $N$ . We observe that the maximum mean throughput is the same for all populations. However, with more nodes this maximum is reached at smaller  $\sigma$ . This is because the mean throughput depends not only on  $\sigma$  alone but on the offered traffic load which is

equal to  $N \cdot \sigma$ . Consequently, with a larger  $N$  the maximum mean throughput occurs at a smaller  $\sigma$ . For  $N \in \{200, 300\}$  the average throughput decreases with increasing  $\sigma$ . This is due to the fact that with increasing  $\sigma$  there are more nodes accessing the control channel. This causes more collisions and fewer scheduled data packets. The congestion becomes more serious for larger  $N$ , resulting in a lower average throughput. Note that with  $N = 100$  the mean throughput is not reduced for increasing  $\sigma$ . In this case,  $N$  is small enough such that the offered load can be handled well by slotted ALOHA. This indicates that slotted ALOHA does not degrade the mean throughput as long as the population is small enough. Analysis and simulation results match very well for small values of  $\sigma$ , for larger  $\sigma$  there is some discrepancy. Note that generally for  $N = 100$  the simulation gives higher throughputs than the analysis; whereas for  $N \in \{200, 300\}$  we observe the opposite. This is due to the assumption of delayed first-time transmissions of control packets that we made in our analysis. For large populations the

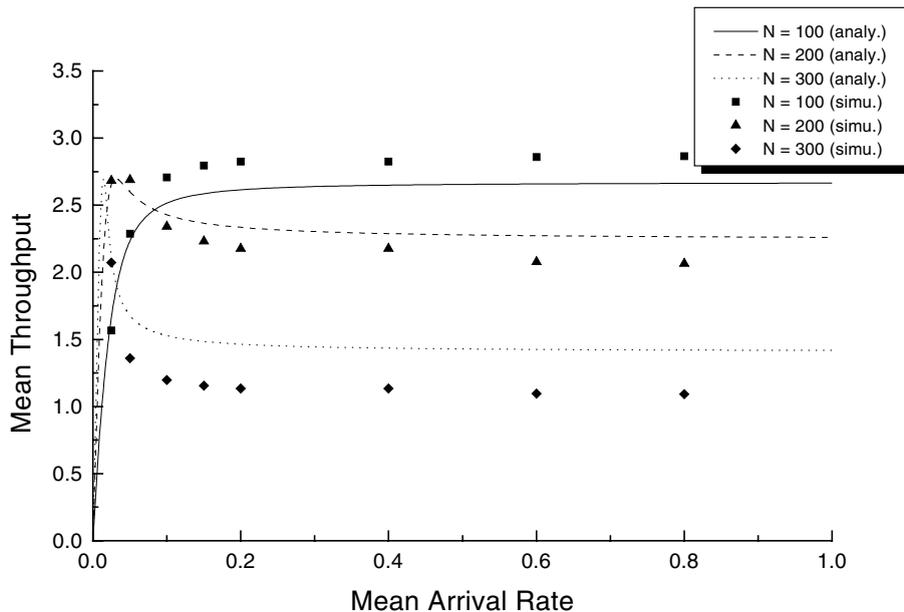


Fig. 13. Mean throughput (mean number of transmitting nodes) vs. mean arrival rate  $\sigma$  (packet/frame) for different numbers of nodes  $N \in \{100, 200, 300\}$ .

number of collisions on the control channel is reduced by delaying also first-time transmissions, resulting in a higher throughput. But for small populations this delay leads to an underutilized control channel and thereby smaller throughput. Hence, this assumption yields accurate results only for moderate loads  $N \cdot \sigma$ .

The mean packet delay (in frames) vs. the mean arrival rate is shown in Fig. 14. For all populations  $N$  the average delay is bounded because we consider single-packet buffers at each node without accounting for queueing delays. New packets arriving at a backlogged node are discarded and do not contribute to the mean delay. With more nodes the slotted ALOHA channel gets congested already at low loads. This causes high delays due to retransmissions of control packets. We can see that with  $N = 100$  the mean delay does not change much with increasing  $\sigma$ . This is because the offered load does not significantly overload slotted ALOHA leading to only a few collisions and retransmissions which in turn keeps the mean delay small. The figure shows, that for large populations ( $N = 300$ ) the simulation gives larger delays than the analysis. This is again due to the fact that in

the simulation first-time transmissions of control packets are not delayed leading to more collisions and more retransmissions of control packets.

Figs. 13 and 14 can be combined as illustrated in Fig. 15. This figure depicts the mean delay vs. mean throughput as  $\sigma$  is varied from 0 to 1. In the following we consider only this kind of graph in order to save space. The subsequent curves are obtained for  $N = 240$ .

A given number of nodes can be connected by AWGs with different physical degree  $D$ . As depicted in Fig. 16, a  $4 \times 4$  AWG yields higher maximum mean throughput and lower mean delay at high traffic loads than a  $2 \times 2$  AWG. Using an  $8 \times 8$  AWG, instead, increases the maximum average throughput only slightly but suffers from a larger average delay for lower arrival rates. A large  $D$  implies that the degree  $S$  of each combiner is small. Accordingly, in a given frame fewer nodes access the slotted ALOHA channel reducing the contention and thereby improving the performance. However, with a  $D$  chosen too large the cycle becomes too long and nodes have to wait a longer time period, resulting in an increased delay. In addition, owing to the longer cycle length more

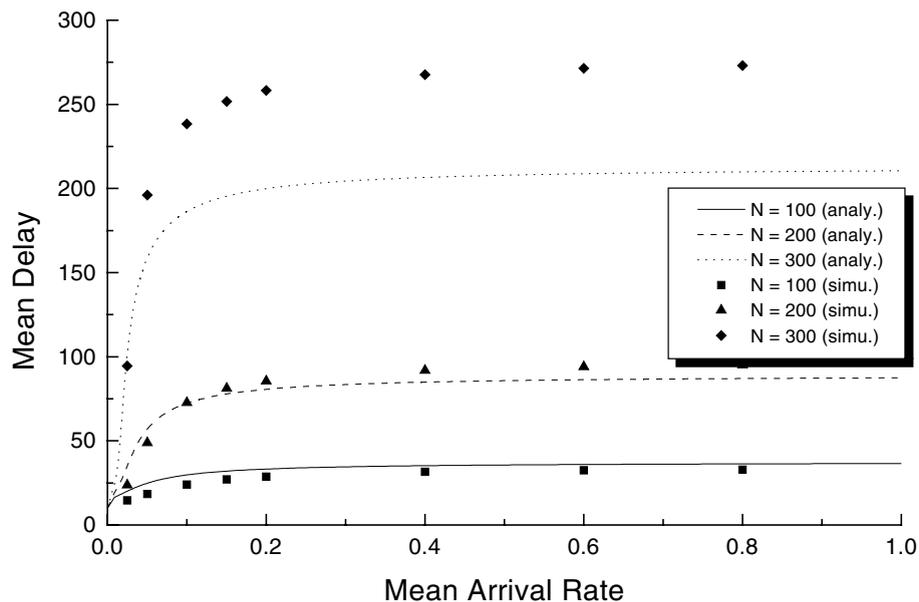


Fig. 14. Mean delay (in frames) vs. mean arrival rate  $\sigma$  (packet/frame) for different numbers of nodes  $N \in \{100, 200, 300\}$ .

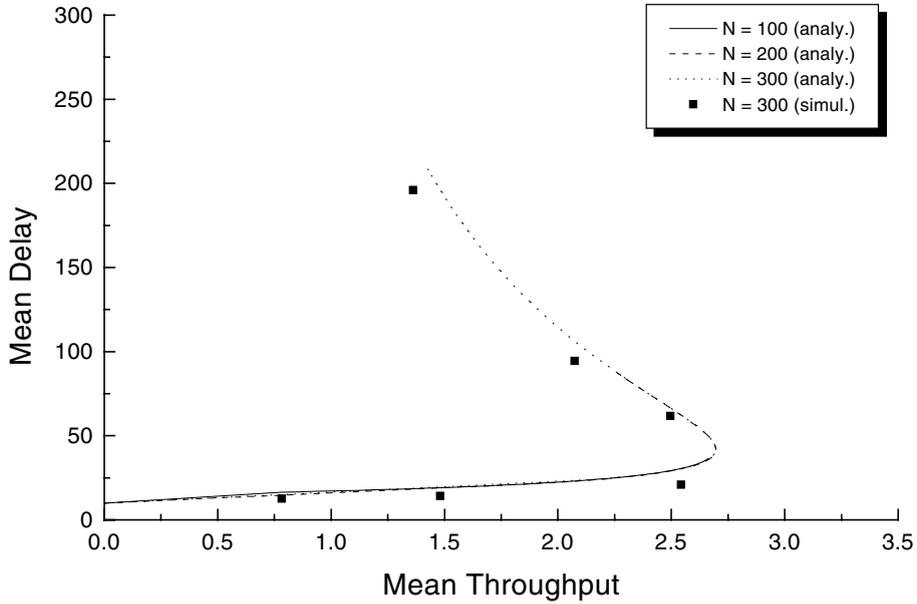


Fig. 15. Mean delay (in frames) vs. mean throughput (mean number of transmitting nodes) for different numbers of nodes  $N \in \{100, 200, 300\}$ .

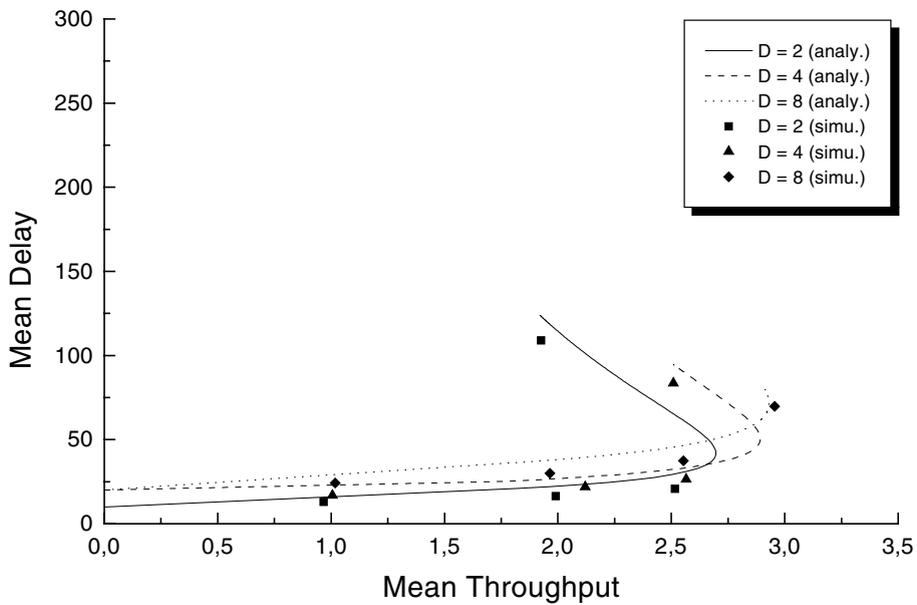


Fig. 16. Mean delay (in frames) vs. mean throughput (mean number of transmitting nodes) for different AWG degrees  $D \in \{2, 4, 8\}$ .

nodes are backlogged and try to access the same frame increasing the number of collisions and

limiting the throughput improvement. Thus, unless running the system under heavy traffic, it is

reasonable to use an AWG with a rather moderate number of ports. Note that this allows to exploit

more FSRs of an AWG for a given transceiver tuning range and channel spacing.

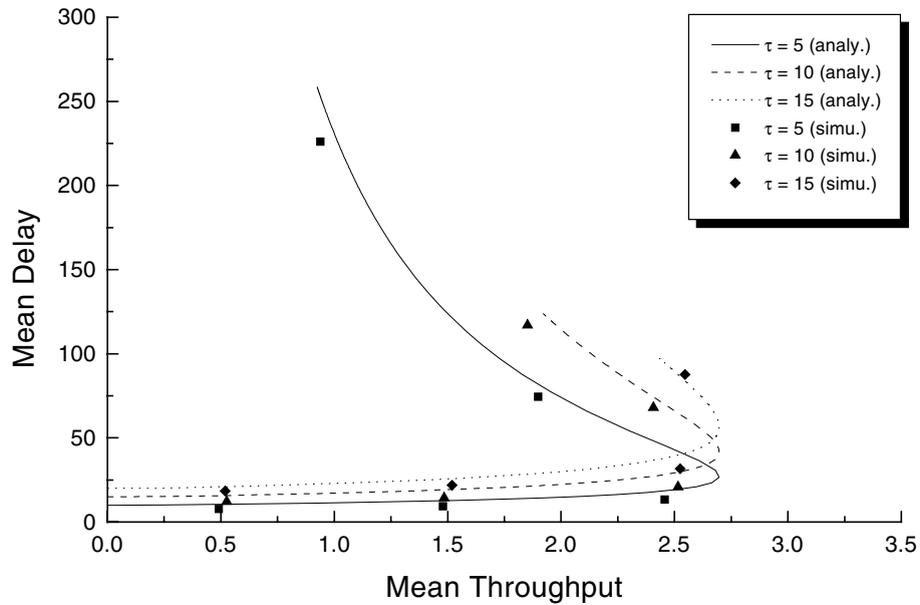


Fig. 17. Mean delay (in frames) vs. mean throughput (mean number of transmitting nodes) for different propagation delays  $\tau \in \{5, 10, 15\}$  frames.

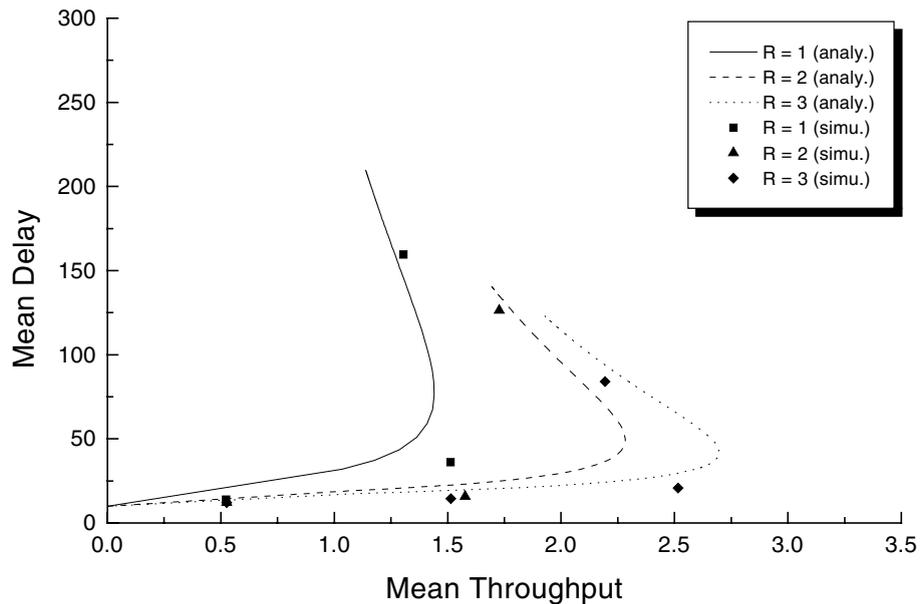


Fig. 18. Mean delay (in frames) vs. mean throughput (mean number of transmitting nodes) for different number of FSRs  $R \in \{1, 2, 3\}$ .

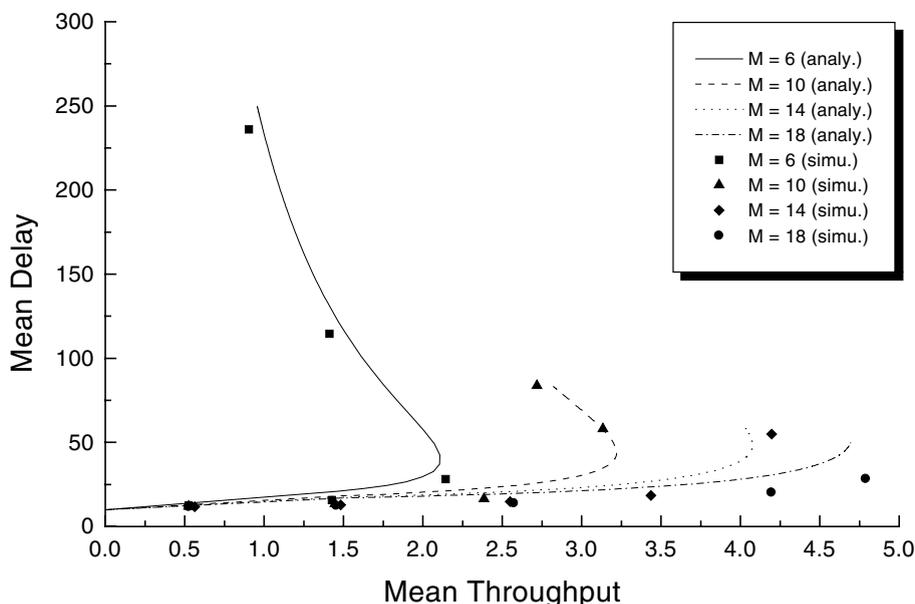


Fig. 19. Mean delay (in frames) vs. mean throughput (mean number of transmitting nodes) for different number of reservation slots  $M \in \{6, 10, 14, 18\}$ .

Fig. 17 depicts the impact of the propagation delay on the network performance. At low traffic loads packets experience less delay for smaller propagation delays. Whereas with increasing traffic larger propagation delays provide a better throughput-delay performance. This is due to the fact that at low traffic loads almost no collisions of control packets occur and nodes receive the successfully transmitted control packets earlier with smaller propagation delays, resulting in a decreased delay. At higher traffic loads the control channel gets more congested. In this case, a larger propagation delay implies that nodes have to wait a longer time interval for the transmitted control packets. During this time period those nodes do not access the control channel, resulting in less contention and an increased throughput and a decreased delay due to fewer retransmissions.

The results in Fig. 18 clearly demonstrate the benefit of using multiple FSRs of an AWG. Each additional FSR increases the degree of concurrency and thereby alleviates the scheduling bottleneck, resulting in a significantly improved throughput-delay performance of the network.

Using three FSRs instead of one improves the maximum mean throughput by approximately 88%. However, using multiple FSRs requires transceivers with a larger tuning range.

Fig. 19 shows that a large reservation window has a positive impact on the system performance. By using many reservation slots the collision probability in each slot is reduced and the number of successful control packets is increased. Consequently, fewer control packets have to be retransmitted improving the throughput-delay performance. However, long reservation windows in conjunction with enough slots available for spatial wavelength reuse might require large data packets. These larger data packets could be realized by means of traffic aggregation (grooming). Note that for  $M = 18$  a normalized mean throughput of up to 78% is achieved.

## 8. Conclusions

The proposed WDM network combines the merits and mitigates the drawbacks of wavelength

routing and broadcast-and-select devices. While the AWG dramatically improves the throughput-delay performance by allowing for spatial spectrum reuse and concurrent transmission over multiple FSRs, the splitters and combiners are required to provide additional ports and to support efficient multicasting. In the proposed architecture we use a broadband light source which is spectrally sliced by the AWG to realize broadcasting of control information. By means of direct sequence spread spectrum no separate control channel is needed and data and control can be sent and received simultaneously. CDMA allows the concurrent transmission of multiple control signals. The proposed MAC protocol enables each node to acquire global knowledge at any time, despite the fact that each node has only one single tunable transceiver. By applying a distributed deterministic scheduling algorithm the signaling overhead is significantly reduced and data packets experience neither channel nor receiver collisions. The system exhibits two bottlenecks: (1) The modified slotted ALOHA reservation channel and (2) the relatively small scheduling window. While the latter one is dramatically alleviated by exploiting multiple FSRs of the underlying AWG, the first one could be relaxed by aggregating data packets. Longer data packets would enlarge the frame size and thereby accommodate more reservation slots in each frame.

Finally, we note that the emphasis of our network and protocol design was primarily on minimizing costs and complexity. Nevertheless, a mean throughput of up to 78% is rather high and could be further improved at the cost of additional hardware and complexity.

### Acknowledgements

This work was supported in part by Deutsche Telekom AG. A shorter version of this paper was presented at Terabit Optical Networking: Architecture, Control, and Management Issues—Part of SPIE Photonics East 2000, Boston, MA, November 2000, and won the Best Paper Award of the conference.

### References

- [1] M. Listanti, V. Eramo, R. Sabella, Architectural and technological issues for future optical Internet networks, *IEEE Commun. Mag.* 38 (9) (2000) 82–92.
- [2] S. Mokbel, Canada's optical research and education network: CA\*net3, *Proc. Design of Reliable Commun. Networks (DRCN)*, Munich, Germany, April 2000, pp. 10–32.
- [3] M. Jaeger, H.-M. Foisel, F.-J. Westphal, J. Chawki, et al., Evaluation of network architectures for the integration of IP over optical networks, *Proc. Design of Reliable Commun. Networks (DRCN)*, Munich, Germany, April 2000, pp. 261–266.
- [4] C. Qiao, M. Yoo, A taxonomy of switching techniques, in: K.M. Sivalingam, S. Subramaniam (Eds.), *Optical WDM Networks—Principles and Practice*, Kluwer Academic Publishers, Dordrecht, 2000, pp. 103–125 (Chapter 5).
- [5] S. Yao, B. Mukherjee, Advances in photonic packet switching: an overview, *IEEE Commun. Mag.* 38 (2) (2000) 84–94.
- [6] C. Qiao, Labeled optical burst switching for IP-over-WDM integration, *IEEE Commun. Mag.* 38 (9) (2000) 104–114.
- [7] T. Matsumoto, H. Ishio, Multiple-access optical network architecture employing a wavelength-and-network-division technique: MANDALA, *IEICE Trans. Commun. E* 82-B (9) (1999) 1439–1445.
- [8] N.P. Caponio, A.M. Hill, F. Neri, R. Sabella, Single-layer optical platform based on WDM/TDM multiple access for large-scale “switchless” networks, *Eur. Trans. Telecommun.* 11 (1) (2000) 73–82.
- [9] A.M. Hill, Unconstrained signalling in the “switchless” WDM/TDMA optical transport network, *Proc. Terabit Optical Networking: Architecture, Control, and Management Issues—Part of SPIE Photonics East*, vol. 4213, Boston, MA, November 2000, pp. 209–219.
- [10] A. Bianco, E. Leonardi, M. Mellia, F. Neri, Network controller design for SONATA—a large-scale all-optical passive network, *IEEE J. Sel. Areas Commun.* 18 (10) (2000) 2017–2028.
- [11] D. Banerjee, J. Frank, B. Mukherjee, Passive optical network architecture based on waveguide grating routers, *IEEE J. Sel. Areas Commun.* 16 (7) (1998) 1040–1050.
- [12] G. Maier, M. Martinelli, A. Pattavina, E. Salvadori, Design and cost performance of the multistage WDM-PON access networks, *IEEE J. Lightwave Technol.* 18 (2) (2000) 125–143.
- [13] M. Maier, A. Wolisz, Demonstrating the potential of arrayed-waveguide grating based single-hop WDM networks, *Opt. Networks Mag.* 2 (5) (2001) 75–85.
- [14] B. Mukherjee, WDM-based local lightwave networks—Part I: single-hop systems, *IEEE Network Mag.* 6 (3) (1992) 12–27.

- [15] T. Pfeiffer, H. Schmuck, B. Deppisch, M. Witte, et al., TDM/CDM/WDM approach for metro networks with 200 optical channels, Proc. ECOC 2000, vol. 3, Munich, Germany, September 2000, pp. 77–78.
- [16] B. Mukherjee, WDM optical communication networks: progress and challenges, IEEE J. Sel. Areas Commun. 18 (10) (2000) 1810–1824.
- [17] F. Ruehl, T. Anderson, Cost-effective metro WDM network architectures, Technol. Digest OFC 2001, Anaheim, CA, paper WL1, March 2001.
- [18] K. Kato, A. Okada, Y. Sakai, K. Noguchi, et al., 10-Tbps full-mesh WDM network based on cyclic-frequency arrayed-waveguide grating router, Proc. ECOC 2000, vol. 1, Munich, Germany, September 2000, pp. 105–107.
- [19] A. Okada, T. Sakamoto, Y. Sakai, 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, Technol. Digest OFC 2001, Anaheim, CA, paper ThG5, March 2001.
- [20] C.F. Lam, K.C. Reichmann, P.P. Iannone, Cascadable modular transmitter and receiver for delivering multiple broadcast services on WDM passive optical networks, Proc. ECOC 2000, vol. 1, Munich, Germany, September 2000, pp. 109–110.
- [21] H. Takahashi, K. Oda, H. Toba, Y. Inoue, Transmission characteristics of arrayed waveguide  $N \times N$  wavelength multiplexer, IEEE/OSA J. Lightwave Technol. 13 (3) (1995) 447–455.
- [22] Y. Tachikawa, Y. Inoue, M. Kawachi, H. Takahashi, K. Inoue, Arrayed-waveguide grating add-drop multiplexer with loop-back optical paths, Electron. Lett. 29 (24) (1993) 2133–2134.
- [23] Y. Tachikawa, Y. Inoue, M. Ishii, T. Nozawa, Arrayed-waveguide grating multiplexer with loop-back optical paths and its applications, IEEE/OSA J. Lightwave Technol. 14 (6) (1996) 977–984.
- [24] B. Glance, I.P. Kaminow, R.W. Wilson, Applications of the integrated waveguide grating router, IEEE/OSA J. Lightwave Technol. 12 (6) (1994) 957–962.
- [25] K.A. McGreer, Arrayed waveguide gratings for wavelength routing, IEEE Commun. Mag. (December 1998) 62–68.
- [26] J. Gripp, P. Bernasconi, C. Chan, K. Sherman, et al., Demonstration of a 1Tb/s optical packet switch fabric ( $80 * 12.5$  Gb/s), scalable to 128 Tb/s ( $6400 * 20$  Gb/s), ECOC 2000, Post Deadline Paper 2.7, Munich, Germany, September 2000.
- [27] M. Zirngibl, Multifrequency lasers and applications in WDM networks, IEEE Commun. Mag. (1998) 39–41.
- [28] A. Franzen, D.K. Hunter, I. Andonovic, A low loss optical packet synchronisation architecture, Proc. SPIE, vol. 3531, Boston, MA, November 1998, pp. 390–395.
- [29] H. Tsuda, H. Takenouchi, A. Hirano, T. Kurokawa, et al., Performance analysis of a dispersion compensator using arrayed-waveguide gratings, IEEE/OSA J. Lightwave Technol. 18 (8) (2000) 1139–1147.
- [30] A. Hirano, K. Yonenaga, Y. Miyamoto, H. Toba, et al., 640 Gbit/s (16 channel  $\times$  42.7 Gbit/s) WDM L-band DSF transmission experiment using 25 nm bandwidth AWG dispersion slope compensator, Electron. Lett. 36 (19) (2000) 1638–1639.
- [31] L. Bersiner, D. Rund, Bidirectional WDM transmission with spectrum sliced LEDs, J. Opt. Commun. 11 (2) (1990) 56–59.
- [32] J.C. Lu, L. Kleinrock, On the performance of wavelength division multiple access networks, Proc. International Conf. Commun. (ICC)'92, June 1992, pp. 1151–1157.
- [33] L. Giehmann, A. Gladisch, N. Hanik, J. Rudolph, O. Ziemann, The application of code division multiple access for transport overhead information in transparent optical networks, Technol. Digest OFC'98, San Jose, CA, February 1998, pp. 228–229.
- [34] B. Mukherjee, Optical Communication Networks, McGraw-Hill, New York, 1997.
- [35] R. Ramaswami, K.N. Sivarajan, Optical Networks—A Practical Perspective, Morgan Kaufmann, San Mateo, CA, 1998.
- [36] H. Takahashi, K. Oda, H. Toba, Impact of crosstalk in an arrayed-waveguide multiplexer on  $N \times N$  optical interconnection, IEEE/OSA J. Lightwave Technol. 14 (6) (1996) 1097–1105.
- [37] E.L. Goldstein, L. Eskildsen, Scaling limitations in transparent optical networks due to low-level crosstalk, IEEE Photon. Technol. Lett. 7 (1) (1995) 93–94.
- [38] L. Gillner, C.P. Larsen, M. Gustavsson, Scalability of optical multiwavelength switching networks: crosstalk analysis, IEEE/OSA J. Lightwave Technol. 17 (1) (1999) 58–67.
- [39] K. Kato, A. Okada, Y. Sakai, K. Noguchi, et al., 10-Tbps full-mesh WDM network based on cyclic-frequency arrayed-waveguide grating router, Proc. ECOC 2000, vol. 1, Munich, Germany, September 2000, pp. 105–107.
- [40] B. Mason, Widely tunable semiconductor lasers, Proc. ECOC 2000, vol. 2, Munich, Germany, September 2000, pp. 157–158.
- [41] K.C. Reichmann, P.P. Iannone, N.J. Frigo, Operational demonstration and filter alignment study of multiple broadcast video delivery on a WDM passive optical network, IEEE Photon. Technol. Lett. 10 (9) (1998) 1331–1333.
- [42] D.K. Jung, S.K. Shin, C.-H. Lee, Y.C. Chung, Wavelength-division-multiplexed passive optical network based on spectrum-slicing techniques, IEEE Photon. Technol. Lett. 10 (9) (1998) 1334–1336.
- [43] S.L. Woodward, P.P. Iannone, K.C. Reichmann, N.J. Frigo, A spectrally sliced PON employing Fabry-Perot lasers, IEEE Photon. Technol. Lett. 10 (9) (1998) 1337–1339.
- [44] T. Pfeiffer, Highspeed optically modulated broadband light source, Proc. ECOC 2000, vol. 3, Munich, Germany, September 2000, pp. 51–52.

- [45] L. Boivin, B.C. Collings, Spectrum-slicing of coherent sources for optical communications, Proc. ECOC 2000, vol. 3, Munich, Germany, September 2000, pp. 49–50.
- [46] S.S. Wagner, T.E. Chapuran, Broadband high-density WDM transmission using superluminescent diodes, *Electron. Lett.* 26 (11) (1990) 696–697.
- [47] I.M.I. Habbab, M. Kavehrad, C.-E.W. Sundberg, Protocols for very high-speed optical fiber local area networks using a passive star topology, *IEEE/OSA J. Lightwave Technol.* LT-5 (12) (1987) 1782–1794.
- [48] N. Mehravari, Performance and protocol improvements for very high speed optical fiber local area networks using a passive star topology, *IEEE/OSA J. Lightwave Technol.* 8 (4) (1990) 520–530.
- [49] M. Chen, T.-S. Yum, A conflict-free protocol for optical WDM networks, Proc. IEEE Globecom'91, Phoenix, AZ, December 1991, pp. 1276–1281.
- [50] H.B. Jeon, C.K. Un, Contention-based reservation protocol in fibre optic local area network with passive star topology, *Electron. Lett.* 26 (12) (1990) 780–781.
- [51] H.W. Lee, Protocols for multichannel optical fibre LAN using passive star topology, *Electron. Lett.* 27 (17) (1991) 1506–1507.
- [52] L. Kleinrock, F.A. Tobagi, Packet switching in radio channels: Part I—carrier sense multiple-access modes and their throughput-delay characteristics, *IEEE Trans. Commun.* COM-23 (12) (1975).
- [53] E. Modiano, Unscheduled multicasts in WDM broadcast-and-select networks, Proc. IEEE INFOCOM'98, 1998, pp. 86–93.
- [54] D.A. Levine, I.F. Akyildiz, PROTON: A media access control protocol for optical networks with star topology, *IEEE/ACM Trans. Networking* 3 (2) (1995) 158–168.
- [55] E. Modiano, A novel architecture and medium access control (MAC) protocol for WDM networks, Technol. Digest OFC'98, San Jose, CA, February 1998, pp. 90–91.
- [56] F. Jia, B. Mukherjee, The receiver collision avoidance (RCA) protocol for a single-hop WDM lightwave network, *IEEE/OSA J. Lightwave Technol.* 11 (5/6) (1993) 1053–1065.
- [57] M.J. Spencer, M.A. Summerfield, WRAP: A media access control protocol for wavelength-routed passive optical networks, Proc. ECOC 2000, vol. 3, Munich, Germany, September 2000, pp. 85–86.
- [58] F. Jia, B. Mukherjee, A high-capacity, packet-switched, single-hop local lightwave network, Proc. IEEE Globecom'93, Houston, TX, December 1993, pp. 1110–1114.
- [59] J. Lu, L. Kleinrock, A wavelength division multiple access protocol for high-speed local area networks with a passive star topology, *Perform. Eval.* 16 (1–3) (1992) 223–239.
- [60] A. Fukuda, S. Tasaka, The equilibrium point analysis—a unified analytic tool for packet broadcast networks, Proc. IEEE GLOBECOM'83, San Diego, CA, November 1983, pp. 1133–1140.



**Martin Maier** received the B.S. degree in electrical engineering and the Dipl.-Ing. degree in electrical engineering with distinctions from the Technical University of Berlin in 1994 and 1998, respectively. He works currently toward the Ph.D. degree as a research and teaching assistant with the TKN group at the Technical University of Berlin. He was a recipient of the Deutsche Telekom scholarship from June 1999 through May 2001. As a visiting researcher he spent spring 1998 at USC in L.A., CA, and winter 2001 at ASU in Tempe, AZ. He is a co-recipient of a best paper award presented at the SPIE Photonics East 2000 conference. His research interests include switching/routing techniques, architectures and protocols for optical WDM networks.



**Martin Reisslein** is an Assistant Professor in the Department of Electrical Engineering at Arizona State University, Tempe. He is affiliated with ASU's Telecommunications Research Center. He received the Dipl.-Ing. (FH) degree from the Fachhochschule Dieburg, Germany, in 1994, and the M.S.E. degree from the University of Pennsylvania, Philadelphia, in 1996. Both in electrical engineering. He received his Ph.D. in systems engineering from the University of Pennsylvania in 1998. During the academic year 1994–1995 he visited the University of Pennsylvania as a Fulbright scholar. From July 1998 through October 2000 he was a scientist with the German National Research Center for Information Technology (GMD FOKUS), Berlin. While in Berlin he was teaching courses on performance evaluation and computer networking at the Technical University Berlin. He is associate editor-in-chief of the *IEEE Communications Surveys and Tutorials* and has served on the Technical Program Committees of *IEEE Infocom*, *IEEE Globecom*, and the *IEEE International Symposium on Computer and Communications*. He has organized sessions at the *IEEE Computer Communications Workshop (CCW)*. He maintains an extensive library of video traces for network performance evaluation, including frame size traces of MPEG-4 and H.263 encoded video, at <http://www.eas.asu.edu/trace>. He is co-recipient of the Best Paper Award of the *SPIE Photonics East 2000—Terabit Optical Networking* conference. His research interests are in the areas of Internet Quality of Service, video traffic characterization, wireless networking, and optical networking.



**Adam Wolisz** is currently a Professor of Electrical Engineering and Computer Science (secondary assignment) at the Technical University Berlin, where he is directing the Telecommunication Networks Group (TKN). Parallely, he is also member of the Senior Board of GMD Fokus, being especially in charge of the Competence Centers GLONE and TIP. His teaching activities encompass courses on Communication Networks and Protocols, High Speed Networks, Wireless Networks and Performance Analysis of Communication Networks. He is acting as a member of the Steering Committee of the Computer Engi-

neering Curriculum at the Technical University Berlin. He participates in the nationally (Deutsche Forschungsgemeinschaft) founded Graduate Course in Communication-Based Systems.

His research interests are in architectures and protocols of communication networks as well as protocol engineering with impact on performance and QoS aspects. Recently he is working mainly on mobile multimedia communication, with special regard to architectural aspects of network heterogeneity

and integration of wireless networks in the Internet. The research topics are usually investigated by a combination of simulation studies and real experiments.

He has authored two books and authored or co-authored over 100 papers in technical journals and conference proceedings. He is Senior Member of IEEE, IEEE Communications Society (including the TCCC and TCPC) as well as GI/ITG Technical Committee on Communication and Distributed Systems.