Towards efficient packet switching metro WDM networks

Martin Maier
Telecommunication Networks Group
Technical University Berlin
Berlin, Germany

Martin Reisslein
Telecommunications Research Center
Department of Electrical Engineering
Arizona State University
Tempe, Arizona USA

Adam Wolisz
Telecommunication Networks Group
Technical University Berlin
Berlin, Germany

ABSTRACT
In this article, we first give a brief overview of the current state-of-the-art of optical WDM networking. In this overview we point out that the ubiquitous SONET/SDH-WDM metro rings are likely to become a serious bottleneck between high-speed local area networks/broadband access technologies and WDM backbone networks. Next, we discuss innovative metro WDM network architectures and formulate the requirements of future packet switching metro WDM networks. After reviewing previous work on single-hop WDM networks, several resulting guidelines for the design of Medium Access Control (MAC) protocols are provided. Following these guidelines leads us to a future-proof Arrayed-Waveguide Grating (AWG) based metro WDM network and a novel reservation based MAC protocol, which supports Quality-of-Service (QoS). Each node at the periphery of this passive single-hop WDM network is equipped with a single tunable transceiver for data and a broadband light source, which is spectrally sliced in order to broadcast reservation requests. Each node has global knowledge and schedules variable-size data packets in a distributed fashion by executing the same greedy first-come-first-served and first-fit arbitration algorithm. The network is reliable, scalable, and allows for optical multicasting. All wavelengths are used for data transmission. The protocol supports both packet and circuit switching. The network efficiency is significantly increased by using multiple Free Spectral Ranges (FSRs) of the underlying AWG, spatially reusing wavelengths, and transmitting data and control simultaneously by means of spreading techniques.

1 Introduction

As illustrated in Fig. 1, the hierarchy of communication networks can be viewed as consisting of backbone, metro, and access networks where the latter ones collect (distribute) data from (to) different clients such as wireless stations and LANs. Let us first consider backbone networks. Future optical WDM backbone networks seem to be converging to a two-layer infrastructure in which the transport is by means of interconnected all-optical *islands of transparency* while the remainder of the communication layers are based on IP [1]. The islands of transparency are formed out of wavelength-division multiplexing (WDM) links, stitched together by all-optical cross connects (OXCs) and add-drop multiplexers (OADMs). The optical-electrical-optical (OEO) boundaries between islands will be retained due to management, jurisdiction, billing, and/or signal regeneration issues. Such islands not only provide huge pipes of "infinite bandwidth" but also transparency to data rates, modulation formats, and protocols. Moreover, transparent islands lead to significantly simplified network management and large cost savings which is one of the most important drivers for optical networking. While early OADMs and OXCs provided only frozen paths, Micro-Electro-Mechanical-Systems (MEMS) can now be used to realize reconfigurable OADMs and OXCs. Those devices make the network more flexible and robust against link and/or node failures. Recently, much research has been focusing on controlling such reconfigurable devices. Multiprotocol Label Switching (MPLS) routing and signaling protocols have been extended and modified in order to enable Optical Label Switching (OLS) (using wavelengths as labels leads to a specific kind of OLS termed MPλS) and to enrich optical WDM networks with point-and-click provisioning [2], protection [3], restoration [4], traffic engineering [5], and other future services, e.g., rent-a-wavelength. However, since MEMS have a switching time of about 10 ms only circuit switching can be realized. Those circuits correspond to wavelengths and can be either permanent or reconfigured, for example twice a day in order to provide sufficient bandwidth to offices during the day and to residential areas in the evening. Future OXCs and OADMs might exhibit a significantly improved switching time of a few nanoseconds deploying semiconductor optical amplifiers (SOAs) or lithium-niobate based components.

In the meantime, Optical Burst Switching (OBS) seems to be a viable intermediate step between current lambda switching and future packet switching [6]. OBS aggregates multiple packets, e.g., IP packets, into bursts at the optical network edge and makes a one-way reservation by sending a control packet prior to transmitting the corresponding data burst. OBS does not require buffering at intermediate nodes and allows for statistical multiplexing. In addition, OBS is also able to provide differentiated services by controlling the offset time between control packet and data burst and/or by using Fiber Delay Lines (FDLs) at intermediate nodes [7,8]. In order to further improve the switching granularity, OBS is likely to be followed by Optical Packet Switching (OPS), both in backbone and in metro WDM networks [9-11].

Now, let us take a look at the network periphery as depicted in Fig. 1. Current Gigabit Ethernet (GbE) LANs together with the 10GbE standard completed in early 2002 are expected to provide sufficient bandwidth for at least the next few years. Both the telcos and cable providers are steadily moving the fiber-to-copper discontinuity point out toward the network periphery [1]. Phone companies typically deploy some form of Digital Subscriber Loop (DSL) and cable companies deploy cable modems. While in the past only large businesses have had fiber between their premises, residential areas will be connected by their own fiber in the near future. These broadband access technologies in conjunction with next-generation wireless services, e.g., UMTS, will require a huge amount of bandwidth and Quality-of-Service (QoS) support. To improve response time, an increasing part of IP traffic will be local by placing more proxies in the metro networks. In addition—Napster being simply a precursor—future *peer-to-peer* applications where each attached user will also operate as a server will dramatically increase the amount of local traffic.

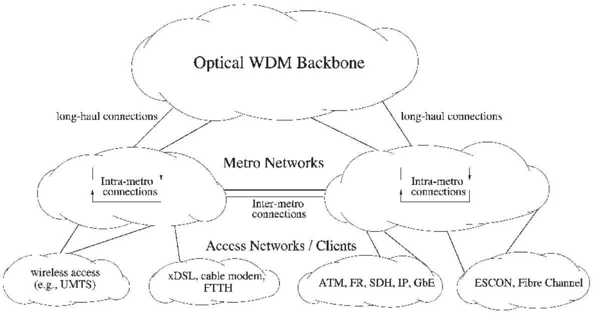

Figure 1: Network hierarchy.

Between those high-speed clients and the huge pipes in the backbone lies a severe bandwidth bottleneck at the metro level. This is due to the fact that initially the bandwidth need was felt mainly in the backbone and consequently most of the WDM research has been focused on upgrading the long-haul transport network. In contrast, current metro networks are mostly SONET/SDH over WDM rings which can carry the ever increasing amount of data traffic only very inefficiently, resulting in the so-called *metro-gap*. This gap prevents the clients from taking benefit of the abundant bandwidth available in the backbone. To bridge this abyss, novel metro architectures and protocols have to be found and applied. In general, future WDM metro networks have to exhibit the following features:

- *Flexibility:* As shown in Fig. 1, metro networks must be able to support a wide range of protocols such as ATM, Frame Relay (FR), SONET/SDH, IP, GbE, Enterprise System Connection (ESCON), and Fibre Channel.
- *Cost-sensitivity:* Metro networks are very cost-sensitive. As a consequence, the network and node architecture has to be simple and protocols must not perform complex operations.
- *Upgradability:* Future-proof metro WDM networks have to be able to incorporate advanced technologies, e.g., tunable lasers with a wider tuning range and a smaller tuning time, in an easy and non-disruptive manner.
- *Scalability:* Nodes must be removed and added in an easy and non-disruptive way.

- *Efficiency:* Future metro WDM networks should be highly efficient by allowing for spatial wavelength reuse and having a small mean hop distance.
- *Connectivity:* To increase efficiency, nodes should be able to communicate directly without any hubs or intermediate nodes between them.
- *Multicasting:* Future metro WDM networks should allow for multicasting in order to efficiently support applications such as videoconferences or distributed games and to efficiently distribute content updates from master to slave proxies.
- *Quality of Service:* QoS is required for mission-critical data and delay-sensitive applications.

Recently, innovative metro WDM networks and components have been presented. Cost-effective, optical add-drop multiplexers for ring metro WDM networks were described in [12]. The OADMs are based on tunable Fiber Bragg Gratings (FBGs) and are suitable for realizing lambda switching. Another photonic ring metro WDM network is HORNET which applies a distributed MAC protocol termed CSMA/CA for supporting photonic packet switching [13]. More efficient and cost-effective alternative metro WDM architectures were presented in [14]. Those networks are based on a passive arrayed-waveguide grating (AWG) and provide a large degree of concurrency by allowing for extensive spatial wavelength reuse. Such networks are basically single-hop WDM networks and have already been realized as optical packet switching network testbeds [15-16]. In this article we concentrate on *single-hop* metro WDM networks since they have the following advantageous properties:

- The mean hop distance is minimum (unity), thereby inherently guaranteeing transparency and increasing the channel utilization which is inversely proportional to the mean hop distance.
- As opposed to multihop networks, no system capacity is wasted due to data forwarding improving the throughput-delay performance of the system.
- Stations have to process only data packets which are addressed to themselves. Hence, protocol processing requirements at each station are reduced which is an important issue in high-speed networks.

The remainder of the article is organized as follows. Section 2 gives an overview of previous work on single-hop WDM networks and Access Control (MAC) protocols. Based on this, in Section 3 we formulate some guidelines for the design of efficient single-hop WDM networks and MAC protocols. Sections 4 and 5 describe a novel metro WDM network architecture and MAC protocol, respectively. Section 6 presents some results and discusses the performance of the proposed protocol. Finally, concluding remarks are given in Section 7.

2 Previous Work

In this section we highlight the pros and cons of some selected previously presented single-hop WDM networks and MAC protocols. The terms station and node are used interchangeably. For an extensive survey the interested reader is referred to [17-18]. Most of those networks are based on a passive star coupler (PSC) forming so-called *broadcast-and-select* networks where all wavelengths are broadcast and the intended destination has to select the corresponding wavelength by tuning its receiver accordingly. Even though the AWG is a *wavelength-sensitive* device as opposed to the PSC, some learnt lessons with respect to PSC based single-hop WDM networks are considered generally valid and helpful for the design of AWG based WDM networks. After discussing the following architectures and protocols, resulting guidelines for the design of single-hop WDM networks and MAC protocols will be formulated in the subsequent section.

The choice of the MAC protocol depends on the components used at the source and destination nodes. These come in four flavors:

- Fixed transmitter(s) and fixed receiver(s) (FT-FR)
- Fixed transmitter(s) and tunable receiver(s) (FT-TR)
- Tunable transmitter(s) and fixed receiver(s) (TT-FR)
- Tunable transmitter(s) and tunable receiver(s) (TT-TR)

According to the notation given in [19], single-hop WDM networks can be described as FT^i-TT^j-FR^m-TR^n systems or CC-FT^i-TT^j-FR^m-TR^n systems if pretransmission coordination takes place over a separate control channel (CC), where $i, j, m, n \geq 0$ denote the number of the respective component present at each node.

As shown in Fig. 2, single-hop MAC protocols can be classified into protocols with fixed channel assignment, random access protocols, and on-demand channel assignment protocols. The latter ones can be further subdivided into so-called tell-and-go and attempt-and-defer access protocols. In tell-and-go protocols the data packet is sent immediately after the corresponding control packet irrespective of the success of the control packet. Whereas in attempt-and-defer protocols data packets are sent only after successfully transmitted control packets.

2.1 Fixed assignment protocols

Frequency-Time Division Multichannel Allocation (FTDMA) protocols were reported in [20]. WDM and

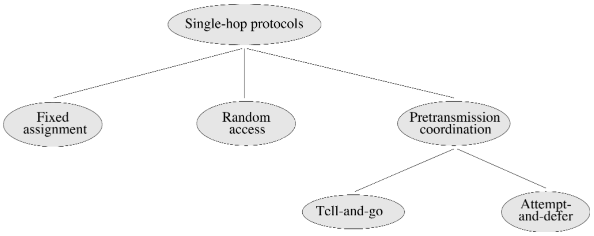

Figure 2: Classification of MAC protocols for single-hop WDM networks.

Time Division Multiplexing (TDM) are combined such that transmission rights are assigned in each slot to source-destination pairs on one of the channels (wavelengths). The protocols offer fixed, partially fixed and random cyclic access policies. They differ in the number of transmission permissions per slot and the degree of channel access control. The basic idea of the FTDMA protocols is the tradeoff between the loss of bandwidth due to channel collisions and destination conflicts and the probability of finding unused channels (idle slots) in a nonempty system. With increasing number of granted permissions per slot the cycle size (number of slots) decreases. All protocols have in common that a node can transmit on and receive from at most one of the channels at any given time. The source/destination allocation protocol is based on a fixed assignment of channels to source/destination pairs by granting W permissions per slot (TT-TR system), where W denotes the number of available wavelengths. Both channel and receiver collisions are prevented. At low to medium loads slots are used only partially. This drawback is alleviated by the destination allocation protocol (TT-FR system). In this case, N permissions per slot are allowed, thus reducing the cycle size. Destination conflicts are eliminated whereas channel collisions can still occur. This protocol provides a more efficient channel access at low to medium loads. The source allocation protocol avoids channel collisions (FT-TR system). In any slot, W sources can transmit each on a different wavelength, to any arbitrary destination. Hence, receiver collisions can still occur. This approach is attractive for low to medium loads, especially when $W \ll N$, where N denotes the number of nodes.

In [21] for a given traffic matrix a fixed cyclic time-wavelength allocation algorithm was proposed which takes into account the number of tunable transceivers, tuning time, and limited number of wavelengths (TT-TR system). The slot duration equals the packet transmission time. The underlying physical topology is PSC based. The protocol prevents channel and receiver collisions with the objective of minimizing the number of tuning operations of the transceivers, thus reducing the tuning penalty (station latency).

This approach was extended in [22]. Since the overall frame duration T is given by

$$T = N_t \cdot \tau + N_s \cdot t \qquad (1)$$

where τ is the tuning time, t is the packet transmission time, N_t is the number of tuning operations and N_s denotes the number of time slots per frame, the objective is to minimize the frame duration by increasing the concurrency of packet transmissions and decreasing the tuning penalty. This problem has been shown to be computationally intractable and has therefore to be split into two subproblems. The first presented algorithm ensures a minimum packet transmission duration while trying to minimize the number of tuning operations whereas the second algorithm ensures a minimum tuning duration while trying to minimize the frame length (number of slots). For small τ/t values the first algorithm exhibits a better performance than the second one, and vice versa for large τ/t values.

The same optimization strategy can be applied to a multiple-star structure where several local stars handle the intra-group communication and one remote star is responsible for the inter-group communication [23]. This kind of configuration is called multilevel network and enables a higher degree of concurrency by means of wavelength reuse.

2.2 Random access protocols

A slotted multichannel random access protocol without control channel for a PSC based topology was introduced in [24]. It is a TT-FR system where each station can transmit on every channel and has its own home channel to receive packets. Two variations of the protocol with different synchronization boundaries are investigated. The first protocol ($SA^{(3)}$) is slotted on minislot boundaries whereas in the second one ($SA^{(4)}$) each slot is longer and contains L minislots, resulting in a data packet length equal to L minislots. $SA^{(4)}$ outperforms $SA^{(3)}$ due to the reduced vulnerable period (L instead of $2L - 1$). The paper also demonstrates that the two non-control channel based protocols are simpler to implement and have superior performance than two previously introduced similar protocols for TT-TR systems [25-26] which are based on a control channel for pretransmission coordination and will be described in the next section. The better performance of $SA^{(4)}$ is due to the longer slot length and the additional data channel, which becomes less beneficial as the number of channels increases. In the analysis of all these protocols the tuning time of transceivers is considered negligible.

A slotted TT-FR^m system which takes into account nodes with limited transmitter tuning capability, different buffer capacities, packet generation rates, and packet destination distributions (asymmetric configuration) was reported in [27]. Since transceivers are chosen such that

$$T_i \cap R_j \neq \emptyset, \quad \forall i, j \text{ and } i \neq j \qquad (2)$$

where T_i denotes the set of wavelengths that the transmitter of node i can tune to, R_j denotes the set of wavelengths that node j can simultaneously receive from and $\emptyset$ denotes the empty set, each node can transmit to any other node in a single hop. Two synchronous channel access protocols were proposed, namely, multichannel slotted ALOHA and random TDMA. In the first protocol, node i transmits—with a certain probability—a buffered packet destined to node j on a channel randomly chosen among $T_i \cap R_j$. In the second protocol, every node, using the same random number generator with the same seed, chooses randomly one wavelength k. Next, node i is chosen randomly among those able to transmit on wavelength *k*. Node *i* chooses randomly from its buffer any of the packets destined to nodes which can receive on channel *k*. This procedure is repeated until all wavelengths are assigned. Thus, transmissions can take place on all wavelengths simultaneously. As expected, in terms of the average packet delay, multichannel slotted ALOHA is superior to random TDMA for low system loads and vice versa for moderate to high system loads. Both protocols achieve the lowest delay and the highest throughput when the contention is reduced. This can be obtained if transmitters are tunable to all wavelengths and the number of receivers per node is equal to the number of wavelengths. This, however, is not economical if the number of wavelengths is large.

An approach to physically prevent channel collisions was presented in [28]. In this TT-FR system a node can transmit only if the addressed channel is not busy and no other node tries to transmit on the same channel at the same time. The Protection-Against-Collision (PAC) optical packet network is a PSC based WDM system where each node is equipped with a PAC circuit located at the central PSC. The PAC circuit determines the state of the respective channel by simple optical power measurement. The resulting signal controls an optical switch connecting the node to the network. If the addressed channel is idle, the PAC circuit closes the corresponding optical switch and blocks other stations. Blocked packets are reflected back to the station and have to be retransmitted. Each packet is preceded by a carrier burst of *n*-bit duration used for probing the state of the addressed channel. This random access protocol is similar to the Carrier Sense Multiple Access (CSMA) protocol. However, the PAC optical packet network does not require a small network propagation delay (relative to the packet transmission time) since all PAC circuits are located near the central PSC.

2.3 Pretransmission coordination protocols

Pretransmission coordination is used to dynamically allocate a channel to a source-destination pair and thereby to prevent channel and receiver collisions. A control packet is sent by the source node on the control channel to inform the target station (and all other stations) of its intention to transmit a data packet. An idle station monitors the control channel and tunes its receiver to the specified data channel during the corresponding time period if a control packet contains its address. Pretransmission coordination protocols are beneficial, especially for TT-TR systems.

2.3.1 Protocols with receiver collision

A control channel based TT-TR system with distributed access protocols was reported in [25]. The control channel is shared by all stations on a contention basis. The length of a data packet is L (integer) times the length of a control packet. Several pairs of protocols X/Y were proposed where X denotes the control channel and Y denotes the data channel access protocol, respectively. Four protocol combinations make use of (slotted or unslotted) ALOHA and (nonpersistent) CSMA, namely, ALOHA/ALOHA, slotted ALOHA/ALOHA, (slotted or unslotted) ALOHA/CSMA and CSMA/(slotted or unslotted) ALOHA. It was shown that the most efficient protocol is CSMA/N-Server switch where N represents the number of data channels [25]. In this protocol idle stations monitor the control channel over one data packet length. In doing so, each station knows the exact status of the channels and the other stations. This approach works without collisions, but stations can be blocked when all channels (servers) are busy.

Those protocols in [25] which are independent of the normalized packet propagation time, i.e., which do not require carrier sensing, were modified in [26]. Especially, for very high-speed optical networks propagation delay independent MAC protocols are needed. The slotted ALOHA/ALOHA protocol was improved as follows. If the control packet is transmitted successfully, then the station transmits the corresponding data packet immediately after the control packet. This saves bandwidth in case of control packet collision. The CSMA/N-Server switch is replaced with a slotted ALOHA/N-Server switch. This approach does not require carrier sensing and yields higher throughput. Both modified protocols exhibit better performance for all values of the system parameters.

Similarly, the *pure* ALOHA/Slotted CSMA access protocol for TT-TR systems allows data transmission over one of the slotted data channels only after a successfully transmitted control packet [29]. Again, no bandwidth is wasted in case of control packet collision. Compared to the original ALOHA/CSMA protocol in [25], under heavy traffic higher throughput and lower average packet delay was achieved for a reasonable range of the system parameters.

In the improved slotted ALOHA/ALOHA protocol, channel collisions can still occur when two or more control packets selecting the same data channel are successfully transmitted within L slots. The slotted ALOHA/*polite* access protocol prevents these channel collisions thereby achieving better performance [30]. If the control packet is successful and if there were no other successful control packets having the same data channel number during the $(L - 1)$ control slots prior to its own control packet transmission, i.e., the selected channel is idle, the station transmits its data packet over the chosen data channel immediately after the control packet is received. Otherwise, the station repeats the same procedure after a random delay.

Due to the nonzero propagation delay nodes using the slotted ALOHA/N-Server switch protocol obtain information which is a (normalized propagation delay) slots old. Consequently, a station *X* may choose the same data channel chosen by another station *Y* whose control packet was transmitted earlier but not yet received by station *X*. The resulting data channel collision is avoided by the so-called slotted ALOHA/*synchronous* protocol which works independently of the network propagation time [30]. The control channel is divided into contiguous frames of *L* control slots. Prior to data packet transmission a station sends a control packet in one of the *L* control slots. The control packet does not include a channel number, i.e., stations do not select data channels. Upon receiving the frame, the data channels are assigned to the stations according to a certain algorithm, e.g., on a First-Come-First-Served (FCFS) basis. Unless there was a control packet collision or no idle data channel available, the station starts transmitting immediately after the channel assignment. This approach provides higher maximum throughput.

Several modifications of the slotted ALOHA/ALOHA protocol [25] and its improved version [26] were reported in [31]. The novel idea is that time is divided into cycles including both control and data packets. In the simplest proposed protocol there is one control slot assigned to each wavelength. The data packets are transmitted on the data channels immediately after the control slot sequence. To use the bandwidth more efficiently, control and data slot sequences overlap in the proposed protocol. Thus, control and data packets are sent in two successive cycles relaxing the need for fast tunable transceivers. Other analyzed protocols differ in the number of control slots per cycle, which are either fixed preassigned to channels or not. The so-called Reservation ALOHA (R-ALOHA) protocol is a candidate for transmitting messages, which are split up into packets. The selected control slot is jammed in every cycle until the end of the session is reached. It was shown that the performance depends on the number of data channels, the length of the data packets and the number of slots on the control channel. In general, the proposed slotted ALOHA and Reservation ALOHA protocols achieve better throughput and delay characteristics than the ones in [25] and [26]. It turned out that an increasing number of control slots has a positive impact on the throughput and a negative on the average packet delay. In accordance with [26], the performance can be improved if data packets are transmitted only after successful control packets. But this holds only for those protocols with fixed preassigned control slots.

Bandwidth can also be utilized more efficiently if the control slot sequence is distributed among several wavelengths [32]. Thus, the actual duration of the control slot sequence is reduced—leading to a significantly improved throughput-delay performance. This *multicontrol channel* approach divides the stations into groups, each with its own control channel. Immediately after the control slot sequence, all wavelengths are used for data packet transmission, i.e., there is no separate control channel. Again, time is divided into cycles containing control and data slots. Note that data channel collisions can occur.

As opposed to [26], receiver collisions were taken into consideration in the slotted ALOHA/*delayed*-ALOHA protocol for a finite population [33]. Again, it is a TT-TR system where a station transmits only after it learns about the successful transmission of its control packet and the transceiver tuning time is assumed to be zero. Apart from the detrimental impact of receiver collisions it was found that an overdimensioned control channel (i.e., the number of data channels is smaller than a certain threshold) causes *bimodal* throughput and *non-monotonic* average packet delay for small values of the backoff delay. The throughput drops due to increasing data channel collisions, then goes up again since there are fewer successfully transmitted control packets, and finally approaches zero because at heavy traffic almost all control packets collide. Note that for a large number of data channels (overdimensioned data channels) the control channel becomes the bottleneck. Hence, control channel based systems have to be dimensioned properly according to the following rule of thumb

$$N = \left\lfloor \frac{2L - 1}{e} \right\rfloor \qquad (3)$$

where N denotes the number of data channels and L stands for the ratio of data to control packet length.

2.3.2 Protocols without receiver collision

A MAC protocol with low mean packet delay and throughput up to 100% was investigated in [34]. It is a CC-FT^2-TR^2-FR system, which prevents channel and receiver collisions completely. However, the station structure is quite complex. The basic idea of this approach is that the backlog information of each station is broadcast via a common control channel. By applying the same scheduling algorithm all *N* stations can transmit data packets in the next slot without any conflict. Every station has two fixed tuned laser sources, one for control and one for data packet transmission. The fixed tuned receiver monitors the control channel. The tuning penalty can be eliminated by alternately using two tunable receivers in pipeline operation. Time is slotted with a slot duration equal to the data packet transmission time. The control slot is partitioned into *N* minislots where the *i*-th minislot is allocated to station *i*. All new packets are announced through the broadcasting of their destination addresses on the control channel. Each station maintains a backlog matrix, a backlog indication matrix and a transmission matrix. The backlog matrix contains the number of backlogged packets in the entire network. The backlog indication matrix is used as input for the so-called Maximum Remaining Sum (MRS) algorithm. The MRS algorithm finds the transmission matrix with the maximum number of nonzero elements, which specify the source destination pairs for the consecutive slot. The algorithm solves a linear integer program (LIP) with the following objective

$$\max \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} \cdot t_{ij}, \qquad (4)$$

where d_{ij} and t_{ij} denote the elements of the backlog indication matrix and the transmission matrix, respectively. The same system was used to reduce the number of required buffers per station and to diminish the mean packet delay in [35]. A central scheduler collects the backlog information of all stations, computes, and finally broadcasts the transmission schedules. The scheduler distinguishes between sequenced and non-sequenced packets. For the latter ones the sequence does not need to be maintained. Congested stations are given permissions by the scheduler to transmit non-sequenced packets to uncongested stations. Particularly for nonuniform traffic this leads to a lower packet loss and a higher network throughput due to load balancing. The scheduler executes the above mentioned MRS algorithm to maximize the aggregate throughput and determines the load balancing from stations with the longest queues to stations with the shortest queues.

A CC-TT-FT-TR-FR contention-based reservation protocol was reported in [36]. Again, this system is based on a PSC. Each node has a transceiver fixed tuned to the control channel and a tunable transceiver for data transmission. All channels are slotted. The control channel is further divided into minislots and one end-to-end propagation delay. Thus, all control packets (which contain only the destination address) arrive at the other nodes in the same slot. A station with a data packet to send, randomly selects a minislot to reserve a data channel (slotted ALOHA on the control channel). The reservation is successful if the control packet (1) does not collide with others, (2) does not have the same destination as those of earlier successful reservations, and (3) the number of currently successful reservations is smaller than the number of data channels. If the reservation is successful and the number of successful reservations before this one is $(i - 1)$, the transmitter and the addressed receiver tune to the wavelength λ_i and the data packet is transmitted in the next slot. If the reservation fails, the station attempts to reserve in the next slot, repeating the same procedure. This is an example for an *attempt-and-defer* protocol. Note, that especially in large high-speed networks bandwidth is wasted during the propagation delay portion on the control channel. The protocol is made more efficient by replacing the propagation delay part with additional minislots [37]. Thus, stations can transmit control packets during the entire slot. After a round-trip propagation delay each node knows whether the corresponding control packet was successful or not. To reduce the number of control packet retransmissions, all control packets which have not experienced channel collisions but have found the destination and/or all channels busy are put in a queue. As soon as the required resources are released the corresponding data packet is transmitted. This scheme can support data packets of variable size. The packet size is announced in an additional control packet field. However, each station has to maintain status information about channel and receiver utilization. Processing requirements can be reduced if stations do not maintain any status tables [38]. Suppose, after a successful control packet a station is assigned channel λ_i for data transmission. In the subsequent slot, control minislot *i* can be used explicitly by this station to reserve another data slot if the message contains more than one data packet. The reserved minislot is released after transmitting the entire message. Thus, multiple control packet transmissions replace the need for status informations at each node. The basic approach in [36] can be extended to incorporate multipriority traffic, e.g., real time and nonreal time traffic [39]. Each control packet contains the destination address as well as the priority of the data packet. Channels are assigned sequentially from the highest priority to the lowest priority successful control packets. Among the control packets of the same priority, the left most packet in a slot is selected first and the right most packet is selected last. If a control packet has the same destination address as a previously selected control packet it is discarded to prevent destination conflicts.

A PSC based CC-TT-TR protocol which can detect and avoid receiver collisions was discussed in [40]. This so-called Receiver Collision Avoidance (RCA) protocol keeps the station structure at a low cost and is scalable since the control channel applies the slotted ALOHA access scheme. Moreover, nonzero tuning time and nonzero propagation delay are taken into account. Since a node cannot monitor the control channel at all times, it is impossible to have total knowledge of the status of the other nodes. Nevertheless, the proposed protocol is able to guarantee data packet transmission without receiver collision. The basic idea is that a data packet is sent (after a tuning time) on the assigned data channel immediately after the control packet transmission is found successful. Here, successful means that the control packet is received without channel and receiver collision. All channels are partitioned into slots equal to the data packet transmission time and are further subdivided into control slots each with the length of a control packet. On the control channel slot *i* is used for reservation of wavelength λ_i. The control packets contain only the destination address. Each node maintains two data structures. The so-called Reception Scheduling Queue (RSQ) is a list of entries containing reception time and assigned data channel and is used to schedule data packet receptions. The so-called Node Activity List (NAL) is used to record information received on the control channel. By using these two data structures properly each node can decide whether it can transmit a control packet or not. While waiting for the return of the transmitted control packet the node monitors the control channel to find out if there are other control packets destined to the same destination. If no other competing control packets arrive at the node prior to the own successfully transmitted control packet, the corresponding data packet is sent in the next control slot. Otherwise, the node aborts and restarts the procedure until it succeeds. Thus, receiver collisions can be detected and avoided. Note that a data packet can be transmitted at the beginning of any control slot instead of at the beginning of each data slot. To further increase bandwidth efficiency one node's tuning time can be overlapped with another node's transmission time if both nodes transmit on the same channel. The maximum throughput efficiency is about 36%. A limitation of the RCA protocol is that the packet delay can fluctuate over a relatively wide range under heavy load, which makes it unsuitable for real time traffic. The Extended RCA (E-RCA) protocol incorporates nonuniform node distances from the PSC [41]. Simulations under various distance distributions show that the E-RCA protocol perform almost as well as the RCA protocol under the same average distance.

A minor variation of the RCA scheme is the MultiS-Net [42]. In contrast to the RCA approach, the protocol in MultiS-Net is synchronous in the sense that data packets can only be transmitted at the beginning of a data slot. Node i randomly selects a control slot and transmits a control packet with the destination address of j in slot $(t + 1)$ if (1) in slot t, node i generates a packet destined to node j, and (2) there are no other control packets destined to i or j. These two conditions guarantee that i and j monitor the control channel in slot $(t + 1 + R)$, where R denotes the constant round-trip delay between a station and the PSC. In slot $(t + 1 + R)$ the control packet returns; if the control packet is successful, station i transmits the data packet in the subsequent data slot. A control packet is successful if (1) no other station has used the same control slot, (2) no prior successful control packet is destined to the same destination, and (3) the number of prior successful control packets is smaller than the number of data channels.

The Receiver Conflict Avoidance Learning Algorithm (RCALA) uses learning automata in order to reduce the number of receiver collisions [43]. It is a CC-TT-FT-TR-FR system based on a PSC. Each node is equipped with a learning automaton; the learning automaton decides which of the packets waiting for transmission will be transmitted in the next time slot. Each source node selects a packet according to a destination distribution $P_j(t)$ (probability that a packet destined for node j is selected at time t) which is updated continuously. All packets are preannounced on the control channel before transmission. Thus, every station receives a network feedback information after the round-trip delay D. This feedback information is utilized to update the destination distribution $P_j(t + 1)$ as follows:

$$P_j(t + 1) = P_j(t) - L \cdot P_j(t), \quad L \in (0; 1), \qquad (5)$$

if a receiver collision occurred at destination j during the last D slots and

$$P_j(t + 1) = P_j(t) + \epsilon \cdot L \cdot [1 - P_j(t)], 0 < \epsilon \ll 1, \quad (6)$$

if no receiver collision occurred. The parameter L must be appropriately selected in order to maximize the network performance. The choice of L is a trade-off between adaptation speed and accuracy. The RCALA protocol can be considered an extension of the random transmission strategy which takes into account the network feedback information. This adaptive transmission strategy shows better throughput-delay performance than the FIFO and random transmission strategies. The RCALA protocol operates more efficiently in networks with low propagation delay and relatively high correlation of packet destinations. A similar PSC based TT-FR system was introduced in [44]. On all channels slotted ALOHA is deployed. Each station monitors the status of all channels by using an array of learning automata, which are able to determine whether a channel is idle or a packet transmission was successful or not. Each station sends a packet with a variable transmission probability. This probability is increased if in the previous slot there was either a successful or no packet transmission at all. It is decreased in case of collisions. Thus, this protocol achieves high throughput and low delay under any load conditions.

Two reservation protocols with varying signaling complexity, which avoid both channel and receiver collisions were reported in [45]. They are for a PSC based CC-FT^2-TR-FR system. Each station has one transceiver, which is fixed tuned to the common control channel, one fixed tuned transmitter for data transmission, and one tunable receiver for data reception. The first proposed protocol is called Dynamic Allocation Scheme (DAS). Every node executes an identical arbitration algorithm by using a common random number generator with the same seed. Thus, all transmitters arrive at the same final conclusion. A transmitter i is randomly selected among all the transmitters. Among all the nonempty receiver queues (one for each destination) at transmitter i, one queue r is randomly chosen. In the upcoming slot, transmitter i sends a packet to receiver r. If all receiver queues are empty the slot remains unused. On the control channel, the receiver queue status of all stations is permanently broadcast using a fixed TDM scheme. Thus, each station obtains all relevant informations for executing the arbitration algorithm. The algorithm procedure is repeated every data slot until all transmitters are served. At any step, transmitters and receivers that have already been scheduled are excluded from the arbitration algorithm. Hence, fairness is provided while higher priorities to the queues with larger arrival rates are implicitly given by selecting only nonempty receiver queues.

The other protocol is called Hybrid TDM (HTDM) scheme which reduces the signaling overhead of the DAS protocol. The HTDM is a combination of TDM and DAS, i.e., it supports both preassigned and dynamic slot assignment. One part of the slots is fixed assigned and the remaining slots are dynamically allocated by using the DAS protocol only for these slots. Thus, receiver queue status information have to be broadcast only for these slots. The HTDM protocol can be considered a trade-off between flexibility and signaling overhead. For bursty and nonuniform traffic, HTDM and particularly DAS are superior to the conventional TDM scheme.

The TDMA/C-server protocol is based on a CC-TT-TR-FR system and prevents both channel and receiver collisions [46]. On the control channel each node is assigned one control slot in a static cyclic fashion. Control packets are composed of four fields, namely source address, destination address, data channel number, and packet size. The protocol supports packets of variable length *L.* Each station maintains two status tables. The channel status table keeps track of the status of the data channels and is used to avoid channel collisions. The node status table prevents receiver collisions by keeping track of the status of the tunable receiver at each node. The table entries indicate the number of slots during which resources are busy. The tables are updated after each control slot by utilizing the receiver that is fixed tuned to the control channel. The variable α denotes the maximum of the time required by the transmitter of the source node to tune to the selected channel and the time required by the destination node to receive and decode the control packet and switch the tunable receiver to the chosen wavelength. If node j transmits a control packet preannouncing a data packet to node i on the wavelength k, then all nodes add $(L + \alpha)$ to entry k in the channel status table and to entry i in the node status table. All positive entries are decremented at the end of each control slot. A station sends a control packet if it is backlogged and the destination and at least one data channel are idle. A destination and a data channel are considered idle if the corresponding entries equal zero and are less or equal to α, respectively.

A reservation based multichannel access protocol without control channel was proposed in [47]. This protocol is reasonable for single-hop networks with a relatively small number of wavelengths. The transmission is organized in cycles. Each cycle is composed of a control phase and a data phase. Time is slotted on the control packet boundaries. A control packet contains the destination address, the data channel number, and the data packet size, which can be variable. During the control phase each node is assigned one control slot per wavelength in a fixed TDM scheme preventing collisions. Each node may reserve access for exactly one channel in the data phase. Every node has a laser array and a fixed tuned receiver operating on an individual home channel. During the control phase the laser array is fully activated to broadcast the control information to all other nodes. The corresponding data packet is sent during the subsequent data phase. Nodes that reserved access on the same channel transmit data in the order in which they transmitted their control packets (distributed arbitration algorithm). This protocol combines the advantages of preallocation and reservation access strategies. It provides lower latencies under light loads and works stable under heavy loads, unlike random access schemes. It trades off maximum capacity and latency reduction. By allowing a node to reserve access for more than one channel during the data phase the performance can be improved [48].

A time/wavelength division multiple access protocol using acoustooptic receivers was reported in [49]. Their relatively long tuning time can be offset by subframe tuning. Each frame is divided into subframes, each containing several slots. During a subframe one receiver is receiving while the other is retuning for the next subframe (pipelining). The minimum subframe length is determined by the receiver tuning time. Due to the larger tuning range of acoustooptic receivers more wavelengths are available. Flexibility is increased if the acoustooptic receivers are tuned to more than one wavelength during a subframe [50].

An interesting approach to deploy fast tunable transmitters and yet to support a large total number of wavelengths was investigated in [51]. Each node has a transmitter that can send on one of several different wavelength groups. Each wavelength group encompasses a limited number of consecutive wavelengths. A wavelength group may be shared by more than one transmitter. In addition, every station has a fixed tuned receiver, each receiving from a separate set of interleaved wavelengths. Every set contains one wavelength from each wavelength group. The receiver has a periodic multiwavelength-pass receiver filter similar to a Fabry Perot interferometer, which receives on one wavelength at a time. Channels are assigned to source destination pairs in a fixed, contention-free manner. Full connectivity is provided by tuning the transmitters and periodically selecting a different pass wavelength at the receiver.

Several access protocols that avoid channel and receiver collisions and incorporate transceiver tuning time and propagation delay were investigated in [52]. These protocols are able to schedule variable-size messages with a single control packet transmission resulting in a reduced signaling overhead. It is a PSC based CC-TT-FT-TR-FR system where one transceiver is fixed tuned to a common control channel and another tunable transceiver is responsible for data transmission/reception. Data channels are divided into fixed data slots whose length equals the data packet transmission time. Variable-size messages may require one or more data slots for transmission. The control channel is divided into frames that are composed of control slots. The control slots are fixed assigned to the stations and do not necessarily have to be synchronized to the data slot boundaries. Thus, the network is easily scalable by adding further control slots to each frame. However, the network has to be reinitialized each time a node is added or removed. Each control packet contains the destination address, the length of the corresponding message (number of data slots) and in one of the proposed protocols also the channel number. Since all stations continuously monitor the control channel and execute the same access arbitration algorithm no channel and receiver collisions occur. Each station keeps track of the usage of the channels and receivers. Each station records the channel that each receiver is tuned to after receiving its last message. The proposed arbitration algorithms try to reduce the impact of tuning overhead by avoiding unnecessary receiver tuning and the message delay by using the above mentioned status informations.

A hybrid multiaccess scheme with wavelength and code concurrency (WCDMA) was reported in [53]. Several stations share a wavelength through code multiplexing. Thus, the number of required channels is reduced resulting in lower packet delays since tunable transceivers with smaller tuning latencies can be applied. The inherent limitation on the number of stations that can be supported by optical Code Division Multiple Access (CDMA) is eliminated because code words can be reused on different wavelengths. It is a CC-FT^2-TR-FR system where a fixed transceiver is deployed for monitoring the control channel, the fixed tuned transmitter and the tunable receiver are responsible for data transmission. The control channel is divided into frames, which are further subdivided into *N* slots, for each station. In slot *i,* station *i* broadcasts the addresses of the receivers to whom it has packets to transmit as well as the status of its transmitter and receiver. The transmitter (receiver) status indicates the receiver (transmitter) address which the transmitter (receiver) is currently transmitting to (receiving from), if any. The receiver cyclically scans the data channels. Upon tuning to channel f_k receiver R_j sequentially checks from the control channel whether a transmitter on f_k has packets for it, and whether that transmitter is available. If there is an available transmitter on f_k, R_j indicates its readiness in the next control frame and starts the reception phase.

3 Resulting Guidelines

From the above discussion the following conclusions for the design of single-hop WDM networks and MAC protocols can be drawn:

- Fixed time/wavelength assignment avoids channel and receiver collisions but is suitable mainly for uniform and nonbursty traffic at medium to high system loads. In addition, nodes must be synchronized and the system is not scalable due to the fixed slot assignment. At low to medium loads a partially fixed time/wavelength assignment leads to a smaller mean packet delay [20].
- The network is made scalable by deploying a random access control channel. In very high-speed networks slotted ALOHA outperforms CSMA since it is independent from the (normalized) propagation delay.
- Higher transmission concurrency increases the efficiency [22]. This can be achieved for example by means of wavelength reuse [23] or by using the multiwavelength selectivity of acoustooptic receivers [50]. A larger transmitter tuning range and more receivers at each station reduce the contention in random access WDM networks, resulting in an improved throughput-delay performance [27].
- The transceiver tuning penalty can be mitigated by minimizing the number of tuning operations [21], by overlapping data transmission and tuning operation [31], or by using multiple FSRs of a fixed tuned Fabry Perot receiver where each transmitter is assigned a separate receiver FSR, thereby requiring only a small transmitter tuning range [51]. The number of required wavelengths can be reduced by means of channel sharing, e.g., CDMA (other options would be Subcarrier Multiplexing (SCM) or Polarization Division Multiplexing (PDM)) [53].
- For single-hop WDM networks with tunable transmitters and tunable receivers at each node pretransmission coordination via a control channel is reasonable [25]. However, it does not necessarily yield better performance [24]. The throughput-delay can be improved if more than one channel are used for control packet transmission [32]. Misdimensioned control channel based systems can exhibit bimodal throughput and nonmonotonic average packet delay [33]. To avoid data channel and receiver collisions, a node does not have to monitor the control channel continuously, only the last few slots are relevant [40-42].
- Adaptive scheduling algorithms convey better throughput-delay characteristics, especially under varying traffic conditions [42,43]. A common distributed access arbitration algorithm reduces the signaling overhead and avoids collisions. A hybrid fixed and dynamic slot assignment is a good tradeoff between flexibility and signaling overhead [45]. Implicit wavelength allocation reduces the pretransmission coordination overhead [36]. Control slot preallocation and dynamic data slot assignment guarantee latency reduction at low loads and stable operation at high loads [47,48].
- Bandwidth can be saved and performance can be improved if data packets are sent only after successfully transmitted control packets [26,29].
- Variable-size data packets can be announced by using a corresponding control packet field [46]. Variable-size messages without requiring status tables at each node can be supported by jamming the respective control channel slots until the message is completely sent [38]. Supporting variable packet lengths reduces the signaling overhead [52]. Additional control packet fields can

be used to support multipriority (real and nonreal time) traffic [39].

- Receiver collisions are considered more destructive than channel collisions [17]. In terms of throughput, channel collision avoidance is superior to retransmissions [30].
- Network propagation time independent protocols reduce the channel collision probability and thereby exhibit a better throughput [30]. In carrier sensing networks the throughput-delay performance can be increased by centralizing the sensing entities [28].
- Buffer sharing among stations reduces the number of required buffers and the mean packet delay [35].
- Concluding, for single-hop WDM networks in which each wavelength is shared by multiple nodes, TT-TR systems without fixed allocated *home channels* for reception provide a better throughput-delay performance than TT-FR systems with home channels [54]. This is due to the fact that a tunable receiver can receive data from any free wavelength as opposed to a fixed receiver that cannot receive data if its home channel is already busy, even though other wavelengths are not used.

4 Architecture

Based on the guidelines discussed in the previous section, we now outline the architecture of a future-proof metro WDM network. The architecture is based on an Arrayed Waveguide Grating (AWG). The AWG is a wavelength sensitive device. It allows for a large degree of concurrency, primarily through the spatial reuse of wavelengths.

4.1 Underlying principles

In this section we briefly review the salient features of the AWG. Without loss of generality we consider a 2×2 AWG. In Fig. 3 a) six wavelengths are launched into the upper AWG input port. Every second wavelength is routed to the same AWG output port. This period of the wavelength response is called Free Spectral Range (FSR). In general, the FSR of a $D \times D$ AWG, $D \geq 2$, contains D wavelengths. Due to the routing characteristics of an AWG a transmitter attached to the upper AWG input port has to be *tunable* over at least one FSR in order to provide full connectivity. A tunable transmitter with a larger tuning range could access more than one FSR of the underlying AWG, each providing an additional channel for communication between any AWG input/output port pair.

As shown in Fig. 3 b) all wavelengths can be used simultaneously at both AWG input ports without causing collisions at the AWG output ports. Besides using multiple FSRs, spatial wavelength reuse is another possibility to increase the degree of concurrency. In fact, spatial wavelength reuse is the main difference between the wavelength-selective AWG and the wavelength-insensitive PSC. Spatial wavelength reuse increases the degree of concurrency leading to a significantly improved network throughput-delay performance [55]. Note from Fig. 3 b) that a receiver attached to one of the AWG output ports has to be *tunable* over at least one FSR in order to guarantee full connectivity. Thus, AWG based single-hop WDM networks with full connectivity require a *TT-TR node structure.* (Alternatively, each tunable transmitter (receiver) could be replaced with an array of fixed tuned transmitters (receivers). However, a tunable transceiver is superior to an array of fixed transceivers in terms of management, operational costs, and complexity.)

Now, we can benefit from the guidelines presented in Section 3. As we have seen, for TT-TR systems pretransmission coordination is reasonable. To conserve bandwidth and to increase the network efficiency data packets should be sent only after successfully transmitted control packets. In addition, to avoid explicit reservation acknowledgements and thereby to reduce both signaling overhead and response time, and to prevent channel and receiver collisions of data packets, each node should execute the same distributed scheduling algorithm. However, this implies that each node needs global knowledge about the state of network resources and all new reservation requests. This can be achieved by broadcasting each control packet. But the AWG as a wavelength-selective device does not support broadcasting. One solution could be to deploy an additional PSC with one attached transceiver per station for broadcasting control packets. Instead, we follow a more elegant and cost-effective ap-

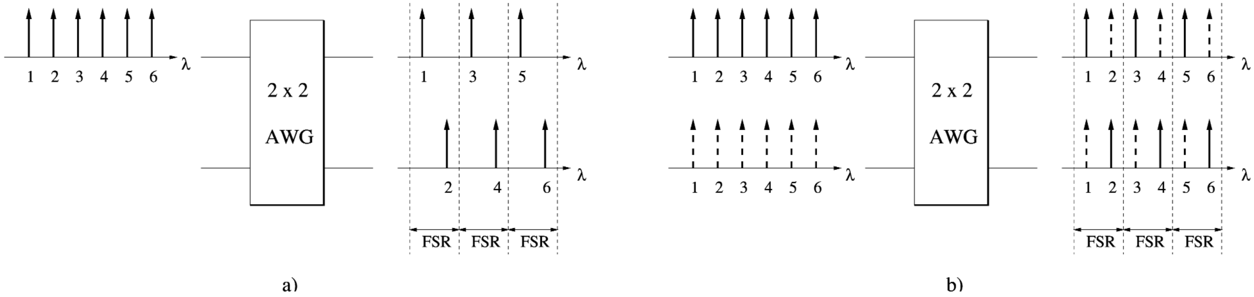

Figure 3: Properties of an AWG: a) Periodic wavelength routing, b) spatial wavelength reuse.

proach as depicted in Fig. 4. Broadcasting can be realized by using a broadband light source, e.g., a Light Emitting Diode (LED), which is spectrally sliced by the AWG, such that slices of the original broadband signal are routed to every AWG output port. Since each slice carries the same (control) information a receiver attached to one of the AWG output ports can be tuned to any slice in order to retrieve the (control) information. As shown in Fig. 4, applying broadband light sources at both AWG input ports simultaneously does not result in *channel collisions* at the AWG output ports. However, a station equipped with a single tunable receiver can monitor only one slice originating either from the upper or lower AWG input port, while the slices coming from the other AWG input port suffer from *receiver collision.* Consequently, if all nodes are required to maintain global knowledge by continuously monitoring all incoming control packets (slices), control packets can be transmitted only from one AWG input port at any given time. Note that the slices and the wavelengths overlap spectrally, as shown in Fig. 4. Thus, using LEDs and transmitters concurrently would lead to collisions. This problem will be addressed shortly.

4.2 Network and node architecture

The network and node architecture is depicted in Fig. 5. The network is based on a $D \times D$ AWG, $D \geq 2$. To increase the network efficiency, all wavelengths are intended to be used at all AWG input ports simultaneously. This requires additional ports for attaching multiple transmitters (receivers) at each AWG input (output) port. As shown in Fig. 5, this can be achieved by attaching a $S \times 1$ combiner at each AWG input port and a $1 \times S$ splitter at each AWG output port ($S \geq 1$). Each node is composed of a transmitting part and a receiving part. The transmitting part of a node is attached to one of the combiner ports. The receiving part of the same node is located at the opposite splitter port. Each node contains a tunable Laser Diode (LD) for data transmission and a tunable Photodiode (PD) for data reception. In addition, each node deploys an LED for broadcasting control packets by means of spectral slicing. No additional receiver is needed if the signaling is done in-band. To increase the degree of concurrency, data and control information are transmitted simultaneously. However, the receiver must be able to distinguish data and control information. This

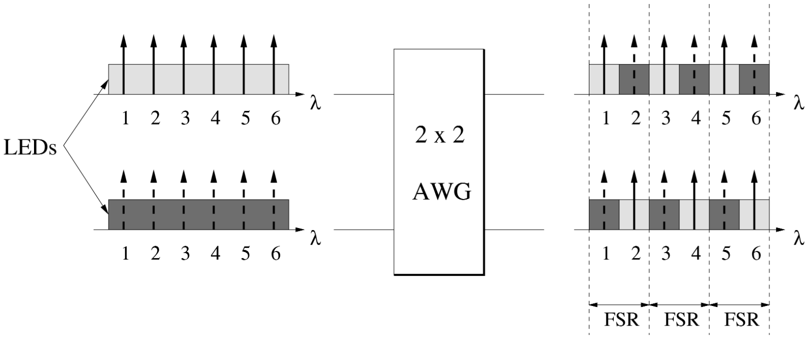

Figure 4: Spectral slicing of broadband Light Emitting Diode (LED) signals.

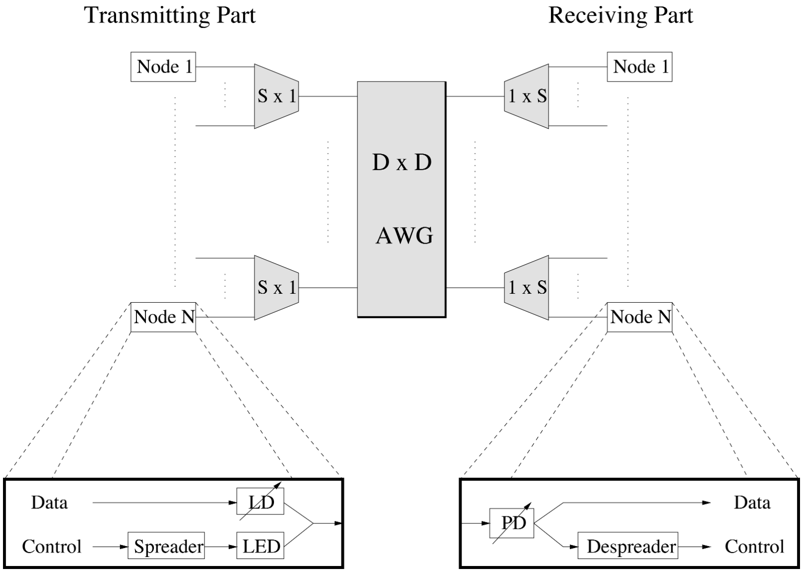

Figure 5: Network architecture.

can be achieved by means of direct-sequence spreading techniques [56]. The control information is spreaded before modulating the LED. Accordingly, at the receiving part the control information is retrieved by despreading a part of the incoming signal. Thus, the network efficiency is significantly improved by transmitting control and data simultaneously and using all wavelengths for data transmission, no additional control channel (wavelength) is needed. (For a more detailed discussion of the physical limitations of this approach the interested reader is referred to [57].)

The combiners have to be wavelength-insensitive since each attached transmitter has to be tunable in order to provide full connectivity. But no matter which wavelength the transmitter is tuned to, all transmitted data packets have to be carried on the fiber between the corresponding combiner output port and the corresponding AWG input port. This requires that the combiners are wavelength-insensitive. Similarly, each splitter has to be wavelength-insensitive. To see this, recall that each receiver has to be tunable in order to guarantee full connectivity. In other words, a given destination node receives data packets from different AWG input ports on different wavelengths. As a consequence, the splitter has to make sure that all wavelengths are distributed to all receivers that are attached to the splitter output ports. This requires that the splitters are wavelength-insensitive. Note that similar to the combiners, the splitters introduce splitting loss. However, the splitters allow for *optical multicasting* which is one of the key features of future metro WDM networks. Other requirements of metro networks are also met. By using only passive components (combiners, splitters, AWG) the network is reliable and cost-effective. Since all active devices are located at the network periphery they can be replaced in case of failure or upgraded in an easy and non-disruptive manner. Moreover, the node structure consisting of one single transceiver and one low-cost LED is simple and economic.

5 MAC Protocol

The wavelength assignment is schematically shown in Fig. 6. The y-axis denotes the wavelengths used for transmission and reception. As illustrated, R adjacent FSRs are exploited. 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 receiver during simultaneous transmissions in different FSRs of the AWG, the FSR of the receivers has to be different 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. 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

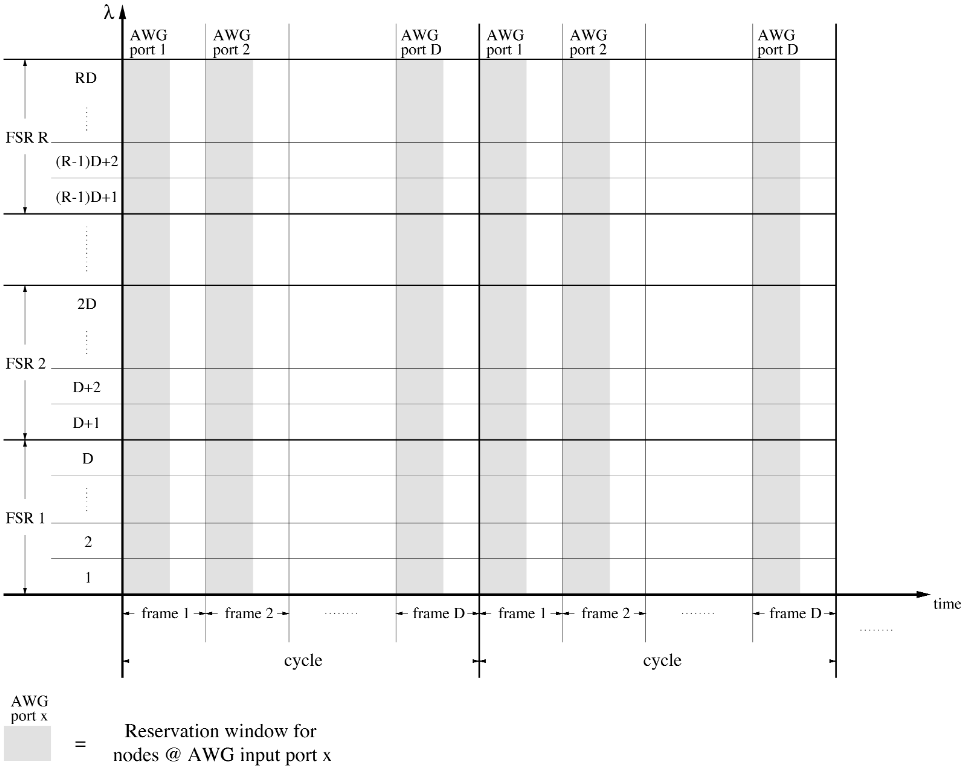

Figure 6: Wavelength assignment.

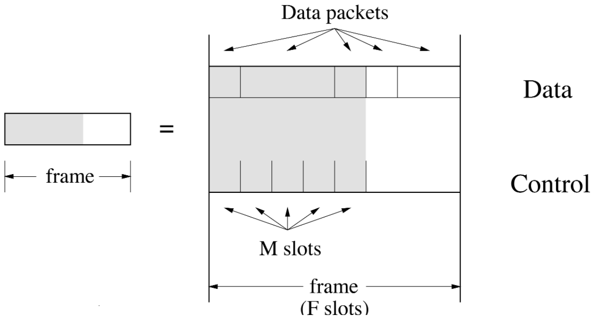

Figure 7: Frame format.

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 large enough to guarantee spatial wavelength reuse. The spatial wavelength reuse 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 exhibit a negligible tuning time of a few nanoseconds. Each frame is partitioned into the first M, $1 \leq M < F$, slots (shaded region) and the remaining $(F - M)$ slots. In the first M slots the pretransmission coordination takes place. During this period control packets are transmitted and all nodes are obliged to tune their receivers to one of the corresponding LED slices (channels) for obtaining the control information in order to acquire global knowledge. Owing to the aforementioned wavelength routing properties of the AWG, in a given frame only nodes that are attached to the same corresponding combiner can transmit control packets. Nodes attached to AWG input port i (via a common combiner) send their control packets in frame i, $1 \leq i \leq D$ (see Fig. 6). Each frame within a cycle accommodates control packets originating from a different AWG input port. Hence, after D frames (one cycle) all nodes have had the opportunity to send their control packets guaranteeing fairness. The M slots are not fixed assigned. Instead, control packets are sent on a contention basis using a slightly modified version of slotted ALOHA. This makes the entire network scalable. Control packets arrive at the receivers after a propagation delay that is equal to half the round-trip time. 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 distributed scheduling algorithm. The parameter M trades off two kinds of concurrency. During the first M slots of each frame, (spreaded) control and data packets can be transmitted simultaneously, but only from nodes that 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 transmitters at each node. 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 three fields, namely destination address, length of the corresponding data packet, and a type field. 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.

Let us now take a look at the tunable receiver at each node. 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. In frame k, $1 \leq k \leq D$, each receiver collects the control packets originating from 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 retransmits the control packet in this next cycle with probability p, and so forth. Successful control packets are put in a distributed queue.

All nodes process the successful control packets by executing the same arbitration (scheduling) algorithm. Consequently, all nodes come to the same transmission and reception schedule. 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. For now, we apply a straightforward greedy algorithm, which schedules the control packets on a first-come-first-served and first-fit basis. 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. Note that a control packet does not have to have a field which indicates the wavelength on which the corresponding data packet is intended to be transmitted.

Instead, wavelengths are dynamically allocated in a distributed fashion resulting in a reduced signaling overhead.

The length of the scheduling window is equal to D frames for two reasons. First, by limiting the scheduling window, length to a small number of frames the computational requirements at each node are kept small as well. Secondly, every D frames all nodes receive control packets from the same set of nodes. At the same time, due to the wavelength routing properties of the AWG and the requirement that all nodes listen to the LED slices, only this set of nodes can transmit data packets. Those data packets were announced by control packets exactly D frames earlier. By making the scheduling window D frames long, data and control packets can be sent simultaneously. This kind of concurrency leads to an improved throughput-delay performance.

Next, we discuss the support of 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. Circuit switching is realized by using the type and length fields of the control packet. The length field gives the required number of slots per cycle. By setting the bit in the type field to 1, the source node indicates that the given number of slots must be reserved in each cycle. After receiving the control packet the circuit is set up by choosing the first possible free slots at the lowest available wavelength. Those slots 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 set up a circuit, i.e., node i is granted a certain number of slots per cycle which was specified in the foregoing control packet. Furthermore, suppose j other nodes attached to the same combiner hold currently circuits, where $1 \leq j \leq M - 1$. 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. Moreover, nodes do not have to maintain and update data structures containing the lifetimes of all currently active circuits, resulting in a reduced nodal complexity. 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, randomly chooses one of the slots $j + 1, j + 2, \ldots, M$. Finally, we have assumed that control packets can be corrupted only due to channel collisions with other control packets transmitted in the same reservation slot. Transmission errors are considered negligible. This assumption is reasonable since the transmission path is all-optical and passive without any intermediate active or switching network elements. A more conservative approach could add a Forward Error Correction (FEC) field to each control packet which enables each receiver to compensate for single transmission bit errors.

6 Results

In this section we present some selected results. To save space we mention the assumptions made in the analysis only briefly. For additional results and a detailed discussion of the system model and the analysis, the interested reader is referred to [57]. We consider *fixed-size* data packets with a length of F slots and *uniform unicast* traffic. Each node has a *single-packet buffer*, and *nonpersistent* control packets, i.e., each collided control packet selects a new destination node before being retransmitted. The parameters are set to the following default values: $D = 2$, $M = 8$, $N = 240$, $R = 3$, and $S = 120$. All curves are obtained for varying the mean arrival rate of new data packets to each node from 0 to 1. The throughput is defined as the mean number of transmitting nodes. The mean delay is measured from the time the data packet arrives at a node until the end of the frame during which it is transmitted.

The results in Fig. 8 clearly demonstrate the benefit of using multiple FSRs of an AWG. Each additional FSR increases the degree of concurrency and thereby significantly improves the throughput-delay performance of the network. Using three FSRs instead of one roughly doubles the maximum throughput. However, using multiple FSRs requires transceivers with a larger tuning range.

Fig. 9 shows that the number of reservation slots has a strong impact on the system performance. By using

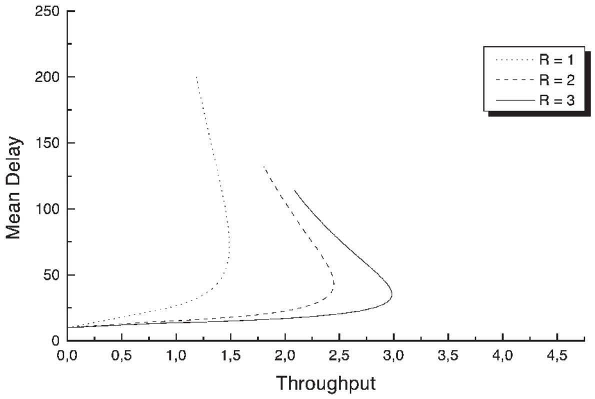

Figure 8: Mean packet delay vs. throughput for different numbers of FSRs $R \in \{1, 2, 3\}$.

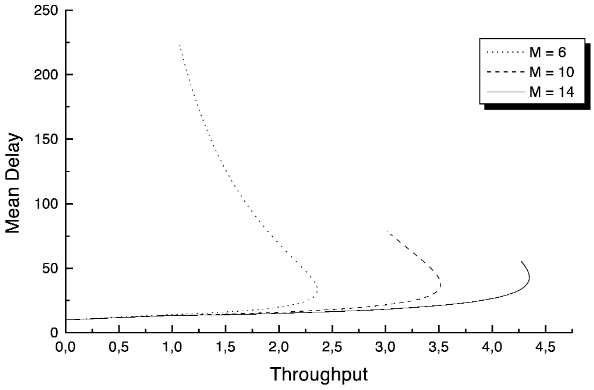

Figure 9: Mean packet delay vs. throughput for different numbers of reservation slots $M \in \{6, 10, 14\}$.

many reservation slots the collision probability in each slot is reduced and the number of successful control packets is increased resulting in an improved throughput-delay performance. We observe that for $M = 14$ all three FSRs are used for data transmission. (Note that using more reservation slots for a given frame size has an impact on spatial wavelength reuse since the white region of each frame (see Fig. 7) decreases. This aspect becomes important for variable-size data packets but not in our case where we consider data packets with a packet length equal to F slots.)

7 Conclusions

We have presented a novel metro WDM network architecture and a MAC protocol. The degree of concurrency is dramatically increased by using multiple FSRs of the underlying AWG, spatially reusing all wavelengths, and simultaneously transmitting data and control packets by means of spreading techniques. All wavelengths are used for data transmission and no additional control channel is required. The network consisting only of passive components is reliable and meets all requirements of future and future-proof metro networks. Due to its inherent transparency the single-hop network provides a large flexibility to a wide range of (legacy) protocols. In addition, the distributed scheduling algorithm avoids explicit acknowledgements and allows data packets to have variable size. Compared to previously reported WDM networks, the node structure with one single tunable transceiver and one low-cost LED is rather economic. Since all active devices are located at the network periphery the network can be maintained and upgraded in an easy and non-disruptive way. The network is scalable by using a random access scheme for reservation. Multicasting is also supported. Circuit switching provides QoS to mission-critical and delay/jitter-sensitive traffic.

8 References

[1] P. Green. Progress in optical networking. *IEEE Commun. Mag.,* 39(1):54-61, January 2001.

[2] C. Awduche and Y. Rekhter. Multiprotocol lambda switching: Combining MPLS traffic engineering control with optical crossconnects. *IEEE Commun. Mag.,* 39(3):111-116, March 2001.

[3] A. Fumagalli and L. Valcarenghi. IP restoration vs. WDM protection: Is there an optimal choice? *IEEE Network,* 14(6):34-41, Nov./Dec. 2000.

[4] R. Doverspike and J. Yates. Challenges for MPLS in optical network restoration. *IEEE Commun. Mag.,* 39(2):89-96, February 2001.

[5] J. Y. Wei, C. Liu, K. H. Liu, J. L. Pastor, and *et al.* IP over WDM network traffic engineering demonstration and experimentation. In *OFC 2001, post-deadline paper PD33,* Anaheim, CA, March 2001.

[6] S. Verma, H. Chaskar, and R. Ravikanth. Optical Burst Switching: A viable solution for Terabit IP backbone. *IEEE Network,* 14(6):48-53, Nov./Dec. 2000.

[7] M. Yoo, C. Qiao, and S. Dixit. QoS performance of Optical Burst Switching in IP-over-WDM networks. *IEEE J. on Sel. Areas in Commun.,* 18(10): 2062-2071, October 2000.

[8] M. Yoo, C. Qiao, and S. Dixit. Optical Burst Switching for service differentiation in the next-generation optical internet. *IEEE Commun. Mag.,* 39(2):98-104, Feb. 2001.

[9] A. Jourdan, D. Chiaroni, E. Dotaro, G. J. Eilenberger, and *et al.* The perspective of optical packet switching in IP-dominant backbone and metropolitan networks. *IEEE Commun. Mag.,* 39(3):136-141, March 2001.

[10] M. J. O'Mahony, D. Simeonidou, D. K. Hunter, and A. Tzanakaki. The application of optical packet switching in future communication networks. *IEEE Commun. Mag.,* 39(3):128-135, March 2001.

[11] S. Yao, S. J. B. Yoo, and B. Mukherjee. All-optical packet switching for metropolitan area networks: Opportunities and challenges. *IEEE Commun. Mag.,* 39(3):142-148, March 2001.

[12] D. Stoll, P. Leisching, H. Bock, and A. Richter. Metropolitan DWDM: A dynamically configurable ring for the KomNet field trial in Berlin. *IEEE Commun. Mag.,* 39(2):106-113, February 2001.

[13] K. V. Shrikhande, I. M. White, D. Wonglumsom, S. M. Gemelos, and *et al.* HORNET: A packet-over-WDM multiple access metropolitan area ring network. *IEEE J. on Sel. Areas in Commun.,* 18(10): 2004-2016, October 2000.

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

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

[16] 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 *OFC 2001. Technical Digest, paper ThG5,* Anaheim, CA, March 2001.

[17] A. Borella, G. Cancellieri, and F. Chiaraluce. *Wavelength Division Multiple Access Optical Networks.* Artech House, 1998.

[18] K. M. Sivalingam and S. Subramaniam, editors. *Optical WDM Networks—Principles and Practice, Chapter 7.* Kluwer Academic Publishers, 2000.

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

[20] I. Chlamtac and A. Ganz. Channel allocation protocols in frequency-time controlled high-speed networks. *IEEE Trans. on Commun.,* 36(4):430-440, April 1988.

[21] A. Ganz and Y. Gao. A time-wavelength assignment algorithm for a WDM star network. In *IEEE INFOCOM '92,* pages 2144-2150, Florence, Italy, May 1992.

[22] 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.

[23] A. Ganz and Y. Gao. Traffic scheduling in multiple WDM star systems. In *IEEE ICC '92,* pages 1468-1472, Chicago, IL, 1992.

[24] 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.

[25] 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.

[26] 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.

[27] A. Ganz and Z. Koren. WDM passive star—protocols and performance analysis. In *IEEE INFOCOM '91,* pages 991-1000, Bal Harbor, FL, April 1991.

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

[29] H. Shi and M. Kahverad. ALOHA/Slotted-CSMA protocol for a very high-speed optical fiber local area network using passive star topology. In *IEEE INFOCOM '91,* pages 1510-1515, Bal Harbor, FL, April 1991.

[30] H. W. Lee. Protocols for multichannel optical fibre LAN using passive star topology. *Electronics Letters,* 27(17):1506-1507, August 1991.

[31] 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.

[32] G. N. M. Sudhakar, M. Kavehrad, and N. D. Georganas. Multi-control channel very high-speed optical fiber local area networks and their interconnection using a passive star topology. In *IEEE GLOBECOM '91,* pages 624-628, Phoenix, AZ, December 1991.

[33] F. Jia and B. Mukherjee. Bimodal throughput, non-monotonic delay, optimal bandwidth dimensioning, and analysis of receiver collisions in a single-hop WDM local Lightwave network. In *IEEE GLOBECOM '92,* pages 1896-1900, Orlando, FL, December 1992.

[34] M. Chen and T.-S. Yum. A conflict-free protocol for optical WDMA networks. In *IEEE GLOBECOM '91,* pages 1276-1281, Phoenix, AZ, December 1991.

[35] M. Chen and T.-S. Yum. Buffer sharing in conflict-free WDMA networks. *IEICE Trans. on Commun.,* E77-B(9):1144-1151, September 1994.

[36] H. B. Jeon and C. K. Un. Contention-based reservation protocol in fibre optic local area network with passive star topology. *Electronics Letters,* 26(12): 780-781, June 1990.

[37] H. B. Jeon and C. K. Un. Contention-based reservation protocols in multiwavelength optical networks with a passive star topology. *IEEE Trans. on Commun.,* 43(11):2794-2802, November 1995.

[38] J. H. Lee and C. K. Un. Dynamic scheduling protocol for variable-sized messages in a WDM-based local network. *IEEE/OSA J. of Lightwave Technol.,* 14(7):1595-1600, July 1996.

[39] H. S. Kim, B. C. Shin, J. H. Lee, and C. K. Un. Performance evaluation of reservation protocol with priority control for single-hop WDM networks. *Electronics Letters,* 31(17):1472-1473, August 1995.

[40] F. Jia and B. Mukherjee. The receiver collision avoidance (RCA) protocol for a single-hop WDM Lightwave network. *IEEE/OSA J. of Lightwave Technol.,* 11(5/6):1053-1065, May/June 1993.

[41] F. Jia and B. Mukherjee. Performance analysis of a generalized receiver collision avoidance (RCA) protocol for single-hop WDM local Lightwave networks. In *Proc., SPIE '92,* pages 229-240, Boston, MA, September 1992.

[42] F. Jia and B. Mukherjee. A high-capacity, packet-switched, single-hop local Lightwave network. In

IEEE GLOBECOM '93, pages 1110-1114, Houston, TX, December 1993.

[43] G. I. Papadimitriou and D. G. Maritsas. Learning automata-based receiver conflict avoidance algorithms for WDM broadcast-and-select star networks. *IEEE/ACM Trans. on Networking,* 4(3):407-412, June 1996.

[44] G. I. Papadimitriou and D. G. Maritsas. Self-adaptive random-access protocols for WDM passive star networks. *IEE Proc.-Comput. Digit. Tech.,* 142(4): 306-312, July 1995.

[45] R. Chipalkatti, Z. Zhang, and A. S. Acampora. Protocols for optical star-coupler network using WDM: Performance and complexity study. *IEEE J. on Sel. Areas in Commun.,* 11(4):579-589, May 1993.

[46] K. Bogineni and P. W. Dowd. A collisionless multiple access protocol for a wavelength division multiplexed star-coupled configuration: Architecture and performance analysis. *IEEE/OSA J. of Lightwave Technol.,* 10(11):1688-1699, November 1992.

[47] K. M. Sivalingam and P. W. Dowd. A multilevel WDM access protocol for an optically interconnected multiprocessor system. *IEEE/OSA J. of Lightwave Technol.,* 13(11):2152-2167, November 1995.

[48] K. M. Sivalingam and J. Wang. Media access protocols for WDM networks with on-line scheduling. *IEEE/OSA J. of Lightwave Technol.,* 14(6):1278-1286, June 1996.

[49] M. Kovačević and M. Gerla. Analysis of a T/WDMA scheme with subframe tuning. In *IEEE ICC '93,* pages 1239-1244, Geneva, Switzerland, May 1993.

[50] M. Kovačević and M. Gerla. Time and wavelength division multiaccess with acoustooptic tunable filters. *Fiber and Integrated Optics,* 12:113-132, November 1992.

[51] J. H. Laarhuis and A. M. J. Koonen. An efficient medium access control strategy for high-speed WDM multiaccess networks. *IEEE/OSA J. of Lightwave Technol.,* 11(5/6):1078-1087, May/June 1993.

[52] F. Jia, B. Mukherjee, and J. Iness. Scheduling variable-length messages in a single-hop multichannel local lightwave network. *IEEE/ACM Trans. on Networking,* 3(4):477-488, August 1995.

[53] A. Mokhtar and M. Azizoglu. Hybrid multiaccess for all-optical LANs with nonzero tuning delays. In *IEEE ICC '95,* pages 1272-1276, Seattle, WA, June 1995.

[54] K. M. Sivalingam and S. Subramaniam, editors. *Optical WDM Networks—Principles and Practice, Chapter 9.* Kluwer Academic Publishers, 2000.

[55] M. Maier and A. Wolisz. Single-hop WDM network with high spectrum reuse based on an arrayed waveguide grating. In *Optical Network Workshop,* University of Dallas, TX, Jan./Feb. 2000.

[56] L. Giehmann, A. Gladisch, N. Hanik, and *et al.* The application of code division multiple access for transport overhead information in transparent optical networks. In *OFC 1998 Technical Digest, paper WM42,* pages 228-229, San Jose, CA, February 1998.

[57] M. Maier, M. Reisslein, and A. Wolisz. High-performance switchless WDM network using multiple free spectral ranges of an arrayed-waveguide grating. In *Terabit Optical Networking: Architecture, Control and Management Issues, Part of SPIE Photonics East 2000,* volume 4213, pages 101-112, Boston, MA, November 2000.

Martin Maier
maier@ee.tu-berlin.de

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 thru 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
reisslein@asu.edu

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 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. He is particularly interested in traffic management for multimedia services with statistical Quality of Service in the Internet and wireless communication systems.

Adam Wolisz
wolisz@ee.tu-berlin.de

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 a member of the Senior Board of GMD Fokus being especially in charge of the Competence Centers GLONE and TIP.

He is acting as a member of the Steering Committee of the Computer Engineering 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 Quality of Service 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 2 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.