Wavelength Reuse for Efficient Transport of Variable–Size Packets in a Metro WDM Network

Martin Maier Michael Scheutzow Martin Reisslein Adam Wolisz

Abstract— **Metro WDM networks play an important role in the emerging Internet hierarchy; they interconnect the backbone WDM networks and the local access networks. The current SONET/SDH–over–WDM–ring metro networks are expected to become a serious bottleneck — the so–called metro gap — as they are faced with an increasing amount of bursty data traffic and quickly increasing bandwidths in the backbone networks and access networks. Innovative metro WDM networks that are highly efficient and able to handle variable–size packets are needed to alleviate the metro gap. In this paper we study an AWG–based single–hop WDM metro network. We analyze the photonic switching of variable–size packets with spatial wavelength reuse. We derive computationally efficient and accurate expressions for the network throughput and delay. Our extensive numerical investigations — based on our analytical results and simulations — reveal that spatial wavelength reuse is crucial for efficient photonic packet switching. In typical scenarios, spatial wavelength reuse increases the throughput by 60% while reducing the delay by 40%.**

Keywords— **Arrayed Waveguide Grating; Medium Access Control; Metro WDM Network; Multiple Free Spectral Ranges; Photonic Packet Switching; Spatial Wavelength Reuse; Variable–Size Packets.**

I. INTRODUCTION

THE INTERNET of the future may be viewed as a three level hierarchy consisting of backbone networks, metro networks, and access networks. Backbone networks will provide almost infinite bandwidth based on Wavelength Division Multiplexing (WDM) links. These WDM links are connected with reconfigurable Optical Add–Drop Multiplexers (OADMs) and Optical Cross Connects (OXCs) controlled by Multiprotocol Optical Lambda Switching (MPλS), Optical Burst Switching (OBS), and Optical Packet Switching (OPS) mechanisms. Access networks transport data to (and from) individual users. By employing advanced LAN technologies, such as Gigabit Ethernet, broadband access, such as xDSL and cable modems, as well as high–speed next–generation wireless systems, such as UMTS, access networks provide an ever increasing amount of bandwidth. Metro networks interconnect the high–speed WDM backbone networks and the high–speed access networks. Current metro networks are typically based on SONET/SDH–over–WDM rings which carry the ever increasing amount of bursty data traffic only inefficiently. In addition, content providers increasingly place proxy caches in metro networks. These proxies further increase the load on metro networks. Metro networks are therefore expected to become a serious bottleneck — the so–called metro gap — in the future Internet [1]. For these reasons there is an urgent need for innovative metro network architectures and protocols [2].

Two key requirements for metro networks are (1) flexibility, and (2) efficiency. Flexibility is required since metro networks have to support a wide range of heterogeneous protocols, such as ATM, Frame Relay, SONET/SDH, and IP. This requires, in particular, that the metro networks are able to transport *variable–size packets*. Efficiency is required because metro networks are highly cost sensitive. Therefore, the deployed WDM networking components and the WDM networking resources (in particular wavelengths) must be utilized efficiently. As we demonstrate in this paper, a crucial technique for achieving high efficiency is *spatial wavelength reuse*. By spatial wavelength reuse we mean that in our AWG–based single–hop network (outlined in Section II) all wavelengths are used at all AWG ports simultaneously.

This paper builds on earlier work [3], in which we have proposed an Arrayed–Waveguide Grating (AWG)–based single–hop WDM network. This earlier work focused primarily on the network architecture and the Medium Access Control (MAC) protocol. The elementary analysis conducted in [3] provided very limited insights into the performance of the proposed network. The performance analysis in [3] is limited in that it considered only fixed–size packets and did not consider spatial wavelength reuse. However, the efficient transmission of variable–size packets is

This work was supported in part by Deutsche Telekom AG.

Please direct correspondence to M. Reisslein.

M. Maier and A. Wolisz are with the Telecommunication Networks Group, Technical University Berlin, 10587 Berlin, Germany, (e-mail: maier@ee.tu-berlin.de, wolisz@ee.tu-berlin.de).

M. Scheutzow is with the Dept. of Mathematics, Technical University Berlin, (e-mail: ms@math.tu-berlin.de).

M. Reisslein is with the Telecommunications Research Center, Dept. of Electrical Engineering, Arizona State University, P.O. Box 877206, Tempe AZ 85287–7206, Phone: (480)965–8593, Fax: (480)965–8325, (email: reisslein@asu.edu, web: http://www.eas.asu.edu/˜mre).

of paramount importance for future metro networks. In this paper we study (1) the photonic switching of variable–size packets, and (2) the spatial wavelength reuse in the AWG–based network proposed in earlier work. The main contribution of this paper is to develop a stochastic model to evaluate the performance of the transport of variable–size packets with spatial wavelength reuse in the AWG–based network. Our analytical model gives computationally efficient and accurate expressions for the throughput and delay in the network. Our numerical results indicate that the AWG–based network can efficiently transport variable–size packets. We also find that spatial wavelength reuse is crucial for efficient photonic packet switching. For typical scenarios, spatial wavelength reuse increases the throughput by 60% while reducing the delay by 40%.

This paper is organized as follows. In the following subsection we give a quick overview of related work. In Section II we briefly review the architecture of the studied AWG–based single–hop network as well as the reasoning for selecting this architecture. In Section III we briefly review the MAC protocol for the studied network. In Section IV we develop a stochastic model for the performance evaluation of the transmission of variable–size packets with spatial wavelength reuse. This model and performance evaluation are our main contributions. In Section V we use our analytical results to conduct numerical investigations. We also conduct simulations to verify the accuracy of our analytical results. We conclude in Section VI.

A. Related Work

Metro WDM networks have just recently begun to attract the interest of the research community [4]. A metro network based on optical add–drop multiplexers (OADMs) is studied in [5]. This network is geared towards optical circuit switching. The HORNET metro network [6], [7] allows for photonic packet switching. Both, HORNET [6], [7] and the OADM–based metro network [5] have a ring topology, i.e., transmissions typically have to traverse multiple network nodes. These networks are therefore fundamentally different from the single–hop network studied here. We note that an AWG–based single–hop WDM network is also studied in [8]. However, this network has higher hardware requirements and is envisaged as a wide–area network.

II. Architecture

Initial and operational costs of metro WDM networks can be dramatically reduced by deploying passive wavelength–selective arrayed–waveguide gratings (AWGs) [9]. Moreover, AWGs with a crosstalk as low as -40 dB significantly increase the network capacity by spatially reusing all wavelengths at each AWG input port [10], [11]. Deploying *athermal* AWGs which do not require any temperature control further reduces network costs and simplifies network management [12]. Due to their inherent frequency–cyclic nature, AWG–based networks can be upgraded gracefully [13]. To capitalize on these advantages, the considered metro WDM network is based on a $D \times D$ AWG, as shown in Fig. 1. At each AWG input port a wavelength–insensitive $S \times 1$ combiner collects data from S attached nodes. Similarly, at each AWG output port signals are distributed to S nodes by a wavelength–insensitive $1 \times S$ splitter (note that these splitters can also be used for optical multicasting). 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. Thus, the network connects $N = D \cdot S$ nodes. Each node contains a tunable laser diode (LD) and a tunable photodiode (PD) for data transmission and reception, respectively. In addition, each node uses a broadband light source, e.g., a light emitting diode (LED), for broadcasting control information by means of spectral slicing. By spreading the control information before externally modulating the LED at the transmitting side of a node, data and control can be transmitted simultaneously without requiring an additional receiver at each node. At the receiving part of a node the control information is retrieved by despreading a part of the incoming signal. At each AWG input port we exploit R adjacent FSRs of the AWG, each FSR consisting of D contiguous wavelengths. For a more detailed discussion of the architecture the interested reader is referred to [3].

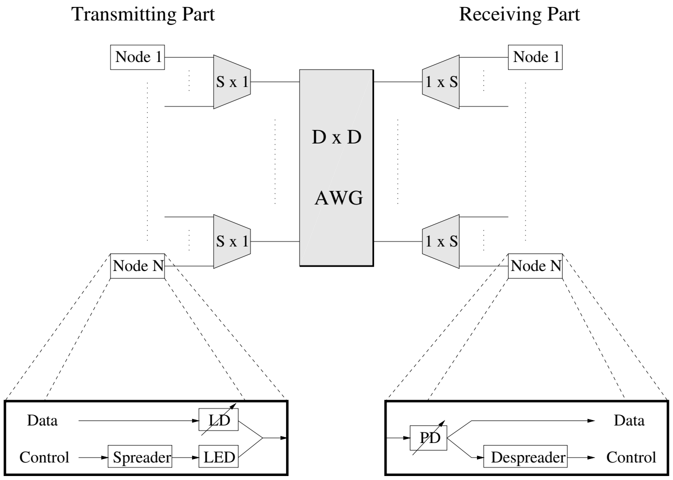

Fig. 1. Network and node architecture

III. MAC Protocol

In this section we give a brief overview of our MAC protocol; we refer the interested reader to [3] for more details. Our MAC protocol is an attempt–and–defer type of protocol, i.e., a data packet is only transmitted after the corresponding control packet has been successful. In our MAC protocol time is divided into *cycles*. Each cycle consists of D *frames*. Each frame contains F *slots*. The slot length is equal to the transmission time of a control packet. Each frame is partitioned into the first M, $1 \leq M < F$ slots and the remaining $(F - M)$ slots. In the first M slots, control packets are transmitted and all nodes must be tuned (locked) to one of the LED slices carrying the control information. In every frame within a cycle, the nodes attached to a different AWG input port send their control packets. Specifically, all nodes attached to AWG input port o, $1 \leq o \leq D$, (via a common combiner) send their control packets in frame o of the cycle. During the first M slots of frame o control and data packets can be transmitted simultaneously by the nodes attached to AWG input port o. Transmissions from the other AWG input port can not be received during this time interval. In the last $(F - M)$ slots of each frame no control packets are sent. The receivers are unlocked, allowing transmission between any pair of nodes. This allows for spatial wavelength reuse — the main focus of this paper.

When a data packet arrives to a node attached to AWG input port o, the node's LED broadcasts the corresponding control packet in one of the first M slots of the frame assigned to AWG input port o. The control packet has four fields: destination address, length, type (packet/circuit switched, see [3]), and forward error correction. The control packet is transmitted on a contention basis using a modified version of slotted Aloha. Every node collects all control packets by locking its receiver to one of the LED slices carrying the control information during the first M slots of every frame. Thus, each node maintains global knowledge of all the other nodes' activities (and also learns whether its own control packet collided in the slotted Aloha contention or not). If a control packet collides, it is retransmitted in the next cycle with probability p; with probability $(1 - p)$ the retransmission is deferred by one cycle. The successfully received control packets are processed by all nodes. Each node applies the same scheduling algorithm and thus comes to the same conclusion. We assume, for simplicity, first–come–first–served–first–fit scheduling. The algorithm tries to schedule the data packets within the scheduling window of D frames (i.e., one cycle). If the scheduling fails, the source node retransmits the control packet.

IV. Analysis

A. System and Traffic Model

In our analysis we consider a system with a large S. Our analysis is approximate for finite S and exact in the asymptotic limit $S \to \infty$. However, our numerical investigations (see Section V) indicate that our analysis is very accurate even for moderate S, e.g., $S = 25$. Throughout our analysis we assume that the propagation delay is no larger than one cycle. Thus, if a control packet is sent in a given frame, the corresponding data packet could be scheduled for transmission one cycle later. We assume that all nodes are equidistant from the AWG, i.e., the propagation delay is the same for all nodes.

We assume that each node has a buffer that can hold a single data packet and a single control packet. We make the following assumptions about the traffic generation process. Suppose that a node's control packet has just been (1) successfully transmitted, and (2) the corresponding data packet has been successfully scheduled (within the scheduling window of one cycle; see Section IV-C). With probability σ this node then generates the control packet for the next data packet right before the beginning of the next frame in which the node can send the next control packet (i.e., one cycle after the previous control packet was sent). If no control packet is generated, then the node waits for one cycle and then generates a new control packet with probability σ, and so on. The node's buffer may hold the scheduled (but not yet transmitted) data packet and the next control packet at the same time. This next control packet is sent with probability one in the next frame assigned to the node's AWG input port (possibly simultaneously with the scheduled data packet). A data packet is purged from the node's buffer at the end of the frame during which it is transmitted. After a data packet is purged from the buffer, the next data packet is placed in the buffer, provided the corresponding control packet is already in the buffer.

If a control packet fails in the slotted Aloha contention or the data packet scheduling, then the node retransmits the control packet in the next frame assigned to the node's AWG input port with probability p, with probability $(1-p)$ it defers the retransmission by one cycle. In this next cycle the node transmits the control packet with probability p and defers the transmission with probability $(1 - p)$, and so on. We define

$$\tilde{\alpha} := \frac{S\sigma}{M} \text{ and } \alpha := \frac{Sp}{M}. \qquad (1)$$

We conduct an approximate analysis for large S. Our analysis becomes asymptotically exact when $S \to \infty$ and $\tilde{\alpha}$ as well as α (and also M) are fixed (with σ and p chosen so as to satisfy (1)).

We assume uniform unicast traffic. A data packet is destined to any one of the N nodes (including the sending node, for simplicity) with equal probability $1/N$. Without loss of generality we consider two packet sizes. Let L denote the length of a data packet in slots. A data packet is long (has size $L = F$) with probability q, i.e., $P(L = F) = q$. A data packet is short (has size $L = K,\ 1 \leq K \leq F - M$) with probability $(1 - q)$, i.e., $P(L = K) = 1 - q$. If a control packet fails (either in the slotted Aloha or the scheduling) the size of the corresponding data packet is not changed. However, we do assume nonpersistency [14] for the destination, i.e., a new random destination is drawn for each attempt of transmitting a control packet.

Now consider the nodes attached to a given (fixed) AWG input port $o,\ 1 \leq o \leq D$. These nodes send their control packets in frame o of a given cycle. We refer to the nodes that at the beginning of frame o hold an old packet, that is, a control packet that has failed in slotted Aloha or scheduling, as *"old"*. We refer to all the other nodes as *"new"*. Note that the set of "new" nodes comprises both the nodes that have generated a new (never before transmitted) control packet as well as the nodes that have deferred the generation of a new control packet. Let η be a random variable denoting the number of "new" nodes at AWG input port o, and let

$$\nu := \frac{E[\eta]}{S}. \tag{2}$$

Let λ_l be a random variable denoting the number of nodes at port o that are to send a control packet corresponding to a long data packet next (irrespective of whether a given node is "old" or "new", and keeping in mind that the set of "new" nodes also comprises those nodes that have deferred the generation of the next control packet; those nodes are accounted for in λ_l if the next generated control packet corresponds to a long data packet). Let λ_k be a random variable denoting the number of nodes at port o that are to send a control packet corresponding to a short data packet next. By definition, $\lambda_k = S - \lambda_l$. Let

$$\tilde{q} := \frac{E[\lambda_l]}{S} \tag{3}$$

denote the expected fraction of long packets to be sent. We expect that $\tilde{q}$ is typically larger than q since long packets are harder to schedule and thus typically require more retransmissions (of control packets).

B. Analysis of Slotted Aloha Contention

First, we calculate the number of control packets from nodes attached to AWG input port $o,\ 1 \leq o \leq D$, that are successful in the slotted Aloha contention in frame o. Let $Y_i^n,\ i = 1, \ldots, M$, be a random variable denoting the number of control packets that were randomly transmitted in slot $i,\ i = 1, \ldots, M$, by "new" nodes. Recall that each of the η "new" nodes sends a control packet with probability σ in the frame. Thus,

$$P(Y_i^n = k) = \binom{\eta}{k}\left(\frac{\sigma}{M}\right)^k\left(1 - \frac{\sigma}{M}\right)^{\eta-k}, \quad k = 0, 1, \ldots, \eta. \tag{4}$$

Throughout our analysis we assume that S is large and that $\tilde{\alpha}$ and α are fixed. We may therefore reasonably approximate the BIN($\eta,\ \sigma/M$) distribution with a Poisson ($\eta\sigma/M$) distribution, that is,

$$P(Y_i^n = k) \approx e^{-\eta\sigma/M}\frac{(\eta\sigma/M)^k}{k!}, \quad k = 0, 1, \ldots, \tag{5}$$

which is exact for $\eta \to \infty$ with $\eta\sigma/M$ fixed. We now recall the definition $\tilde{\alpha} := \sigma S/M$. We also approximate η/S by its expectation ν; this is reasonable since η/S has only small fluctuations in steady state for large S. Thus,

$$P(Y_i^n = k) \approx e^{-\tilde{\alpha}\nu}\frac{(\tilde{\alpha}\nu)^k}{k!}, \quad k = 0, 1, \ldots. \tag{6}$$

We note that for $S \to \infty$ the random variables $Y_1^n, Y_2^n, \ldots, Y_M^n$ are mutually independent. This is because a given node places with the miniscule probability σ/M a control packet in a given slot, say slot 1. (Note in particular that the expected value of Y_1^n is small compared to the number of "new" nodes, that is, $\tilde{\alpha}\nu \ll \eta$; this is because in the considered asymptotic limit $S \to \infty$ with α fixed, we have $1 \gg \sigma/M = \tilde{\alpha}\nu/\eta$.) Thus, Y_1^n has almost no impact on $Y_2^n, \ldots, Y_M^n$ (see [15] for a formal proof, which we can not include here because of page limitations).

Let $Y_i^o,\ i = 1, \ldots, M$, be a random variable denoting the number of control packets in slot $i,\ i = 1, \ldots, M$, that originate from "old" nodes. Each of the $(S - \eta)$ "old" nodes sends a control packet with probability p in the frame. With an analysis similar to the analysis for Y_i^n, we find that Y_i^o is approximately distributed according to a Poisson distribution with parameter $\alpha(1 - \nu)$. We note again that for $S \to \infty$ the random variables $Y_1^o, Y_2^o, \ldots, Y_M^o$ are mutually independent. They are also independent of $Y_1^n, Y_2^n, \ldots, Y_M^n$. Hence, we obtain for $Y_i = Y_i^n + Y_i^o$

$$P(Y_i = k) \approx e^{-[\tilde{\alpha}\nu + \alpha(1-\nu)]}\frac{[\tilde{\alpha}\nu + \alpha(1 - \nu)]^k}{k!},$$

$$k = 0, 1, \ldots . \quad (7)$$

Henceforth, we let for notational convenience

$$\beta := \tilde{\alpha}\nu + \alpha(1 - \nu), \quad (8)$$

i.e.,

$$P(Y_i = k) \approx e^{-\beta} \frac{\beta^k}{k!}, \; k = 0, 1, \ldots . \quad (9)$$

Let $X_i, \; i = 1, \ldots, M$, be a random variable indicating whether or not slot i contains a successful control packet. Specifically, let

$$X_i = \begin{cases} 1 & \text{if } Y_i = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (10)$$

¿From (9) clearly, $P(X_i = 1) = \beta e^{-\beta}$ and $P(X_i = 0) = 1 - \beta e^{-\beta}$ for $i = 1, \ldots, M$. The total number of successful control packets in the considered frame is $\sum_{i=1}^{M} X_i$, which has a BIN$(M, \; \beta e^{-\beta})$ distribution, that is,

$$P\left(\sum_{i=1}^{M} X_i = l\right) = \binom{M}{l} \left(\beta e^{-\beta}\right)^l \left(1 - \beta e^{-\beta}\right)^{M-l}, \quad l = 0, 1, \ldots, M. \quad (11)$$

Recall from Section IV-A that with the assumed uniform unicast traffic each packet is destined to any one of the D AWG output ports with equal probability $1/D$. Let Z denote the number of successful control packets — in the considered frame — that are destined to a given (fixed) AWG output port $d, \; d = 1, \ldots, D$. Clearly, from (11)

$$P(Z = k) = \binom{M}{k} \left(\frac{\beta e^{-\beta}}{D}\right)^k \left(1 - \frac{\beta e^{-\beta}}{D}\right)^{M-k}, \quad k = 0, 1, \ldots, M. \quad (12)$$

Let Z_l denote the number of successful control packets that correspond to long data packets destined to a given AWG output port d. Recall that $\tilde{q}$ is the expected fraction of long packets to be sent. Hence, $Z_l \sim BIN(M, \; \beta e^{-\beta} \frac{1}{D} \tilde{q})$. Similarly, let Z_k denote the number of control packets that are successful in the slotted Aloha contention and correspond to short data packets destined to a given AWG output port d. Clearly, $Z_k \sim BIN(M, \; \beta e^{-\beta} \frac{1}{D}(1 - \tilde{q}))$.

C. Analysis of Packet Scheduling

In this section we calculate the expected number of packets that are successfully scheduled. Recall from the previous section that the total number of long packets that (1) originate from a given AWG input port $o, \; 1 \leq o \leq D$, (2) are successful in the slotted Aloha contention of frame o (of a given cycle), and (3) are destined to a given AWG output port $d, \; 1 \leq d \leq D$, is $Z_l \sim BIN(M, \; \beta e^{-\beta} \frac{1}{D} \tilde{q})$. For short packets we have $Z_k \sim BIN(M, \; \beta e^{-\beta} \frac{1}{D}(1 - \tilde{q}))$. Note that these two random variables are not independent. Let $\mathcal{L}$ ($\mathcal{S}$) be a random variable denoting the number of long (short) packets that (1) originate from a given AWG input port $o, \; 1 \leq o \leq D$, (2) are successful in the slotted Aloha contention of frame o (of a given cycle), (3) are destined to a given AWG output port $d, \; 1 \leq d \leq D$, and (4) are successfully scheduled within the scheduling window of D frames (i.e., one cycle).

Consider the scheduling of packets from a given (fixed) AWG input port o to a given (fixed) AWG output port d over the scheduling window (i.e., D frames). Clearly, we can schedule at most R long packets (i.e., $\mathcal{L} \leq R$) because the receivers at output port d must tune to the appropriate spectral slices during the first M slots of every frame. Thus they can tune to a node at AWG input port o for F consecutive slots, only in the frame, during which the nodes at AWG input port o send their control packets.

Now, suppose that $\mathcal{L}$ ($\leq R$) long packets are scheduled (how $\mathcal{L}$ is determined is discussed shortly). With $\mathcal{L}$ long packets already scheduled, we can schedule at most

$$\mathcal{S} \leq (D - 1) \cdot R \cdot \left\lfloor \frac{F - M}{K} \right\rfloor + (R - \mathcal{L}) \left\lfloor \frac{F}{K} \right\rfloor \quad (13)$$

short packets. To see this, note that in the frame during which the nodes at AWG input port o send their control packets, there are $(R - \mathcal{L})$ FSRs — channels between AWG input port o and AWG output port d — still free for a duration of F consecutive slots. Furthermore, there are $(D - 1)$ frames in the scheduling window during which the nodes at AWG output port d must tune (are locked) to the nodes sending control packets from the other AWG input ports for the first M slots of the frame. During each of these frames, the receivers are unlocked for $(F - M)$ slots. The R utilized FSRs provide R parallel channels between AWG input port o and AWG output port d. Note that the $(D - 1)R\lfloor(F - M)/K\rfloor$ component in (13) is due to the spatial reuse of wavelengths at the considered AWG input port. Without spatial wavelength reuse this component would be zero and we could schedule at most $(R - \mathcal{L})\lfloor F/K \rfloor$ short packets. Continuing our analysis for a network with spatial wavelength reuse, we have

$$\mathcal{S} = \min\left\{ Z_k, \; (D - 1) \cdot R \cdot \left\lfloor \frac{F - M}{K} \right\rfloor + (R - \mathcal{L}) \left\lfloor \frac{F}{K} \right\rfloor \right\}. \quad (14)$$

In (14) we neglect receiver collisions, that is, we do not account for situations where a packet can not be scheduled because its receiver is already scheduled to receive a different packet. This assumption is reasonable as receiver collisions are rather unlikely for large S, which we assume throughout our analysis.

In this paper we consider for simplicity a first–come–first–served–first–fit scheduling policy. Data packets are scheduled for the first possible slot(s) at the lowest available wavelength. To arbitrate the access to the long (F slots) transmission slots and the short ($(F-M)$ slots) transmission slots we adopt the following *arbitration policy*. Our arbitration policy proceeds in one round if there are R or less successful control packets in the slotted Aloha contention. In case there are more than R successful control packets in the slotted Aloha contention, our arbitration policy proceeds in two rounds. First, consider the case where R or less control packets are successful in the slotted Aloha contention and we have one round of arbitration. In this case all the successful packets are scheduled in the long (F slot) transmission slots. Next, consider the case where more than R control packets are successful in the slotted Aloha contention and we have two rounds of arbitration. In this case we scan the M slotted Aloha slots from index 1 through M. In the first round we schedule the first R successful packets out of the slotted Aloha contention in the R long (F slot) transmission slots. In this round we schedule only one packet for each of the long transmission slots, irrespective of whether the packet is long or short. At this point (having filled each of the long transmission slots with one data packet) all the remaining successful control packets that correspond to long data packets fail in the scheduling and the transmitting node has to re–transmit the control packet. We then proceed with the second round. In the second round we schedule the remaining successful control packets that correspond to short data packets. Provided $F/K > 2$, we schedule these short data packets for the long transmission slots that hold only one short data packet from the first round. We also schedule these short data packets for the short ($(F-M)$ slot) transmission slots. After all the long and short transmission slots have been filled, the remaining short data packets fail in the scheduling and the transmitting node has to retransmit the control packet. We note that our adopted arbitration policy is just one out of many possible arbitration policies, all of which can be analyzed in a similar fashion.

With the adopted arbitration policy the expected number of scheduled long packets is

$$E[\mathcal{L}] \quad = \quad \sum_{k=0}^{M} E[\mathcal{L}|Z=k]\cdot P(Z=k), \qquad (15)$$

which after some analysis (see [15] for the details, which we can not include here because of page limitations) gives

$$E[\mathcal{L}] = \tilde{q}\cdot\varphi(\beta), \qquad (16)$$

where

$$\varphi(\beta) := R - \sum_{k=0}^{\min(R,\,M)} \binom{M}{k}\left(\frac{\beta e^{-\beta}}{D}\right)^{k} \cdot \left(1-\frac{\beta e^{-\beta}}{D}\right)^{M-k}(R-k). \qquad (17)$$

We now calculate the expected number of scheduled short packets. Generally,

$$E[\mathcal{S}] \quad = \quad \sum_{k=0}^{M} E[\mathcal{S}|Z=k]\cdot P(Z=k). \qquad (18)$$

We obtain after some analysis (see [15]) that

$$\begin{aligned} E[\mathcal{S}] \quad &= \quad (1-\tilde{q})\left[R-\sum_{k=0}^{R}(R-k)\cdot P(Z=k)\right] + \\ &\sum_{j=1}^{M-R}\gamma_j \sum_{m=j}^{M-R}\sum_{k=m+R}^{M}\binom{k-R}{m}(1-\tilde{q})^m \cdot \\ &\tilde{q}^{k-R-m}\cdot P(Z=k) \\ =: \quad & h(\tilde{q},\,\beta), \end{aligned} \qquad (19)$$

where γ_j is computed as follows. If $\lfloor F/K \rfloor - 1 > 0$ then

$$\gamma_j \quad = \quad \sum_{\{m:\, m\le v_j\}} \binom{R}{m}\tilde{q}^{m}(1-\tilde{q})^{R-m}, \qquad (20)$$

where

$$v_j := \min\left(R,\ \frac{(D-1)R\left\lfloor \frac{F-M}{K}\right\rfloor - j}{\left\lfloor \frac{F}{K}\right\rfloor - 1} + R\right). \qquad (21)$$

If $\lfloor F/K \rfloor = 1$ then

$$\gamma_j = \begin{cases} 1 & \text{if } j \le (D-1)R\left\lfloor \frac{F-M}{K}\right\rfloor \\ 0 & \text{otherwise.} \end{cases} \qquad (22)$$

We note that this result applies to a network *with spatial wavelength reuse*. For a network *without spatial wavelength reuse* the term $(D-1)R\lfloor (F-M)/K \rfloor$ has to be replaced by zero in (21) and (22).

D. Network/System Analysis

In this section we put the analyses for the individual components of the considered network, namely traffic model, slotted Aloha contention, and packet scheduling, together. We establish two equilibrium conditions and solve for the two unknowns $\tilde{q}$ and β. (Alternatively, we may consider the two unknowns $\tilde{q}$ and ν, noting that $\nu = (\beta - \alpha)/(\tilde{\alpha} - \alpha)$ for $\tilde{\alpha} \neq \alpha$; the case $\tilde{\alpha} = \alpha$ is discussed at the end of this section.)

In steady state the system satisfies the equilibrium condition

$$E[\mathcal{L}] = q(E[\mathcal{L}] + E[\mathcal{S}]). \tag{23}$$

To see this, note that in equilibrium the mean number of scheduled long packets from a given (fixed) AWG input port destined to a given (fixed) AWG output port (LHS) is equal to the mean number of newly generated long packets (RHS). Inserting (16) and (19) in (23) gives

$$\begin{aligned} \tilde{q} \cdot \varphi(\beta) &= q[\tilde{q} \cdot \varphi(\beta) + h(\tilde{q},\, \beta)] && (24)\\ \Leftrightarrow (1-q) \cdot \tilde{q} \cdot \varphi(\beta) &= q \cdot h(\tilde{q},\, \beta). && (25) \end{aligned}$$

The second equilibrium condition is

$$\frac{\sigma}{D} \cdot E[\eta] = E[\mathcal{L} + \mathcal{S}]. \tag{26}$$

This is because $\sigma \cdot \eta$ new packets are generated in each frame at the nodes attached to a given AWG input port. With probability $1/D$ each of the generated packets is destined to a given (fixed) AWG output port. On the other hand, $E[\mathcal{L} + \mathcal{S}]$ packets are scheduled (and transmitted) on average from a given AWG input port to a given AWG output port in one cycle; in equilibrium as many new packets must be generated. Inserting (1) and (2) in the LHS of (26), and (16) and (19) in the RHS of (26) we obtain

$$\frac{\tilde{\alpha} \cdot M}{D} \cdot \frac{\beta - \alpha}{\tilde{\alpha} - \alpha} = \tilde{q} \cdot \varphi(\beta) + h(\tilde{q},\, \beta). \tag{27}$$

Inserting (27) in (25) we obtain

$$\tilde{q} = \frac{q \cdot \tilde{\alpha} \cdot M}{D \cdot \varphi(\beta)} \cdot \frac{\beta - \alpha}{\tilde{\alpha} - \alpha}. \tag{28}$$

Inserting (28) in (27) we obtain

$$(1-q) \cdot \frac{\tilde{\alpha} \cdot M}{D} \cdot \frac{\beta - \alpha}{\tilde{\alpha} - \alpha} = h\left(\frac{q \cdot \tilde{\alpha} \cdot M}{D \cdot \varphi(\beta)} \cdot \frac{\beta - \alpha}{\tilde{\alpha} - \alpha},\, \beta\right). \tag{29}$$

We solve Equation (29) numerically to obtain β (noting that by (8), $\min(\tilde{\alpha},\, \alpha) \leq \beta \leq \max(\tilde{\alpha},\, \alpha)$). We then insert β in (28) to obtain $\tilde{q}$. With β and $\tilde{q}$ we calculate $E[\mathcal{L}]$ (16) and $E[\mathcal{S}]$ (19). We define the mean throughput as the mean number of successfully transmitted data packets per frame. The mean throughput from a given (fixed) AWG input port to a given (fixed) AWG output port is then given by

$$TH_{\text{port}} := \frac{F \cdot E[\mathcal{L}] + K \cdot E[\mathcal{S}]}{F \cdot D}. \tag{30}$$

The *mean aggregate throughput* of the network is

$$TH_{\text{net}} = D^2 \cdot TH_{\text{port}}. \tag{31}$$

We note that in case $\tilde{\alpha} = \alpha$, i.e., $\sigma = p$, we have from (8) $\beta = \alpha$. Inserting this in (25) gives an equation for $\tilde{q}$, which we solve numerically.

We now espouse the mean packet delay in the network. We define the mean delay as the average time period in cycles from the generation of the control packet corresponding to a data packet until the transmission of the data packet. Recall from Section IV-C that $E[\mathcal{L}] + E[\mathcal{S}]$ is the expected number of data packets that the nodes at a given AWG input port transmit to the nodes at a given AWG output port per cycle. Now, consider a given (fixed) node m, $1 \leq m \leq N$. In the assumed uniform packet traffic scenario, this node m transmits on average $(E[\mathcal{L}]+E[\mathcal{S}])/S$ data packets to the nodes at a given AWG output port per cycle. Thus, node m transmits on average $(E[\mathcal{L}] + E[\mathcal{S}])D/S$ data packets to the N nodes attached to the D AWG output ports per cycle. The average time period in cycles from the generation of a control packet at node m until the generation of the next control packet is therefore $S/[D \cdot (E[\mathcal{L}] + E[\mathcal{S}])]$. Note that the time period from the successful scheduling of a data packet until the generation of the control packet for the next data packet is geometrically distributed with mean $(1 - \sigma)/\sigma$ cycles. Hence, the average delay in the network in cycles is

$$\text{Delay} = \frac{S}{D \cdot (E[\mathcal{L}] + E[\mathcal{S}])} - \frac{1 - \sigma}{\sigma}. \tag{32}$$

where $E[\mathcal{L}]$ and $E[\mathcal{S}]$ are known from the evaluation of the throughput (30).

V. Numerical Results

In this section, we demonstrate the benefit of spatial wavelength reuse and investigate the impact of the system parameters on the throughput–delay performance of the network. Recall from Section IV-A that a data packet is F slots long with probability q, and $K = (F - M)$ slots long with probability $(1 - q)$. By default the parameters take on the following values: number of used FSRs $R = 2$, fraction of long data packets $q = 0.25$, number of reservation slots per frame $M = 30$, physical degree of the AWG $D = 4$, number of nodes $N = 200$, retransmission probability $p = 0.8$, number of slots per frame $F = 200$, and length of short packets $K = 170$ slots. Each cycle is assumed to have a constant length of $D \cdot F = 800$ slots. We also provide extensive simulation results of a more realistic network in order to verify the accuracy of the analysis. As opposed to the analysis, in the simulation a given node can not transmit data packets to itself. Furthermore, in the simulation both length and destination of a given data packet are not renewed (i.e., are persistent) when retransmitting the corresponding control packet (the analysis assumes that the length of the data packet is persistent, while the destination is non–persistent). In addition, the simulation takes receiver collisions into account, i.e., a given data packet is not scheduled if the receiver of the intended destination node is busy. Each simulation was run for 10^7 slots including a warm–up phase of 10^6 slots. Using the method of batch means we also calculated the 98% confidence intervals for the mean aggregate throughput and the mean delay whose widths are less than 1% of the corresponding sample means for all simulation results.

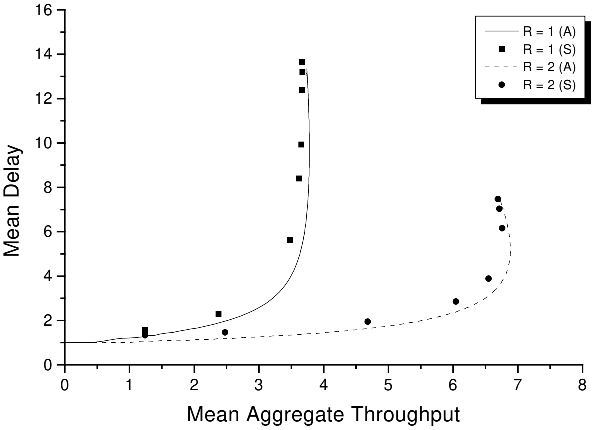

Fig. 2. Mean delay (cycles) vs. mean aggregate throughput (packets/frame) for different number of used FSRs $R \in \{1,\ 2\}$

Fig. 2 shows the mean delay vs. the mean aggregate throughput as the mean arrival rate σ is varied from 0 to 1. We observe that the simulation provides slightly larger delay values than the analysis. This is because the simulation takes also the transmission time of data packets into account as opposed to the analysis. In the analysis the mean delay is equal to the time interval between the generation of a given data packet and the end of the cycle in which the given data packet is successfully scheduled but not yet transmitted. Apparently, using two FSRs instead of one (leaving all other parameters unchanged) dramatically increases the mean aggregate throughput while decreasing the mean delay. This is due to the fact that an additional FSR increases the degree of concurrency and thereby mitigates the scheduling bottleneck resulting in more successfully transmitted data packets and fewer retransmissions. Note that the number of used FSRs is limited and is determined by the transceiver tuning range, the degree of the underlying AWG, and the channel spacing. To avoid tuning penalties we deploy fast tunable transceivers whose tuning range is typically $10 - 15$ nanometers (nm). All results presented in this section assume a channel spacing of 200 GHz, i.e., 1.6 nm at 1.55 μm. Thus, we can use $7 - 10$ wavelengths at each AWG input port depending on the transceiver tuning range. For all subsequent results the number of wavelengths is assumed to be eight. Consequently, with a 4×4 AWG we deploy two FSRs for concurrent transmission/reception of data packets.

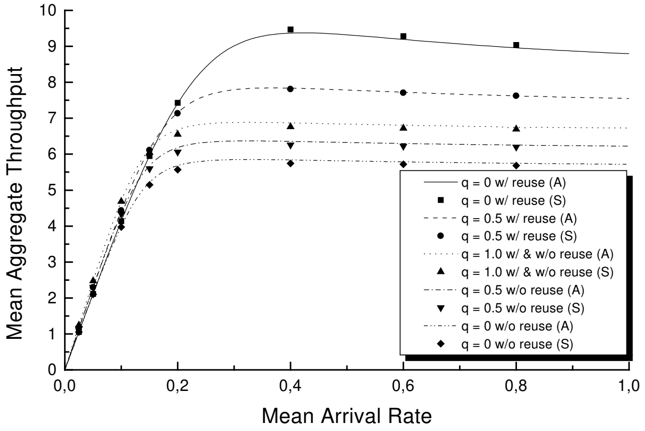

Fig. 3. Mean aggregate throughput (packets/frame) vs. mean arrival rate σ with and without wavelength reuse for different fraction of long data packets $q \in \{0,\ 0.5,\ 1.0\}$

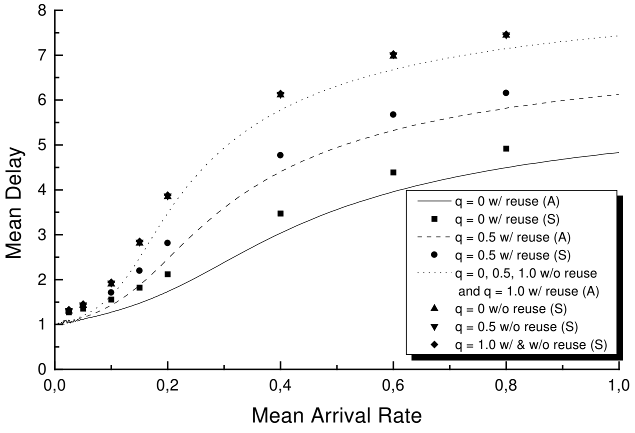

Fig. 4. Mean delay (cycles) vs. mean arrival rate σ with and without wavelength reuse for different fraction of long data packets $q \in \{0,\ 0.5,\ 1.0\}$

Figs. 3 and 4 illustrate that spatial wavelength reuse dramatically improves the throughput–delay performance of the network for variable–size data packets. Fig. 3 shows the mean aggregate throughput vs. the mean arrival rate σ with and without spatial wavelength reuse for different fraction of long data packets $q \in \{0,\ 0.5,\ 1.0\}$. Simulation and analysis results match very well. For $q = 1.0$, i.e., all data packets have a length of F slots, the mean aggregate throughput is the same no matter whether wavelengths are spatially reused or not. This is because the data packets are too long for being scheduled in the $(D - 1)$ frames in which the corresponding nodes do not send control packets and spatial wavelength reuse would be possible in the last $(F - M)$ slots of the frame. Thus, these frames remain unused for $q = 1.0$. For $q = 0.5$, 50% of the data packets are long (F slots) and the other 50% are short (K slots). Allowing for spatial wavelength reuse the latter ones can now be scheduled in all frames including the aforementioned $(D - 1)$ frames. Consequently, with wavelength reuse more data packets are successfully transmitted resulting in a higher throughput. In contrast, without wavelength reuse data packets can be scheduled only in one frame per cycle in which the corresponding nodes also transmit their control packets. Furthermore, since for $q = 0.5$ some successfully transmitted data packets are short (K slots), wavelengths are not fully utilized resulting in a lower throughput compared to $q = 1.0$. For $q = 0$ the benefit of spatial wavelength reuse becomes even more dramatic. In this case there are only short data packets (K slots) which fill up a large number of frames leading to a further increased mean aggregate throughput. Note that for $q = 0$ spatial wavelength reuse significantly increases the maximum aggregate throughput by more than 60%. All curves in Fig. 3 run into saturation since for increasing σ no additional data packets can be scheduled due to busy channels and receivers and an increasing number of colliding control packets.

Fig. 4 depicts the mean delay vs. σ with and without wavelength reuse for different fraction of long data packets $q \in \{0,\ 0.5,\ 1.0\}$. Again, the simulation gives slightly larger delay values because of the aforementioned reason. All curves have in common that at very light traffic the mean delay is equal to one cycle owing to the propagation delay of the control packet. With increasing σ the mean delay increases due to more unsuccessful control packets. These control packets have to be retransmitted resulting in an increased mean delay. Note that we obtain the largest delay if the aforementioned $(D - 1)$ frames per cycle can not be used for data transmission. This holds not only for the cases where wavelength reuse is not deployed but also for $q = 1.0$ with spatial wavelength reuse. This is due to the fact that for $q = 1.0$ the data packets are too long and do not fit in the last $(F - M)$ slots of the aforementioned $(D - 1)$ frames. As a consequence, for these cases fewer data packets can be successfully scheduled and the corresponding control packets have to be retransmitted more often, leading to a higher mean delay. With decreasing q there are more short data packets which can easily be scheduled in the aforementioned $(D - 1)$ frames. Due to the resulting wavelength reuse more data packets can be successfully scheduled. Therefore, fewer control packets have to be retransmitted, leading to a decreased mean delay. In particular, for $q = 0$ wavelengths are used very efficiently resulting in the lowest mean delay.

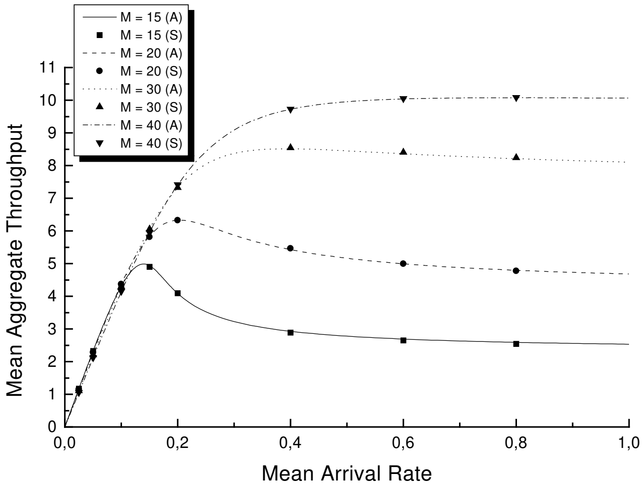

Fig. 5. Mean aggregate throughput (packets/frame) vs. mean arrival rate σ for different number of reservation slots $M \in \{15,\ 20,\ 30,\ 40\}$

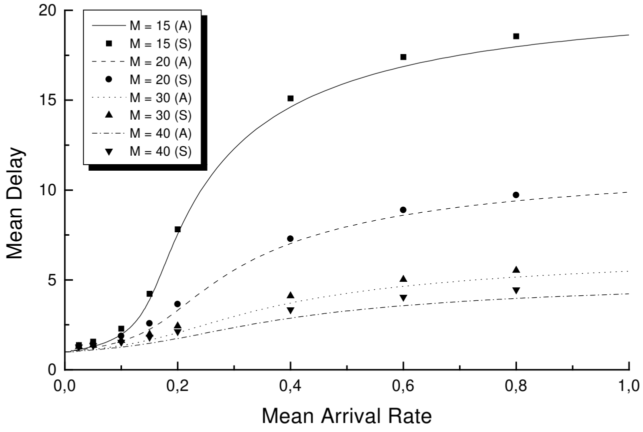

Fig. 6. Mean delay (cycles) vs. mean arrival rate σ for different number of reservation slots $M \in \{15,\ 20,\ 30,\ 40\}$

The impact of the number of reservation slots M per frame on the network throughput–delay performance is shown in Figs. 5 and 6. The mean aggregate throughput and the mean delay are depicted as a function of σ for $M \in \{15,\ 20,\ 30,\ 40\}$. Recall that by default the frame length F is set to 200 slots. Each frame is composed of M reservation slots and $K = (F - M)$ slots which can be used for transmitting short packets by means of spatial wavelength reuse. Clearly, for a fixed F increasing M decreases the length of short packets K, and vice versa for decreasing M. As shown in Figs. 5 and 6, increasing the number of reservation slots significantly improves the throughput–delay performance of the considered network. Due to the reduced contention more control packets are successfully transmitted resulting in an increased mean aggregate throughput and a decreased mean delay. Thus, in terms of the throughput–delay performance it is advantageous to use more reservation slots per frame even though this implies a smaller K. This also indicates that the random access reservation scheme can be a severe bottleneck. Note that the network throughput–delay performance could be easily improved by replacing the random access of the reservation slots with a dedicated assignment of the reservation slots. However, such a dedicated assignment does not scale very well.

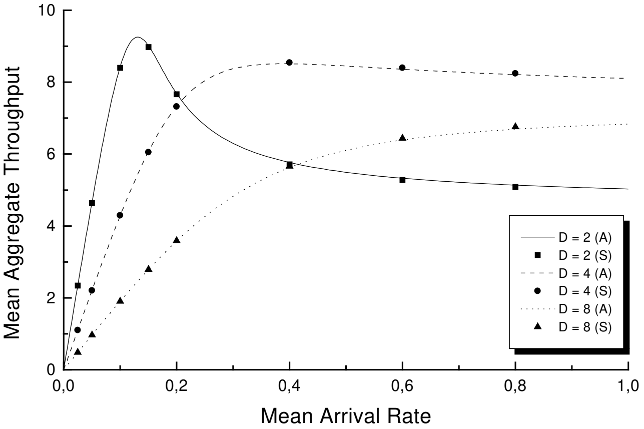

Fig. 7. Mean aggregate throughput (packets/frame) vs. mean arrival rate σ for different AWG degree $D \in \{2,\ 4,\ 8\}$

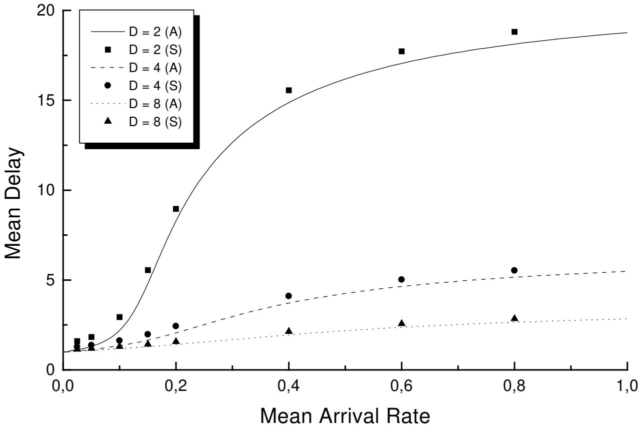

Fig. 8. Mean delay (cycles) vs. mean arrival rate σ for different AWG degree $D \in \{2,\ 4,\ 8\}$

A given number of nodes can be connected by AWGs with different physical degree D. Figs. 7 and 8 depict for $D \in \{2,\ 4,\ 8\}$ the mean aggregate throughput and the mean delay as a function of σ, respectively. Recall that we have chosen the transceiver tuning range and the channel spacing such that we make use of eight wavelengths. The number of used FSRs R is then determined only by the physical degree D of the underlying AWG and is given by $R = 8/D$. Consequently, for a smaller D more FSRs can be exploited, and vice versa for a larger D. Furthermore, for a smaller D each cycle contains fewer but longer frames, and vice versa for a larger D.

As shown in Fig. 7, $D = 2$ provides the largest maximum mean aggregate throughput at light traffic. However, with increasing σ the mean aggregate throughput decreases. This is due to the fact that for $D = 2$ short data packets are rather long ($K = 800/D - M = 370$ slots) resulting in a higher channel utilization and thereby a higher throughput at small traffic loads. But a small D also implies that for a given population N more nodes are attached to the same combiner since $S = N/D$. All these S nodes make their reservations in the same frame. For an increasing σ this leads to more collisions of control packets resulting in a lower mean aggregate throughput and a higher mean delay due to more retransmissions of the corresponding control packets (Fig. 8).

This problem is alleviated by deploying a 4×4 or a 8×8 AWG. For a larger D fewer nodes send control packets in the same frame causing fewer collisions at high traffic loads. However, for $D = 4$ and $D = 8$ only 2 FSRs and 1 FSR can be deployed, respectively. Moreover, a larger D reduces the length of short data packets as well. Fig. 7 shows that for $D = 4$ the mean aggregate throughput is rather high for a wide range of σ. Whereas for $D = 8$ the throughput is rather low due to the small number of control packets per frame and the low channel utilization owing to the reduced length of short data packets. Note that for $D = 4$ the mean aggregate throughput gradually decreases for increasing σ. This is because at high traffic loads control packets suffer from collisions and have to be retransmitted, resulting in a slightly higher mean delay compared to $D = 8$. Concluding, in terms of throughput–delay performance choosing $D = 4$ seems to provide the best solution for a wide range of traffic loads.

Figs. 9 depicts the throughput–delay performance of the network for different population $N \in \{40,\ 100,\ 200,\ 300\}$. As shown in Fig. 9, increasing N improves the mean aggregate throughput due to more reservation requests and successfully scheduled data packets. However, for $N = 200$ and especially $N = 300$ the throughput decreases for increasing σ. This is because for large populations more

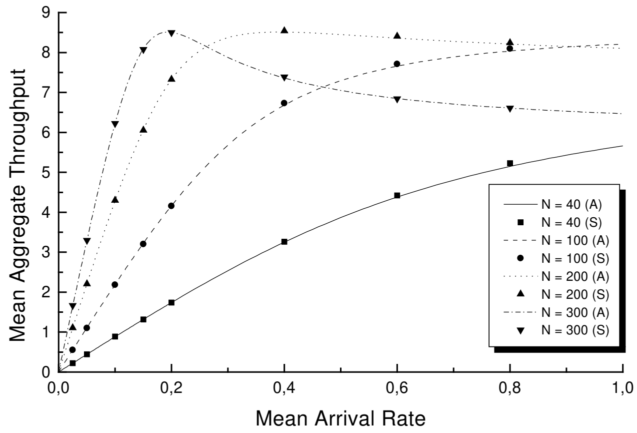

Fig. 9. Mean aggregate throughput (packets/frame) vs. mean arrival rate σ for different population $N \in \{40,\ 100,\ 200,\ 300\}$

control packets suffer from channel collisions resulting in a lower mean aggregate throughput. Note that simulation and analysis results match very well even for small populations despite the fact that (1) we have conducted an asymptotic analysis for large S, and (2) the analysis does not take receiver collisions into account (while the simulation does).

VI. Conclusion

We have analyzed the photonic switching of variable–size packets with spatial wavelength reuse in an AWG–based single–hop WDM network. We have obtained computationally efficient and accurate expressions for the throughput and delay in the network. Based on our analytical results we have conducted extensive numerical investigations of the performance characteristics of the network. We have also conducted extensive simulations to verify the accuracy of the analytical results. Our numerical results indicate that the AWG–based single–hop network, originally proposed in [3], can efficiently transport variable–size packets. We found that spatial wavelength reuse is crucial for efficient photonic packet switching. Spatial wavelength reuse significantly increases the throughput while dramatically reducing the delay.

In our ongoing work we are studying the optimal trade–offs of the network parameters, e.g., the AWG degree that maximizes the throughput (and minimizes the delay) for a given number of nodes (and traffic load). We are also studying multicasting in the AWG–based single–hop network.

References

[1] L. G. Kazovsky, K. Shrikhande, I. M. White, M. Rogge, and D. Wonglumsom, "Optical metropolitan area networks," in *OFC 2001 Technical Digest, paper WU1*, Anaheim, CA, Mar. 2001.

[2] B. Mukherjee, "WDM optical communication networks: Progress and challenges," *IEEE J. on Sel. Areas in Commun.*, vol. 18, no. 10, pp. 1810–1824, Oct. 2000.

[3] 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*, Boston, MA, Nov. 2000, vol. 4213, pp. 101–112, Paper won the *Best Paper Award* of the conference.

[4] S. Yao, S. J. B. Yoo, and B. Mukherjee, "All–optical packet switching for metropolitan area networks: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 39, no. 3, pp. 142–148, March 2001.

[5] 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.*, vol. 39, no. 2, pp. 106–113, Feb. 2001.

[6] K. V. Shrikhande, I. M. White, D. Wonglumsom, S. M. Gemelos, M. S. Rogge, Y. Fukashiro, M. Avenarius, and L. G. Kazovsky, "HORNET: A packet–over–WDM multiple access metropolitan area ring network," *IEEE J. on Sel. Areas in Commun.*, vol. 18, no. 10, pp. 2004–2016, Oct. 2000.

[7] I. M. White, K. Shrikhande, M. S. Rogge, M. Gemelos, D. Wonglumsom, G. Desa, Y. Fukashiro, and L. G. Kazovsky, "Architecture and protocols for HORNET: A novel packet–over–WDM multiple–access MAN," in *Proc., IEEE Globecom*, San Francisco, CA, Nov./Dec. 2001.

[8] N. P. Caponio, A. M. Hill, F. Neri, and R. Sabella, "Single–layer optical platform based on WDM/TDM multiple access for large–scale 'switchless' networks," *European Trans. on Telecommun.*, vol. 11, no. 1, pp. 73–82, Jan./Feb. 2000.

[9] F. Ruehl and T. Anderson, "Cost–effective metro WDM network architectures," in *OFC 2001 Technical Digest, paper WL1*, Anaheim, CA, Mar. 2001.

[10] K. Kato, A. Okada, Y. Sakai, K. Noguchi, T. Sakamoto, A. Takahara, S. Kamei, A. Kaneko, S. Suzuki, and M. Matsuoka, "10–Tbps full–mesh WDM network based on cyclic–frequency arrayed–waveguide grating router," in *ECOC 2000*, Munich, Germany, Sept. 2000, vol. 1, pp. 105–107.

[11] A. Okada, T. Sakamoto, Y. Sakai, K. Noguchi, and M. Matsuoka, "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, Mar. 2001.

[12] N. Keil, H. H. Yao, C. Zawadzki, J. Bauer, M. Bauer, C. Dreyer, and J. Schneider, "Athermal polarization–independent all–polymer arrayed waveguide grating (AWG) multi/demultiplexer," in *OFC 2001 Technical Digest, paper PD7*, Anaheim, CA, March 2001.

[13] M. D. Feuer, S. L. Woodward, C. F. Lam, and M. L. Boroditsky, "Upgradeable metro networks using frequency–cyclic optical add/drop," in *OFC 2001 Technical Digest, paper WBB5*, Anaheim, CA, March 2001.

[14] J. Lu and L. Kleinrock, "A wavelength division multiplexing access protocol for high–speed local area networks with a passive star topology," *Performance Evaluation*, vol. 16, no. 1–3, pp. 223–239, Nov. 1992.

[15] M. Maier, M. Scheutzow, M. Reisslein, and A. Wolisz, "Wavelength reuse for efficient transport of variable–size packets in a metro WDM network (extended version)," Tech. Rep., Arizona State University, Telecommunications Research Center, July 2001, available at `http://www.eas.asu.edu/˜mre`.