Proceedings of the 1998 Conference on Information Sciences and Systems

---

Volume II

Department of Electrical Engineering
Princeton University
Princeton, New Jersey 08544-5263

Scaling WDM Slotted Ring Networks

Alberto Bononi
Università di Parma – Dipartimento di Ingegneria dell'Informazione – Viale delle Scienze, I-43100 Parma, Italy
phone: +39-521-905-760 fax: +39-521-905-758 – email: bononi@tlc.unipr.it

Abstract— Using analogies between the basic theory of cell switching for datagram networks and multiwavelength slotted optical networks, we show that the buffering capabilities of high-speed slotted networks, also known as space-reuse, allows multihop networks such as WDM rings to have larger throughputs than star networks, where space reuse is not possible. By finding correspondences with the classical switching theory, WDM ring networks are extensively analyzed in uniform traffic. The multihop nature of the network, and the clear identification of the meaning of *hop* as it relates to throughput, help identify a novel generalized ring topology, which we call WDM Chordalring, which smoothly scales the throughput of a WDM ring by progressively adding WDM unbuffered routers up to reaching the desired saturation throughput. Routing without buffers is studied in the novel topology by simulation, and an empirical routing scheme is found that gives satisfactory performance. Using such scheme, the topology is optimized and a clear trade-off between number of wavelengths, number of receivers per node and number of routers is identified and quantified.

I. Introduction

This paper deals with the analysis of wavelength routing optical ring networks, and the proposal of a generalized ring topology, for datagram applications in the local/metropolitan area.

First the classical theory of time-slotted cell switches is briefly reviewed, with emphasis on the effect of buffering (input vs output), speed-up and stability, in order to be able to map the problem of slotted WDM ring networks onto the the classical theory. Such mapping allows a better understanding of the impact of number of nodes, number of wavelengths, number of receivers per node, tunability at the transmitter, at the receiver, or both, on the maximum sustainable throughput, called the saturation throughput.

When the number of available wavelengths is a small fraction of the number of nodes, more richly interconnected topologies are necessary to provide a desired saturation throughput to every node. By placing a number of 2x2 unbuffered WDM dynamic routers (each being a stack of independently operated 2x2 switches, one per wavelength) on the unidirectional WDM slotted ring, we enrich the topology with chords (or shortcuts) that decrease the mean number of hops (or active node crossings) and thus increase throughput. Since routers are unbuffered, cells may either be routed on a preferred path or deflected. The routing problem is quite different from previously published work on deflection routing, and new routing strategies must be used.

In such a generalized WDM ring network we will show how routers can be traded for wavelengths to achieve a target saturation throughput. The proposed topology is a variation on chordal rings [1], and is similar to the SMARTnet proposed by I. Rubin [2], although our routers are 2x2, links are unidirectional, and the routing is datagram as opposed to virtual circuit as in [2].

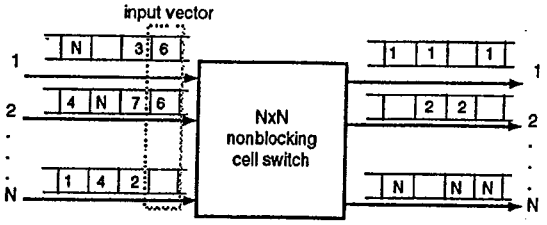

Fig. 1. A time-slotted cell switch.

II. Review of Classical Cell Switches

In this section we will briefly recall the fundamentals of cell switching. Consider the slotted nonblocking space switch shown in Fig. 1, with Bernoulli arrivals with intensity g cells/slot, $0 \le g \le 1$, independent from slot to slot and from port to port, and uniform output port destination.

Contentions for the output ports may arise in the vector of slots at the input of the switch, which we call the *input vector*.

We define the *speedup* L as the maximum number of cells that can be transferred from the input vector to an output port in one time-slot.

With a speedup of 1, the average number of cells delivered to each output port per slot, i.e. the Throughput per destination node, is simply $T = 1 - (1 - g/N)^N$ since the port will receive 0 cells if all N slots in the input vector are not destined to it, and 1 cell otherwise. For N large (in practice $N > 5$) this converges to $T(g) = 1 - e^{-g}$, giving at full load $T(1) \simeq 0.63$, and the arrivals A to each port converge to a Poisson random variable of rate g [3]. At full load, 63% of the cells at the input make it to their output port, and the rest are lost. However, if the switch matrix is speeded up by a factor L, less cells are lost. The number of cells arriving to each output port is then $O = max(A, L)$. The throughput is found as usual as the average number of cells arriving at each output per slot: $T = E[O] = \sum_{i=0}^{L-1} iP\{A = i\} + L(1 - \sum_{i=0}^{L-1} P\{A = i\})$. For large N, such equation gives $T(1) = 0.63, 0.89, 0.97, 0.99$ for L=1,2,3, and 4, respectively: a speedup above 2 gives essentially unity saturation throughput.

In an actual implementation, dropping of blocked packets is undesirable, and hence buffers are provided. These are placed at the input to store cells that are not selected for transmission to the desired output, and at the output in case of speedup $L > 1$ to store the cells transfered to the output, if the line rate of the output link is equal to that of the input link. Such buffers are commonly implemented as first-in-first-out (FIFO) queues [3].

With infinitely large input and output buffers, one can

analytically get [3],[4] the mean waiting time versus offered load, which in the absence of cell loss coincides with the throughput across the switch. One finds [4] that input buffering (L=1) has the worst performance, with a stable input queue up to a maximum arrival rate of $2-\sqrt{2} \cong 0.58$ cells/slot/port. As the load approaches this value, which is called the *saturation throughput* T_{sat} or *maximum sustainable load*, the queue quickly fills up and the waiting time diverges to infinity. Not even an infinitely large buffer can sustain a rate larger that that. In practice, no matter how many buffers we place at the input, loss will occur of all the input flux in excess of T_{sat}. Note that, for infinite input buffers and FIFO queuing, $T_{sat} = 0.58 < 0.63$. That is the throughput with buffers is lower at full load ($g = 1$) than that of a system without buffers [3]. This surprising result is due to the fact that with input buffers the backlog of unselected cells in the input vector creates strong correlations in the destinations of the cells. Such correlations are not present when cells are dropped. Whenever fresh cells with uniform destination distribution step into the input vector at all clock times, the saturation throughput increases to 0.63. This is the case when for example the head-of-line (HOL) cells that are not selected are thrown randomly back in the queue.

The poor performance of FIFO input buffering is due to the well-known HOL blocking, i.e., to cells behind unselected HOLs that would have been able to be transmitted to an otherwise idle output port [3].

It is well known [5],[6] that smarter queuing policies than FIFO can lead to unity saturation throughput even with input buffering, as for instance by having *at each input* a set of N FIFO queues, one per output port, from which one can suitably select cells. Therefore not only the hardware (switch speedup) plays a role in setting the saturation throughput, but also the software (queuing policy).

The arrival process itself plays a key role in setting the saturation throughput. With unbalanced arrivals some queues may saturate well before the others, or bursty arrivals may cause buffers to fill up more quickly.

Finally, the distribution of the destination port of each cell plays also an important role, and as it becomes more localized (cells are more likely to be destined to output ports neighboring the corresponding input port) the saturation throughput may increase even with input buffers, since there are less destination collisions in the input vector. These last two factors that determine T_{sat} are referred to as a "traffic" dependence.

Summarizing, T_{sat} depends on hardware, software and traffic, and can be chosen as a parameter of merit for datagram switching systems, useful for simple comparisons.

A. Applications to Optics

The switch could be implemented by an NxN crossbar (or matrix) switch, as shown in Fig. 2 (top), which consists of input and output buses, each operating at the line rate, with N^2 crosspoints [9]. Examples of optical NxN crossbar switches are also given in Fig. 2. The middle figure shows an implementation with input and output broadcast lines and N^2 selecting gates [10], and the bottom figure an alternative implementation based on N central lines to which inputs and outputs connect by selectors.

All three schemes implement the same crossbar architecture. The bottom scheme has received a lot of attention fo broadcast-and-select WDM shared medium networks [11]-[15]. In such networks, either the transmitting lasers or th receiving filters were tunable, corresponding to the selec tors in the bottom figure, while N wavelength channels (th N central lines in the figure) were available for transmis sion. Users were connected by dedicated fibers to a centra hub where WDM signals get added and broadcast. Whe N channels are available, simultaneous tunability at bot transmitters and receivers is not needed; it may be usefu only when the number of channels is less than N, a situatio typical when the network scales, but the crossbar switch i this case looses its desirable nonblocking properties. W will return to this important point in the next sections.

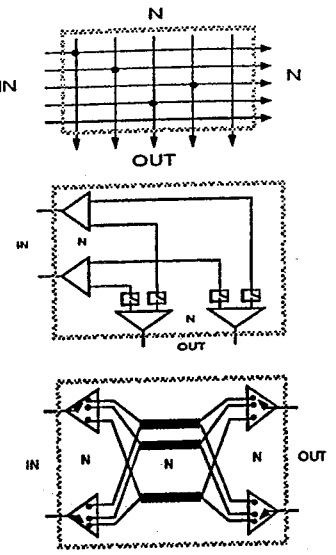

Fig. 2. Different realizations of the matrix switch: (top) classical (middle) with input/output splitting and N^2 gates; (bottom) with N channels and input/output selectors.

Note that some of the proposals adopt a slotted operatio [11]–[13], while some avoid that for simplicity, accepting lower saturation throughput (this is 0.5 in [15] as opposed t 0.58 of the corresponding slotted system). Some proposal adopt input buffering [11], while others adopt various form of output buffering [12][13], being this the most natur form of buffering, given the built-in speedup of N provide by the broadcast optical medium.

To conclude, note that tunability at the periphery, i.e at the access of the network (we call this *peripheral switci ing*), with the transport part of the network being passiv is a peculiar characteristic of local area networks (LAN as opposed to *centralized switching* of classical wide-ar networks.

III. WDM SLOTTED RINGS

We are now ready to analyze WDM slotted ring network and discuss how they relate to classical space switches ar WDM shared medium networks. In high speed cell switc ing networks the duration of the cell, containing a fix number of bits, is becoming shorter and shorter as cor pared to the propagation time, so that a cell, thought stretching in space while propagating, is much smaller th the size of the network it moves in, even in LANs. This lea

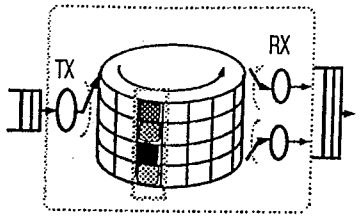

Fig. 3. Mapping WDM slotted rings onto the classical theory.

to the concept of space-reuse often pointed out as the major difference with respect to traditional, low speed networks: several cells can be carried simultaneously through the optical medium. This is nothing but an added *buffering* effect offered for free by the network, as we will see shortly. Space reuse is not possible in shared medium WDM networks, in which there is a hub, a mixing point where signals from the various nodes are combined to be broadcast. However WDM ring networks do allow space reuse, and thus they do have better network performance than star networks, as indicated by larger saturation throughputs.

Thus, let us focus on a slotted WDM ring network, in which there are p wavelength channels, N nodes, and each node has one (possibly tunable) transmitter and one, or more (possibly tunable) receivers. From a logical standpoint the network looks like the switch shown schematically in Fig. 3.

It consists of a rotating drum composed of p slotted and aligned levels, each slot capable of carrying a cell.

A tunable transmitter/receiver will be able to scan the column of the drum passing in front of the node, and inject/absorb cells from it. Here the role of the input vector of HOLs in classical switches is played by each column of the drum. Such HOLs have a service time that differs from the classical case, and is connected to the number of slot-times it takes the drum to complete a rotation.

We want to find out how the problem is related to classical cell switches, and how the number of wavelengths, number of receivers (RX) per node, tunability at the transmitter, at the receiver, or both, affect performance in terms of saturation throughput. We will do this in uniform traffic, to have consistent comparisons with the classical theory.

By our very definition, the speedup here is given by the number of cells a node is able to remove from the network in one slot, i.e., by the number of receivers. Therefore we will have input (electronic) buffers at the TX to store cells which cannot be injected in the drum (because the input vector is full, or because of alternative access policies we could adopt), and output (electronic) buffers if more than one optical receiver is provided at the node. This system has also extra buffering (space reuse) provided by the drum itself.

A. N Wavelengths, 1 Tunable TX, 1 Fixed RX

Let's analyze the case of N wavelengths, 1 tunable TX, and 1 fixed RX per node. Each wavelength can be seen as a bus dedicated to the corresponding receiver, polling for cells destined to it as the bus slots pass in front of each node [16]. Focus the attention on a generic wavelength λ_i. A cell will be injected on the bus only if the corresponding slot is free upon passing the node. Let's follow a test slot on λ_i as it rotates, starting from node i and circulating back to i.

Suppose at first nodes *do not* have electronic input buffers, and uniform traffic is feeding each transmitter. Arrivals are Bernoulli with intensity g, with uniform destinations. At each intermediate node the test slot polls and gets a cell for node i with probability $\frac{g}{N-1}$. The probability that all N-1 independent polls after a cycle are negative is $(1-\frac{g}{N-1})^{(N-1)}$, and therefore the throughput per node, i.e. the probability that the test slot gets a cell in one bus cycle, is the probability that at least one poll was positive: $T(g) = 1-(1-\frac{g}{N-1})^{(N-1)}$. This expression coincides with the throughput of the classic switch without input buffers. At full load this gives $T(1) \cong 0.63$ for N large.

If we include electronic input buffers at the TX to hold cells that cannot be injected, we can define g as the probability that the TX has a cell ready for injection at each slot. However, the destination distribution of the (electronic) HOL cells tends to lose its uniformity, and to be different for the various nodes, since the first node along the bus has the highest priority for injections, while the last has the lowest. Therefore it is less likely to find an electronic HOL destined to node i at the first node of the bus, node $i+1$ if numbering is in the sense of rotation, than to find it at the last, where blocking on bus i due to upstream nodes is more frequent. The same phenomenon occurs in classical input buffered switches if the winning HOL selection, instead of being random as in the previous analysis, is cyclic, with some priority ordering given to the input queues. Simulation results for $N = 20, 40,$ and 60 nodes show that at full load the throughput T is, respectively, $0.602, 0.596, 0.593$ with electronic buffers (converging to $2\text{-}\sqrt{2}$ for large N), and $0.643, 0.637, 0.636$ without such buffers (corresponding to $[1-(1-\frac{1}{N-1})^{N-1}]$). The cyclic service policy of the electronic queues does not alter the full load throughput value.

B. N wavelengths, 1 Tunable RX, 1 Fixed TX

Consider now the case of N wavelengths, 1 tunable RX, and 1 fixed TX per node. Each receiver has access to N-1 dedicated buses from which it receives cells from the other nodes. Since blocking occurs at the TX only when the present slot on its bus has a cell that was not absorbed by its destination receiver, and such blocking is thus not related to the destination of the cells ready at the TX, the performance will be the same both with and without electronic input buffers at the TX. A little thought will indicate that this case logically corresponds to a classical input buffered packet switch, whose input vector is now the column of the drum available at each receiver at each clock. Correlations of cell destinations in the drum columns have the same effect of lowering the per node throughput, as they do in classical switches. The throughput performance is then the same as that of a classical input buffered switch, with a full load value of 0.58 for N large.

As simple trick can be used to break the columnwise correlations in the drum, and thus increase throughput to 0.63. It is enough to insert a device that adds different delays to the various channels. Proposed implementations of access nodes in WDM rings do employ a demux/mux structure for the access nodes[17],[18] that may serve the purpose.

Simulation results for $N = 20, 40,$ and 60 nodes show that at full load the throughput T is, respectively, $0.600, 0.593, 0.591$ for equal length rings (no device inserted, slowly converging to $2\text{-}\sqrt{2}$), and $0.6430, 0.6426, 0.6421$ with unequal rings (device inserted, corresponding to $[1-(1-\frac{1}{N-1})^{N-1}]$). These values are the same as in the previous section.

C. $p < N$ Wavelengths, 1 Tunable TX, 1 Fixed RX per Node

So far there were as many channels as nodes in the network. However scalability requirements dictate that in practice the number of available wavelengths p is much less than the number of nodes N. Since now a wavelength cannot be entirely dedicated to a single node, either the node has a tunable receiver, or its receiver is fixed and a subset of the nodes share the same receiver wavelength. Let's consider the latter case, delaying the analysis of the former to the next section.

Assume there are $N = m * p$ nodes, divided in groups of m equally-spaced nodes receiving on the same wavelength. Assume also that there are no electronic queues at the TXs, and that the traffic is uniform with intensity g.

Focus on a generic wavelength λ_i, with nodes $N_1, .., N_m$ receiving on it. In each segment between consecutive receiving nodes there are $\frac{N}{m} - 1$ active nodes capable of placing cells on λ_i. If we think of each of these sets of nodes as a single generating node co-located with the upstream receiving node, the ring at wavelength λ_i can be seen as a standard ring with m active nodes, capable of both transmitting and receiving. By symmetry, all nodes produce the same amount of traffic, uniformly destined to the other nodes, including themselves. The probability that no cell is generated by all the (independent) active nodes in a segment is $(1-\frac{m\,g}{N-1})^{N/m-1}$, which converges to e^{-g} for $N \to \infty$. Thus the ring at wavelength λ_i is equivalent to a conventional m-node slotted ring, with uniform traffic to all nodes (including the generating node) and Bernoulli arrivals with intensity $g' = 1 - e^{-g}$. By Little's law, the throughput per node is $T = u/H$, where u is the link load, i.e. the fraction of full slots, and H is the average number of hops, in this case $H = (m+1)/2$. The probability r that a cell at a receiving node is destined to it is just the inverse of H: $r = 2/(m+1)$. To find u, we balance injected and absorbed throughput at each node. The average number of injections per slot is $T_{inj} = g'(1-u(1-r))$, since a cell must be ready, and injection occurs only if, after removal of cells destined to the node, the incoming slot is empty, which occurs with probability $(1-u(1-r))$. The average number of absorbed cells per slot is just $T_{abs} = ur$. By equating $T_{inj} = T_{abs}$ we get $u = \frac{g'}{r+g'(1-r)}$. Thus we finally get:

$$\boxed{T(g) = \frac{2}{m+1}\frac{1-e^{-g}}{\frac{2}{m+1}+(1-e^{-g})(1-\frac{2}{m+1})}} \qquad (1)$$

D. $p < N$ wavelengths, 1 Tunable TX, $L < p$ Tunable RX

Consider 1 tunable TX and $L < p$ tunable receivers per node (the speedup). We consider here electronic queues at each TX, and let g be the probability of having at least a cell ready at the TX for injection. To keep the traffic balanced on the rings, nodes inject on the various wavelengths at random whenever more than one wavelength slot is free. Since the injection probability does not depend on the destination of the cells, in uniform traffic the destinations of the electronic HOLs will be uniform. Hence the analysis is identical with and without electronic buffers, provided g has the appropriate meaning.

In the analytical derivation, we consider only the case of unequal length rings, so that columnwise correlations in the drum are broken and we can assume that arrivals to the node from each wavelength in the ring are independent. The load u is the same on each wavelength, because of the load balancing injection policy. Again we let r be the probability that a cell entering the node has reached its destination. Slots entering the node can be: FN= for the node; TG=through-going; E=empty. The probabilities of such events are: P{FN}=ur; P{TG}=u(1-r); P{E}=1-u.

D.1 Absorption Probability a

We seek the probability of absorption a of a FN input *test* cell. Let q be the number of FN cells present on the remaining $p-1$ channels. If $q < L$, then the *test* cell is absorbed with probability one. For $q \geq L$, the probability the test is selected is $\frac{L}{q+1}$, since L cells chosen at random among the q available are absorbed.

The probability of having q FN cells on the remaining $p-1$ channels is a binomial, so that by the total probability law we have:

$$a = 1 - \sum_{q=L}^{p-1}\left(1-\frac{L}{q+1}\right)\binom{p-1}{q}(ur)^q(1-ur)^{p-1-q}. \qquad (2)$$

From (2) we get the following particular cases:

$$\begin{cases} a = \frac{1-(1-ur)^p}{urp} & \text{if } L = 1 \\ a = 1 & \text{if } L = p. \end{cases} \qquad (3)$$

D.2 Injected and Absorbed Throughput

In the assumption of independent input slots, the average number of injected cells per slot is:

$$T_{INJ} = g(1 - (u(1-r))^p) \qquad (4)$$

since a new cell must be ready (g), and its injection is not possible only if all input slots are through-going.

The average number of absorbed cells per slot T_{ABS} can be obtained as follows. Let $A_1, .., A_p$ be indicator functions, taking value 1 if absorption has taken place on the corresponding channel. Let N_a be the number of absorbed cells at the present slot. Let p_{A_j} be the probability of absorption from channel j, $j = 1, .., p$. The expected value of A_j is $E[A_j] = p_{A_j}$ so that

$$E[N_a] = T_{ABS} = \sum_{j=1}^{p} E[A_j] = pura \qquad (5)$$

D.3 Resolution Method

By "hop" we mean the visit of the *test* cell to an active node. In a unidirectional ring where cells at destination are always absorbed, the average number of hops H_0 in uniform

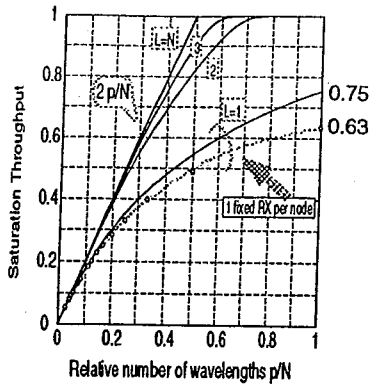

Fig. 4. Saturation throughput versus normalized number of wavelengths. N large. Unequal ring length for the different wavelengths.

traffic (where nodes do not transmit to themselves) is $\frac{N}{2}$. When a node has less receivers than input channels, there is a finite probability $P_m = 1 - a$ of missing a cell. In this case the average number of hops is:

$$H = H_0 + P_m N + P_m^2 N + \cdots = \frac{N(2-a)}{2a}. \quad (6)$$

Given H, Little's law gives: $T = \frac{pu}{H}$. At steady state $T = T_{ABS} = T_{INJ}$, so that from (5) one gets

$$H = \frac{1}{ra}. \quad (7)$$

Using (6) and (7) it is possible to express r as a function of a and viceversa:

$$\begin{cases} r = \frac{2}{N(2-a)} \\ a = 2(1 - \frac{1}{rN}) \end{cases} \quad (8)$$

Equating the second of (8) to (2), a nonlinear equation $r = F(r,u)$ is obtained, where

$$F \triangleq \frac{2}{N} - \sum_{k=L}^{p-1} \left(1 - \frac{L}{k+1}\right) \binom{p-1}{k} (ur)^k r (1-ur)^{p-1-k} \quad (9)$$

Once $r(u)$ is known, going backwards all other parameters are found. Using (5) and (7) we get:

$$T_{ABS} = pu2(r - \frac{1}{N}) \quad (10)$$

so that, equating $T_{INJ} = T_{ABS}$ and using (4) we get:

$$g = \frac{pu2(r - \frac{1}{N})}{1 - (u(1-r))^p}. \quad (11)$$

Note that the offered load $0 \leq g \leq 1$ is indeed the free parameter, while u is an internal variable.

D. Saturation Throughput Comparisons

Fig. 4 summarizes the saturation throughput curves for all the cases considered above, for a ring differing in length among the various channels, and large number of nodes N. The dependence on N can be removed by plotting T_{sat} versus the normalized number of channels p/N.

Consider first a speedup $L = 1$. The networks in sections III-A and III-B refer to the point $p/N = 1$, and both give a value $T_{sat} = 0.63$ (0.58 for equal-length ring). This means that when the added buffering provided by space reuse is not utilized, WDM ring networks perform as well as WDM star networks.

Consider the case $p \leq N, L = 1$. The solid line shows the result for a tunable receiver obtained as in section III-D. The value of T_{sat} at $p/N = 1$ is 0.75, while in the equal-length ring correlations in the drum decrease it to 0.69. Clearly, space reuse gives here the ring a larger T_{sat} than that of star networks. This larger throughput is paied by a larger delay in the drum. The dashed line shows formula (1) for fixed receivers, and is plotted for integer values $m = 1, 2....$ of the number of nodes sharing a wavelength. We can observe that when the number of wavelengths is less than a third of the number of nodes, $p/N < 0.33$, there is no advantage in using tunable receivers. In other words, when the ratio of receivers to channels is more than 3, tunability is not needed.

Fig. 4 also shows the advantage of using more than one tunable receiver, $L > 1$. Clearly, using a speedup larger than 2 does not increase the saturation throughput significantly, especially for $p/N < 0.2$. The limit curve for $L = p$ is the straight line $T = p/(N/2)$, readily obtained by Little's law, considering that no cell is ever missed (hence $H = N/2$) and the load u can be shown to converge to one as $N \to \infty$.

The crucial point here is that the maximum slope of T_{sat} vs p/N equals 2, and such value is due to the multihop ring topology and the uniform traffic assumption (which implies $H \leq N/2$). If the distribution of the destinations is more localized, larger values of the slope can be obtained [19]. Again, T_{sat} is seen to depend on the hardware (the topology) and on the traffic.

Since for typical LANs we are interested in networks in which the ratio p/N is very small, we see that the throughput may be too small for satisfactory performance. If the traffic remains uniform, two ways are available to increase the slope: 1) changing the queuing/injection policy (software), for example using fairness enforcing techniques [16][19], or 2) enriching the topology to reduce the number of hops. In the following we will pursue this second way. The theoretical justification of this way of proceeding comes from a generalization of Little's relation: $T = \frac{pu}{H}$. It can be shown [20] that such relation is valid for slotted lossless networks (i.e. where cells are never lost for buffer overflow within the network) in which not all switching nodes are necessarily *active* (sources/sinks of traffic), provided a *hop* is interpreted as a visit to an active node, and u is the link utilization at the active nodes of in-degree p. Visits to routing switches are not counted in H.

IV. WDM Chordalring

When the number of available wavelengths is a small fraction of the number of nodes, more richly interconnected topologies are necessary to provide a desired saturation throughput to every node.

To gradually enrich the topology starting from the WDM slotted ring, we chose to place a number of 2x2 unbuffered WDM dynamic routers (passive routers, each being a stack

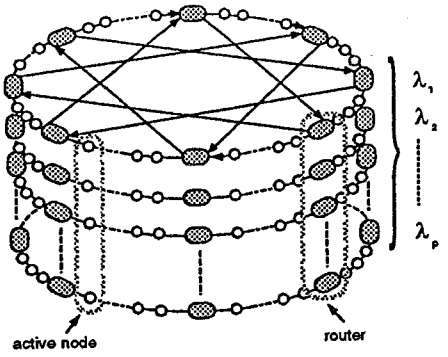

Fig. 5. Logical structure of the WDM Chordalring

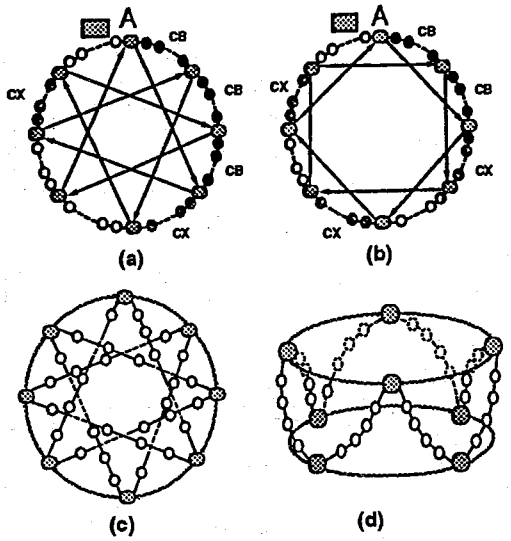

Fig. 6. Chordalring with $R = 8$ routers and (a) K=3, forming 1 group; (b) K=2, forming 2 groups. Preferred switch settings for a cell at router A are illustrated. CB= care bar; CX = care cross. Figs. (c) and (d) are alternative views of (a) and (b).

of independently operated 2x2 switches, one per wavelength) along the ring, and to connect the routers with chords (or shortcuts) that decrease the average number of hops H (or active node crossings) and thus increase throughput. The logical structure of the new network, which we call WDM Chordalring, is shown in Fig. 5.

Since routers are unbuffered, cells may either be routed on a preferred path or deflected [21]. The routing problem is quite different from known work on deflection routing, and new routing strategies must be used.

The aim is to have a smooth topological transition from the ring to the more complex chordal ring, by adding as few routers as needed to achieve the required saturation throughput.

The proposed topology is a variation on chordal rings [1], and is similar to the SMARTnet proposed by I. Rubin [2], although our routers are 2x2, links are unidirectional, and the routing is datagram as opposed to virtual circuit.

A. Routing and Saturation Throughput

We are given an N node WDM slotted ring. We insert R routers in the ring, so that $Q = N/R$ active nodes are present in each arc between consecutive routers. Let K be the number of routers bypassed by each shortcut, including the router reached by the shortcut. Fig. 6(a) shows an example with $R = 8$ and $K = 3$, so that, following a path of shortcuts only, all routers can be reached, i.e. routers belong to a single group. In Fig. 6(b) $K = 2$, and there are two groups of routers. In general, given R, by varying K we obtain a varying number of groups.

Figs. 6 (c) and (d) show an alternative representation of Fig. 6 (a) and (b), in which the shortcut groups form one-way "beltways" connecting one-way lanes of active nodes. Such an alternative representation is more reminiscent of related car traffic problems in urban areas.

If the final goal is to reduce the number of active node crossings H as much as possible, clearly the ideal routing would always follow the shortcuts up to the arc containing the destination node.

Unfortunately, if all cells follow this strategy, plenty of contentions for the beltway occur at the routers. Since buffers are not provided, plenty of cells will be deflected away from the beltway, each time paying a price of Q extra *hops*.

An adaptive routing algorithm makes sense here, preferably routing cells along the beltways at low traffic loads, while choosing on purpose the active-node lanes when close enough to destination at high loads.

Since we were shooting for a routing algorithm maximizing the throughput at full load, we decided to adopt the following load-independent routing strategy, illustrated with an example in Fig. 6(a). Consider a cell reaching router A from the ring. If its destination node is within the arc subtended by the shortcut departing from A, then the cell preferred route is along the ring, and switch A should be set in bar, i.e., the cell is said to be *care bar* (CB). Else, if the destination is within the first arc following each router in the group of A up to but excluding A along the ring, the switch should be set in cross, and the cell is *care cross* (CX). If the cell reaches A from the shortcut, just reverse the switch settings. If the cell has any other destination, it is treated as *don't care*, i.e., there is no preferred switch setting. Having defined the cell preferred settings, if no contention occurs at the switch, the cell actually follows the preferred route; else a coin is flipped to solve the contention and only the winning cell follows the preferred path.

The above algorithm has been found by simulation to give the best throughput/delay results at full load, after testing many other load-independent variations [22]. An accurate analytical model to derive throughput/delay for the specific topological choice $K = R/2$ has been developed in [23], and extended in [22], where the parameter K was optimized for the above routing. It was found in [22] that the optimal K value that minimizes H at high load is between $R/3$ and $R/4$.

Fig. 7 gives simulation results of saturation throughput versus normalized channel number p/N in uniform traffic for WDM chordalrings with unequal link lengths of average value $W = 5$ slots [1], $N = 64$ nodes, optimized value of K, for $L = 1$ tunable receiver per node. The number of routers in the ring R is a parameter, and $R = 0$ corresponds

[1] To eliminate any residual correlation due to routing and topology, as done for the ring. Otherwise the value of W has no effect on throughput and average number of hops.

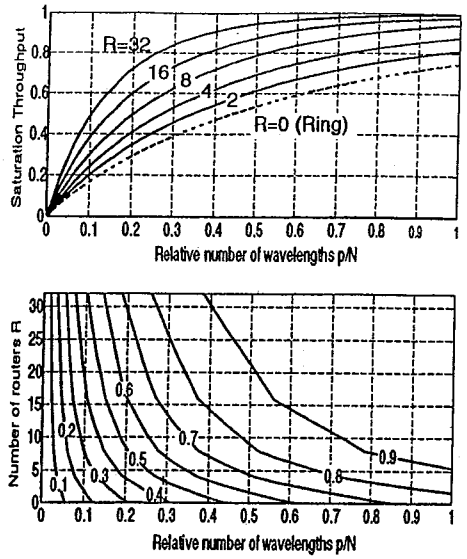

Fig. 7. Top: Saturation Throughput T_{sat} vs channel fraction p/N for a WDM Chordalring with optimized K, $N = 64$, unequal links of average length $W = 5$ and speedup $L = 1$. Bottom: contour plot of T_{sat} on the plane R, p/N.

to the standard slotted ring with unequal channel lengths. The effect of gradually adding routers on the slope of the T_{sat} vs p/N curve can be observed. For N sufficiently large and R within the range shown, such curves are essentially independent of N.

Slicing the T_{sat} vs p/N graph at constant T_{sat} produces the contour plot at the bottom of the figure. From such contour the trade-off between routers and wavelengths in the WDM Chordalring with deflection is clearly understood: when the number of channels p/N is small, adding routers has the greatest impact on throughput. Else, when p/N is large, adding routers does not help improving the throughput, since blocking (deflections) is already low because of the low load per channel. With more than one receiver per node ($L > 1$) the benefits of adding routers is reduced to smaller values of p/N, as expected.

V. Conclusions

This paper has reviewed the basic theory of cell switching for datagram networks, and found instructive analogies with optical implementations, both for star and ring multiwavelength networks. It has been shown that the buffering capabilities of high-speed slotted networks, also known as space-reuse, allows multihop networks such as WDM rings to have larger throughputs than star networks, where space reuse is not possible. WDM ring networks have been extensively analyzed in uniform traffic, and many schemes have been compared. The multihop nature of the network, and the clear identification of the meaning of hop as it relates to throughput have helped designing a novel topology, which has been called WDM Chordalring, which smoothly scales the throughput of a WDM ring by progressively adding WDM unbuffered routers up to reaching the desired saturation throughput. Routing without buffers has been studied in the novel topology by simulation, and an empirical routing scheme has been identified that gives satisfactory performance. Using such routing, the topology has been optimized and a clear trade-off between number of wavelengths, number of receivers per node and number of routers has been identified and quantified. We believe such network can be interesting for high-speed datagram applications, and its smooth scalability properties are believed to be of interest for local/metropolitan area computer/data communications applications.

Acknowledgments

This work was supported partly by the European Community under INCO-DC project No. 950959 "DAWRON", and partly by a grant from CSELT.

References

[1] B. W. Arden and H. Lee, "Analysis of chordal ring network," *IEEE Trans. Comput.*, vol. c-30, pp. 291–295, Apr. 1981.

[2] I. Rubin, and H.-k. Hua, "An all-optical wavelengths-division meshed-ring packet-switching network," in *Proc. IEEE INFOCOM '95*, Boston, MA, pp. 969–976, Apr. 1995.

[3] M. J. Karol, M. G. Hluchyj, and S. P. Morgan, "Input versus output queueing on a space-division packet switch," *IEEE Trans. Commun.*, vol. COM-35, pp. 1347–1356, Dec. 1987.

[4] Y. Oie, M. Murata, K. Kubota, and Hideo Miyahara, "Performance analysis of nonblocking packet switch with input and output buffers," *IEEE Trans. Commun.*, vol. 40, pp. 1294–1297, Aug. 1992.

[5] A. Acampora, *An introduction to broadband networks.* Plenum, 1994.

[6] N. McKeown, V. Anantharam, and J. Walrand, "Achieving 100% throughput in an input-queued switch," in *Proc. IEEE Infocom'96*, San francisco, CA, vol. 1, pp. 296–302, Mar. 1996.

[7] J. Y. Hui, *Switching and traffic theory for integrated broadband networks.* Kluwer, 1990.

[8] M. Schwartz, *Broadband integrated networks.* Prentice Hall, 1996. Chapter 5.

[9] Y.-S. Yeh, M. G. Hluchyj, and A. S. Acampora, "The knockout switch: a simple, modular architecture for high-performance packet switching," *IEEE J. Selected Areas Commun.*, vol. 5, pp. 1204–1283, 1987.

[10] H. S. Hinton, *An introduction to photonic switching fabrics.* Plenum, 1993. Chapter 3.

[11] M.-S. Chen, N. R. Dono, and R. Ramaswami, "A Media Access protocol for packet-switched WDM metropolitan area networks," *IEEE J. Selected Areas Commun.*, vol. 8, pp. 1048–1057, Aug. 1990.

[12] I. Chlamtac, and A. Fumagalli, "Quadro Star: a high performance optical WDM star network," *IEEE Trans. Commun.*, vol. 42, pp. 2582–2591, Aug. 1994.

[13] K. Sasayama et al., "Demonstration of a photonic ATM switch using a frequency-routing-type time-division interconnection network (FRONTIER-NET)," in *Proc. ECOC'94*, pp. 533–536, 1994.

[14] B. Glance, "Protection-against-collision optical packet network," *IEEE J. Lightwave Technol.*, vol. 10, pp. 1323–1328, Sep. 1992.

[15] M. J. Karol, and B. Glance, "Performance of the PAC optical packet network," *IEEE J. Lightwave Technol.*, vol. , pp. 1394–1399, Aug. 1993.

[16] M. Ajmone Marsan, A. Bianco, E. Leonardi, M Meo, and F. Neri, "MAC protocols and fairness control in WDM multirings with tunable transmitters and fixed receivers," *IEEE J. Lightwave Technol.*, vol. 14, pp. 1230–1244, June 1996.

[17] E. L. Goldstein, A. F. Elrefaie, N. Jackman, and S. Zai, "Multiwavelength fiber-amplifier cascade unidirectional interoffice ring networks," in *Proc OFC'93*, pp. 44–46, San Jose, CA, Feb. 1993.

[18] L. Eskildsen, E. L. Goldstein, G. K. Chang, M. Z. Iqbal, and C. Lin, "Self-regulating WDM amplifier module for scalable lightwave networks," *IEEE Photon. Technol. Lett.*, vol. 6, pp. 1321–1323, Nov. 1994.

[19] I. Cidon, and Y. Ofek, "Metaring - A full duplex ring with fairness and spatial reuse," *IEEE Trans. Commun.* vol. 41, pp. 110–120, Jan. 1993.

[20] A. Bononi, "Weakly vs strongly multihop space-division optical networks," *IEEE J. Lightwave Technol.*, vol. 16, Apr. 1998.

[21] P. Baran, "On distributed communications networks," *IEEE Trans. Commun. Syst.* vol. 12, pp. 1–9, Mar. 1964.

[22] L. Mora, "Analisi di traffico in reti ottiche WDM a pacchetto in topologie ad anello generalizzato." Laurea Thesis, Universita' di Parma, March 1997.

[23] S. Savi, "Analisi di traffico in una rete mutifrequenza ad anello generalizzato con instradamento a deflessione." Laurea Thesis, Universita' di Parma, July 1996.