A Geometric Approach to Capacity Provisioning in WDM Networks with Dynamic Traffic

Li-Wei Chen, Eytan Modiano

***Abstract*— In this paper, we use an asymptotic analysis similar to the sphere-packing argument in the proof of Shannon's channel capacity theorem to derive optimal provisioning requirements for networks with both static and dynamic provisioning. We consider an N-user shared-link model where W_s wavelengths are statically assigned to each user, and a common pool of W_d wavelengths are available to all users. We derive the minimum values of W_s and W_d required to achieve asymptotically non-blocking performance as the number of users N becomes large. We show that it is always optimal to statically provision at least enough wavelengths to support the mean of the traffic. We then consider allowing the shared wavelengths W_d to be switched in groups (or *wavebands*) rather than on an individual basis, and show that by employing waveband switching, a link with only a few switches per user can achieve the same performance as a link provisioned with unlimited switches per user using only marginally more wavelengths. We also derive the optimal band size and wavelengths required. Finally, we discuss adaptation of these results to the case of a finite and small number of users.**

I. Introduction

OPTICAL networks are a common solution for high-speed communications. In general, an optical network can consist of nodes connected in arbitrary fashion via many optical fibers (Figure 1). In this paper, we will focus on provisioning a single link in a backbone network. Such a link is shared by traffic between many source-destination pairs in the larger network. Each wavelength on the link can be used to support one lightpath from one of the incoming fibers on the left side of the link to one of the outgoing fibers on the right side of the link.

Wavelength provisioning can be done either statically, by dedicating a wavelength solely to supporting calls from a single source-destination pair along a fixed path in the network, or dynamically, by allowing a wavelength to be switched dynamically (either individually or as part of a bundle of multiple wavelengths called a *waveband*) to serve different source-destination pairs according to traffic demand. These two approaches trade off low cost and simplicity against efficiency of resource usage, respectively.

There has been much investigation of both statically provisioned and dynamically provisioned systems in the literature [1], [2], [3], [4]. Such approaches are well-suited for cases where either the traffic is known a priori and can be statically provisioned, or is extremely unpredictable and needs to be dynamically provisioned. However, in practice, a middle ground is usually more common. Traffic reaching high-bandwidth optical networks typically consists of an amalgamation of a large number of smaller calls, and statistical multiplexing helps reduce the variance of the traffic. As a result, it is common to see traffic demands characterized by a large mean and a small variance around the mean. A hybrid system is well suited to such a scenario. In a hybrid system, a sufficient number of wavelengths are statically provisioned to support the majority of the traffic. Then, on top of this, a smaller number of wavelengths are dynamically provisioned to support the inevitable variation in the realized traffic. Such an approach takes advantage of the relative predicability of the traffic by cheaply provisioning the majority of the wavelengths, but retains sufficient flexibility through the minority of dynamic wavelengths that significant wavelength overprovisioning is not necessary.

The authors are with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (email: lwchen@mit.edu, modiano@mit.edu)

This work was supported by the National Science Foundation (NSF) under Grant Grants ANI-0073730 and ANI-0035217.

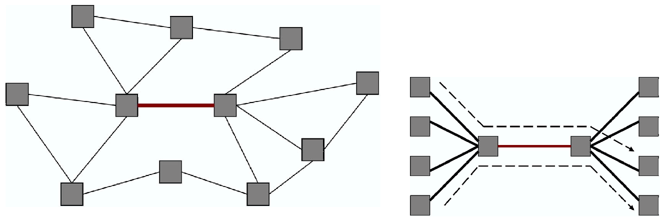

Fig. 1. An example of a mesh optical network consisting of numerous nodes and links, followed by a shared-link model based on the colored link. The dotted lines denote different users of the link. Since each pair of input-output fibers comprises a different user, and there are 4 input fibers and 4 output fibers, there are a total of $4 \cdot 4 = 16$ users in this example.

We will first consider the provisioning of hybrid systems where each dynamic wavelength is individually switched. We use a sphere hardening approach [5]: we allow the number of users to become large, and consider the minimum provisioning in static and dynamic wavelengths necessary to achieve non-blocking performance (i.e., to guarantee that the probability of any call in the snapshot being blocked goes to zero). We will show that it is always optimal to statically provision enough wavelengths to support the traffic mean. We also fully characterize the optimal provisioning strategy for achieving non-blocking performance with minimal wavelength provisioning.

We next consider waveband-switched systems, and again derive a strategy for optimal provisioning. Subsequently, for a fixed number of switches, we determine the appropriate static and dynamic partitioning that results in the minimum number of total wavelengths used. We will show that networks can be nearly optimal in the total number of wavelengths required with very few switches per user if the network is allowed to switch using wavebands. The analysis also provides insight into the proper sizing of wavebands and the amount of static provisioning required in hybrid waveband-switched systems.

A. System Model

In the shared link context, we can consider each incoming-outgoing pair of fibers to be a different *user* of the link. Each lightpath request (which we will henceforth term a *call*) can therefore be thought of as belonging to the user corresponding to the incoming-outgoing fiber pair that it uses. We can similarly associate each static wavelength with the corresponding user. Under these definitions, a call belonging to a given user cannot use a static wavelength belonging to a different user – it must either use a static wavelength belonging to its own user, or employ a dynamic wavelength.

When a user requests a new call setup, the link checks to see if a static wavelength for that user is free. If there is a free static wavelength, it is used. If not, then the link checks to see if any of the shared dynamic wavelengths are free – if so, then a dynamic wavelength is used. If not, then no resources are available to support the call, and it is blocked. There have been several approaches developed in the literature for blocking probability analysis of such systems under Poisson traffic models [6], including the Equivalent Random Traffic (ERT) model [7], [8] and the Hayward approximation [9]. These approximations, while often able to produce good numerical approximations of blocking probability, are purely numerical in nature and do not provide good intuition for guiding the dimensioning of the wavelengths. Furthermore, they assume that the dynamic wavelengths must be individually switched, and do not consider waveband switching.

In this paper, we adopt a snapshot traffic model that leads to closed-form asymptotic analysis and develop guidelines for efficient dimensioning of hybrid networks. We consider examining a "snapshot" of the traffic demand at some instant in time. The snapshot is composed of the vector $\mathbf{c} = [c_1, \ldots, c_N]$, where c_i is the number of calls that user i has at the instant of the snapshot, and N is the total number of users.

We model each variable c_i as a Gaussian random variable with mean μ_i and variance σ_i^2. This is reasonable since each "user" actually consists of a collection of source-destination pairs in the larger network that all use the link from the same source fiber to the same destination fiber. In this paper, we will assume that each user has the same mean μ and variance σ^2; the results are extensible to general μ_i and σ_i but the extension is beyond the scope of this paper (see [10]). Although the traffic for each individual source-destination pair for the user may have some arbitrary distribution, as long as the distributions are well-behaved, the sum of each traffic stream will appear Gaussian by the Central Limit Theorem.

II. Wavelength-Granularity Switching

In this section, we consider a shared link, and assume that there are N users that are the source of calls on the link. Each user is statically provisioned W_s wavelengths for use **exclusively** by that user. In addition to this static provisioning, we will also provide a total of W_d dynamically switched wavelengths. These wavelengths can be shared by any of the N users.

As previously described, we will use a snapshot model of traffic. The traffic is given by a vector $\mathbf{c} = [c_1, \ldots, c_N]$, where each c_i is independent and identically distributed as $\mathsf{N}(\mu, \sigma^2)$. We assume that the mean μ is significantly large relative to σ that the probability of "negative traffic" (a physical impossibility) is low, and therefore does not present a significant modeling concern. We will primarily be concerned with a special blocking event that we call *overflow*. An overflow event occurs when there are insufficient resources to support all calls in the snapshot and at least one call is blocked. We will call the probability of this event the *overflow probability*. An overflow event occurs if the total number of calls exceeds the ability of the static and dynamic wavelengths to support. This can be expressed mathematically as

$$\sum_{i=1}^{N} \max\{c_i - W_s \ , \ 0\} > W_d \tag{1}$$

where $\max\{c_i - W_s \ , \ 0\}$ is the amount of traffic from each user that exceeds the static provisioning; if the total amount of excess from each user exceeds the available pool of shared dynamic wavelengths, a blocking event occurs.

If we consider the N-dimensional vector space occupied by $\mathbf{c}$, the constraint given by (1) represents a collection of hyperplanes bounding the admissible traffic region:

$$\begin{array}{rcll} c_i & \leq & W_s + W_d & \\ c_i + c_j & \leq & 2W_s + W_d & , \quad i \neq j \\ c_i + c_j + c_k & \leq & 3W_s + W_d & , \quad i \neq j \neq k \\ & \vdots & & \end{array}$$

Each constraint reflect the fact that the sum of the traffic from any subset of users clearly cannot exceed the sum of the static provisioning for those users plus the entire dynamic provisioning available. Note that there are a total of N sets of constraints, where the n^{th} set consists of $C(N,n) = \frac{N!}{(N-i)!n!}$ equations, each involving the sum of n elements of the traffic vector $\mathbf{c}$. If the traffic snapshot $\mathbf{c}$ falls within the region defined by the hyperplanes, all calls are admissible; otherwise, an overflow event occurs. The bolded lines in Figure 2 show the admissible region for $N = 2$ in two dimensions.

A. Asymptotic Analysis

We will consider the case where the number of users N becomes large, and use the law of large numbers to help us draw some conclusions. We can rewrite the call vector in the form

$$\mathbf{c} \quad = \quad \mu \cdot \mathbf{1} + \mathbf{c}'$$

where $\mathbf{1}$ is the length-N all-ones vector, and $\mathbf{c}' \sim \mathbf{N}(\mathbf{0}, \sigma^2\mathbf{1})$ is a zero-mean Gaussian random vector with i.i.d. components. Conceptually, we can visualize the random traffic vector as a random vector $\mathbf{c}'$ centered at $\mu\mathbf{1}$. The length of this random vector is given by

$$\|\mathbf{c}'\| \quad = \quad \sqrt{\sum_{n=1}^{N} c_i^2}$$

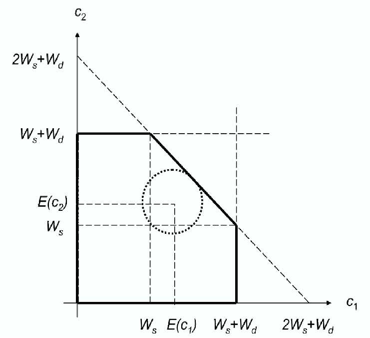

Fig. 2. The admissible traffic region, in two dimensions, for $N = 2$. Three lines form the boundary constraints represented by (1). There are two lines each associated with a single element of the call vector $\mathbf{c}$, and one line associated with both elements of $\mathbf{c}$. The traffic sphere must be entirely contained within this admissible region for the link to be asymptotically non-blocking.

We will use an approach very similar to the sphere packing argument used in the proof of Shannon's channel capacity theorem in information theory [5]. We will show that asymptotically as the number of users becomes large, the traffic vector falls onto a sphere centered at the mean, and the provisioning becomes a problem of choosing the appropriate number of static and dynamic wavelengths so that this traffic sphere is completely contained within the admissible region. From the law of large numbers, we know that

$$\frac{1}{N}\sum_{n=1}^{N} c_i^2 \to \sigma^2$$

as $N \to \infty$. This implies that asymptotically, as the number of users becomes large, the call vector $\mathbf{c}$ becomes concentrated on a sphere of radius $\sqrt{N}\sigma$ centered at the mean $\mu\mathbf{1}$. Therefore, in order for the overflow probability to converge to zero, a necessary and sufficient condition is that the hyperplanes described by (1) enclose the sphere entirely (Figure 2).

B. *Minimum Distance Constraints*

Next, we will derive necessary and sufficient conditions for the admissible traffic region to enclose the traffic sphere. Our goal is to ensure that we provision W_s and W_d such that the minimum distance from the center of the traffic sphere to the boundary of the admissible region is at least the radius of the sphere, therefore ensuring that all the traffic will fall within the admissible region.

Due to the identical distribution of the traffic for each user, the mean point $\mu\mathbf{1}$ will be equidistant from all planes whose description involves the same number of elements of $\mathbf{c}$. We define a *distance function* $f(n)$ such that $f(n)$ is the minimum distance from the mean $\mu\mathbf{1}$ to any hyperplane whose description involves n components of $\mathbf{c}$.

Lemma 1: The distance function $f(n)$ from the traffic mean to a hyperplane involving n elements of the traffic vector $\mathbf{c}$ is given by

$$f(n) = \sqrt{n}\left(W_s + \frac{W_d}{n} - \mu\right), \quad n = 1, \ldots, N \qquad (2)$$

Proof: The geometric proof is essentially a simplified version of the proof of Lemma 2 and omitted for brevity. ■

We define the *minimum boundary distance* to be

$$F_{min} = \min_{n=1,\ldots,N} f(n)$$

A necessary and sufficient condition for the overflow probability to go to zero asymptotically with the number of users is

$$F_{min} \geq \sqrt{N}\sigma$$

We would like to determine the index n such that $f(n)$ is minimized. Unfortunately, this value of n turns out to depend on the choice of provisioning W_s. Let us consider the derivative of the distance function $f'(n)$:

$$f'(n) = \frac{1}{2\sqrt{n}}\left(W_s - \frac{W_d}{n} - \mu\right)$$

We can divide W_s into three regimes of interest, corresponding to different ranges of values for W_s and W_d, and characterize $f(n)$ in each of these regions:

Regime 1: If $W_s \leq \mu$:

In this region, $f'(n) < 0$ for all n. This implies that $f(n)$ is a decreasing function of n, and $F_{min} = f(N)$, giving a minimum distance of

$$F_{min} = \sqrt{N}\left(W_s + \frac{W_d}{N} - \mu\right)$$

Regime 2: If $\mu < W_s \leq \mu + W_d$:

In this region, $f'(n)$ starts out negative and ends up positive over $1 \leq n \leq N$. This implies that $f(n)$ is convex and has a minimum. Neglecting integrality concerns, this minimum occurs when $f'(n) = 0$, or

$$n^* = \frac{W_d}{W_s - \mu}$$

Therefore $F_{min} = f(n^*)$ in this regime. Substituting the appropriate values, it can be shown that the minimum distance is given by

$$F_{min} = 2\sqrt{W_d(W_s - \mu)}$$

Regime 3: If $W_s > \mu + W_d$:

In this region, $f'(n) > 0$ for all n. This implies that $f(n)$ is an increasing function of n, and $F_{min} = f(1)$, giving a minimum distance of

$$F_{min} = W_s + W_d - \mu$$

C. *Optimal Provisioning*

In the preceding section, we derived the minimum distance criteria for the hybrid system. Given a fixed number of statically allocated wavelengths W_s, we can use the equation $F_{min} \geq \sqrt{N}\sigma$ to calculate the minimum number of dynamic wavelengths W_d to achieve asymptotically non-overflow performance. We can also draw a few additional conclusions about provisioning hybrid systems.

Theorem 1: A minimum of μ static wavelengths should always be provisioned per user.

Proof: For $W_s \leq \mu$, we know from Case 1 above that the minimum distance constraint is

$$F_{min} = \sqrt{N}\left(W_s + \frac{W_d}{N} - \mu\right) \quad \geq \quad \sqrt{N}\sigma$$

$$\Rightarrow W_{tot} = NW_s + W_d \quad \geq \quad (\mu+\sigma)N$$

Note that the total number of wavelengths $W_{tot} = NW_s + W_d$ is independent of W_s and W_d in this regime, suggesting that the same total number of wavelengths are required regardless of the partitioning between static and dynamic wavelengths. Since static wavelengths are less expensive to provision than dynamic wavelengths, this shows that there is never any reason to provision less than $W_s = \mu$ wavelengths. ■

An interesting corollary to this theorem follows from the observation that the case where $W_s = 0$ (i.e. all wavelengths are dynamic) also falls in this regime (i.e. Regime 1). Since fully dynamic provisioning is obviously the least-constrained version of this system, we can use it as a bound on the minimum number of wavelengths required by **any** asymptotically overflow-free system.

Corollary: For non-overflow operation, a lower bound on the number of wavelengths required is given by

$$W_{tot} \geq (\mu+\sigma)N$$

We can also consider a system that is fully static, with no dynamic provisioning. This is the most inflexible wavelength partitioning, and provides us with an upper bound on the number of wavelengths required by any hybrid system.

Theorem 2: For a fully static system with no dynamic provisioning, the minimum number of wavelengths required is given by

$$W_{tot} = (\mu+\sigma)N + \left(\sqrt{N}-1\right)N\sigma$$

Proof: Let $W_d = 0$. Then, for overflow-free operation, we obviously need $W_s > \mu$. This puts us in Regime 3 where $W_s > \mu + W_d$, and the minimum distance condition gives us

$$F_{min} = W_s + W_d - \mu \quad > \quad \sqrt{N}\sigma$$
$$W_{tot} = NW_s \quad = \quad (\mu+\sigma)N + \left(\sqrt{N}-1\right)N\sigma$$

■

Note that this exceeds the lower bound on the minimum number of wavelengths by $(\sqrt{N}-1)N\sigma$. We can therefore regard this quantity as the **maximum switching gain** that we can achieve in the hybrid system. This gain is measured in the maximum number of wavelengths that could be saved if all wavelengths were dynamically switched. Combining the upper and lower bounds, we can make the following observation:

Corollary: For efficient overflow-free operation, the total number of wavelengths required by any hybrid system is bounded by

$$(\mu+\sigma)N \leq W_{tot} \leq (\mu+\sigma)N + (\sqrt{N}-1)N\sigma$$

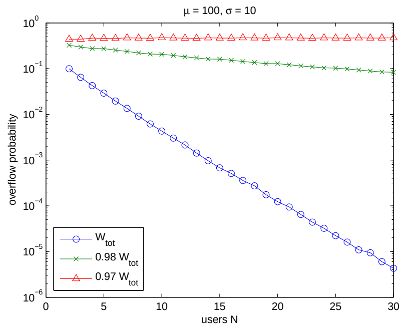

Fig. 3. Curves show decrease in overflow probability with increasing number of users N. The mean has been statically provisioned. The curve with the circles shows a link provisioned with the theoretical minimum number of wavelengths W_{tot} needed to achieve asymptotically non-overflowing operation. Note that if fewer than W_{tot} wavelengths are provisioned, the overflow probability no longer converges to zero as the number of users increases.

D. Numerical Example

Simulations were conducted to verify the accuracy of the provisioning results derived. Figure 3 verifies the results of the preceding discussion for the case of $\mu = 100$ and $\sigma = 10$. The rapidly descending curve shows that if the theoretical minimum of $W_{tot} = (\mu+\sigma)N$ wavelengths are provisioned, then as the number of users N increases, the overflow probability drops off quickly and eventually the system becomes asymptotically non-blocking. The other two curves show that if less than W_{tot} wavelengths are provisioned, the overflow probability no longer converges to zero as the number of users increases.

Note also that the convergence occurs fairly rapidly if the W_{tot} wavelengths calculated in the preceding sections are provisioned. In a system with just 30 users, the overflow probability has already decreased to the order of 10^{-5}. Since the number of users is equal to the number of input-output fiber pairs, this corresponds to a link with as few as 5 input fibers and 6 output fibers, for example. Therefore, the results are useful in designing for good network performance even when N is finite and small.

III. Waveband-Granularity Switching

A. Asymptotic Analysis

In this section, we also consider a shared link, but now we allow for waveband switching. Again, each user is statically provisioned W_s wavelengths for its own use. Additionally, we assume there are b wavebands, each of size W_b. Each waveband can be assigned to serve calls from any user, but all W_b wavelengths within the same waveband must serve the same user.

This banded approach is interesting because, compared to wavelength switching, it allows for more wavelengths to be dynamically allocated, at the cost of coarser switching granularity. For example, given the same number of switches, a waveband approach with bands of $W_b = 2$ wavelengths will have twice the total number of dynamic wavelengths as a wavelength-switched network. However, these dynamic wavelengths are less flexible in the banded case, since two dynamic wavelengths must be assigned to a user at a time.

Note that the wavelength-switched network is a special case of the waveband-switched network where $W_b = 1$. In such a

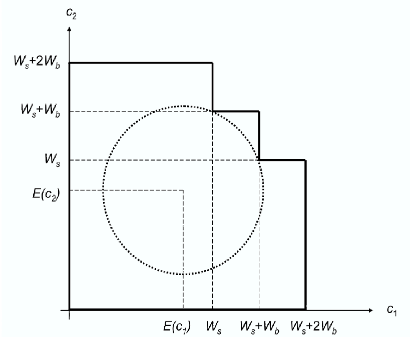

Fig. 4. The admissible region for a link with $N = 2$ users. The traffic sphere must be entirely enclosed within the admissible region in order for the link to be asymptotically non-blocking.

case, $W_d = b \cdot W_b = b$. In this section, we will analyze the performance of banded networks in general, and make some comparisons about the performance improvement gained over wavelength-switched networks by allowing waveband switching.

Again, we assume that the traffic vector $\mathbf{c}$ is composed of normally distributed i.i.d. entries with mean μ and variance σ^2. An overflow event occurs if there exists at least one call that is blocked due to insufficient wavelengths being available to service it. This can be written mathematically as

$$\sum_{i=1}^{N} \max \left\{ \left\lceil \frac{c_i - W_s}{W_b} \right\rceil \ , 0 \right\} > b \tag{3}$$

(3) can be written as N sets of boundary constraints, where the n^{th} set consists of $C(N,n) = \frac{N!}{(N-i)!\ n!}$ equations, each involving n elements of the traffic vector $\mathbf{c}$:

$$\begin{aligned} \left\lceil \frac{c_i - W_s}{W_b} \right\rceil &\leq b \quad , \quad \forall\, i \\ \left\lceil \frac{c_i - W_s}{W_b} \right\rceil + \left\lceil \frac{c_j - W_s}{W_b} \right\rceil &\leq b \quad , \quad \forall\, i \neq j \\ &\vdots \end{aligned}$$

Figure 4 illustrates an example of the admissible traffic region for $N = 2$ users.

We will next determine the distance from the mean of the traffic vector to each boundary, and derive a sufficient mathematical condition for the traffic vector to be admissible (i.e. not in overflow). Consider the n^{th} set of boundary constraints, and suppose that the first n elements of $\mathbf{c}$ are active. Then we require

$$\sum_{i=1}^{n} \left\lceil \frac{c_i - W_s}{W_b} \right\rceil \leq b \tag{4}$$

We observe that since all the terms in the summation are integers, and b is an integer, (4) holds if and only if

$$\frac{c_1 - W_s}{W_b} + \sum_{i=2}^{n} \left\lceil \frac{c_i - W_s}{W_b} \right\rceil \leq b \tag{5}$$

Therefore we can equivalently consider provisioning W_s and W_b to satisfy (5), and (4) will follow. Using the inequality $\lceil x \rceil < x + 1$, we observe that

$$\sum_{i=2}^{n} \left\lceil \frac{c_i - W_s}{W_b} \right\rceil \leq \sum_{i=2}^{n} \left(\frac{c_i - W_s}{W_b} + 1 \right)$$

from which we can conclude that (5) holds if we choose W_s and W_b such that

$$\frac{c_1 - W_s}{W_b} + \sum_{i=2}^{n} \left(\frac{c_i - W_s}{W_b} + 1 \right) \leq b \tag{6}$$

This is equivalent to

$$\frac{c_1 - W_s}{W_b} + \sum_{i=2}^{n} \frac{c_i - W_s}{W_b} \leq b - (n-1)$$

Rearranging the above, we obtain:

$$\sum_{i=1}^{n} c_i \leq nW_s + (b - (n-1))\, W_b \tag{7}$$

By the above reasoning, we have that (7) → (6) → (5) → (4). Therefore (7) is a sufficient condition for the traffic vector being admissible.

Recall that we derived this expression assuming that the first n elements of $\mathbf{c}$ were active. In general, for n active elements, the sum on the left of (7) will involve the sum of those active elements. We also point out here that by using the upper bound $\lceil x \rceil < x + 1$ as the basis for our provisioning, we have been more conservative in our estimate of the traffic. We therefore expect that this will result in a small amount of overprovisioning of the link.

B. *Minimum Distance Constraints*

We define the minimum distance from the traffic mean to any boundary involving n active constraints to be $f(n)$. This minimum distance expression will later be useful in determining sufficient provisioning for overflow-free operation.

Lemma 2: The distance $f(n)$ from the traffic mean to any hyperplane involving n elements of the traffic vector $\mathbf{c}$ is given by:

$$f(n) = \sqrt{n} \left(W_s + \frac{b - (n-1)}{n} W_b - \mu \right) \tag{8}$$

Proof: For a fixed n, the hyperplane has a normal vector consisting of n unity entries and $N - n$ zero entries. Since by symmetry the mean of the traffic is equidistant from all hyperplanes with the same number of active constraints, without loss of generality, assume that the first n constraints that are active. Then the closest point on the hyperplane has the form

$$[\mu + x, \ldots, \mu + x, \mu, \ldots, \mu]$$

where the first n entries are $\mu + x$, and the remainder are μ. The value of x is obtained applying (7), which requires that

$$\sum_{i=1}^{n}(\mu + x) = nW_s + (b - (n-1))W_b$$

$$\Rightarrow x = W_s + \frac{b-(n-1)}{n}W_b - \mu$$

The distance from the point $[\mu, \ldots, \mu]$ to this point on the hyperplane is $\sqrt{n}\, x$, where, after substituting for x, we obtain

$$f(n) = \sqrt{n}\left(W_s + \frac{b-(n-1)}{n}W_b - \mu\right)$$

■

By the same law of large numbers reasoning as in the previous section, we observe that asymptotically as the number of users N becomes large, the traffic will converge to a sphere of radius $\sqrt{N}\sigma$ centered at the mean. Let the minimum distance from the mean to the closest hyperplanar boundary be $F_{min} = \min_n f(n)$. Then the system will be asymptotically non-blocking if $F_{min} > \sqrt{N}\sigma$.

In the preceding wavelength-switched networks, fixing the number of switches b was equivalent to fixing W_d. This left only 1 free parameter, W_s, and uniquely determined the optimal provisioning for asymptotically non-blocking operation. However, in this section, if b is fixed, *two* parameters remain to be chosen: the number of static wavelengths W_s and the waveband size W_b. Therefore, the choice of these parameters is not unique unless we specify additional optimization criteria.

If in addition we require that the optimal provisioning also minimize the total number of wavelengths $W_{tot} = NW_s + bW_b$ for a fixed b, there will exist a unique choice of W_s and W_b for each b. We now proceed to derive this optimal choice.

We would like to first determine the index n that minimizes $f(n)$. We consider the first derivative of $f(n)$:

$$f'(n) = \frac{1}{2\sqrt{n}}\left[W_s - \left(\frac{b+1}{n} + 1\right)W_b - \mu\right] \quad (9)$$

From (9), we observe that the behavior of $f(n)$ can be characterized in three regimes corresponding to different ranges of values for W_s and W_b. By analysis very similar to Section II-B, the minimum number of wavelengths required in each regime, along with the choice of W_s and W_b that achieves this, is listed in Table I.

C. Optimal Provisioning

For a given number of switches b, Section III-B gives the total number of wavelengths in each of three regimes. To determine the optimal operating regime, these expressions should be evaluated for the particular values of b, μ, and σ, and the regime that requires the fewest total wavelengths W_{tot} should be chosen as the operating regime. The results in Section III-B for that regime will then provide the optimal choice of W_s and W_b to achieve asymptotically non-blocking operation.

Figure 5 plots the dropoff in overflow probability as the number of users N increases. The different curves show the effect of underprovisioning the total number of wavelengths relative

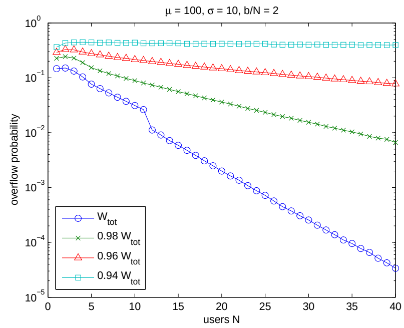

Fig. 5. Dropoff in overflow probability as the number of users N increases. In this example, two switches per user are available. The different curves show the effect of underprovisioning the total number of wavelengths relative to the theoretical minimum. Note that if less than 94% of the calculated wavelengths are provisioned, the overflow probability does not decrease even as the number of users increases.

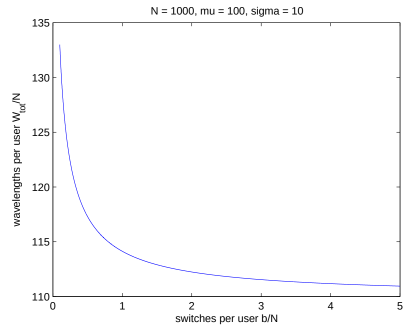

Fig. 6. A plot of the number of wavelengths required as a function of the number of switches. Note that the initial wavelength savings is significant, but the marginal gain in wavelength savings decreases rapidly as the number of switches gets larger. Theorem 1 gives a lower bound on the number of wavelengths required as $\mu + \sigma$ = 110 wavelengths per user in this example.

to the predicted W_{tot}. As expected, W_{tot} slightly overestimates the total number of wavelengths required, but not by much. To see this, observe that if only $0.94W_{tot}$ is provisioned (the curve with the square points), the overflow probability stays constant at a high value even as the number of users increases and does not diminish.

Figure 6 shows an example of the number of wavelengths per user required as a function of the number of switches per user for a 1000-user network. Recall from Theorem 1 that a strict lower bound on the number of wavelengths required is $\mu + \sigma$ wavelengths per user, achieved using full switching requiring $\mu + \sigma$ switches per user. From the figure, we observe that with a relatively small number of switches per user, we can come close to this lower bound – that is, the minimum number of $\mu + \sigma = 110$ wavelengths can be approached very closely with only 3 to 5 switches per user instead of 110 switches per user.

D. Typical Operating Regimes

In this section, we seek a closed-form, approximate expression for the total number of wavelengths required W_{tot} and the static and dynamic provisioning W_s and W_b that applies in typical network operating regimes. This will give us an intuitive sense of the relationship between these quantities and the network parameters such as the traffic mean μ and variance σ^2 and

TABLE I
WAVELENGTH REQUIREMENTS FOR ASYMPTOTICALLY NON-BLOCKING WAVEBAND-SWITCHED NETWORKS

Regime 1: $W_s \leq \mu + \left(\frac{b+1}{N}+1\right) W_b$

b	W_{tot}	W_s	W_b
$b \leq \frac{N\sigma}{2} - 1$	$N\mu + N\sigma + \left[\frac{N(N-1)}{2(b+1)}\right]\sigma$	$\mu + \frac{1}{2}\left(1+\frac{N}{b+1}\right)\sigma$	$\frac{1}{2}\left(\frac{N}{b+1}\right)\sigma$
$\frac{N\sigma}{2} - 1 < b \leq N(\mu+\sigma+1)-1$	$N\left[\mu+\sigma+\frac{N-1}{N}\right]$	$\mu+\sigma+1-\frac{b+1}{N}$	1
$b > N(\mu+\sigma+1)-1$	b	0	1

Regime 2: $\mu + \left(\frac{b+1}{N}+1\right) W_b < W_s \leq \mu + (b+2)W_b$

b	W_{tot}	W_s	W_b
$b \leq N\sigma$	$N\mu + N\left[\frac{b+N}{\sqrt{(b+1)(b+N)}}\right]\sigma$	$\mu + \left[\frac{(b+2N)}{2\sqrt{(b+1)(b+N)}}\right]\sigma$	$\left[\frac{N}{2\sqrt{(b+1)(b+N)}}\right]\sigma$
$b > N\sigma$	$N\left(\mu + \frac{N}{4(b+1)}\,\sigma^2\right)+b$	$\mu + \frac{N}{4(b+1)}\,\sigma^2$	1

Regime 3: $W_s > \mu + (b+2)W_b$

b	W_{tot}	W_s	W_b
$b \leq \frac{\sqrt{N}}{2}\sigma - 1$	$N\left(\mu + \frac{b+2}{b+1}\frac{\sqrt{N}}{2}\,\sigma\right) + \frac{b\sqrt{N}}{2(b+1)}\sigma$	$\mu + \frac{b+2}{b+1}\frac{\sqrt{N}}{2}\,\sigma$	$\frac{1}{b+1}\frac{\sqrt{N}}{2}\,\sigma$
$b > \frac{\sqrt{N}}{2}\sigma - 1$	$N\left(\mu+\sqrt{N}\sigma\right)$	$\mu+\sqrt{N}\sigma$	0

the number of switches b.

We will assume that in a typical environment, $b/N \leq \sigma/2$. From Figure 6, we see that this is a reasonable assumption, since we can approach the minimum number of wavelengths required with a relatively small number of switches per user. We also assume that N is large. We next consider the three regimes to determine which regime is optimal under these assumptions.

In Regime 1, we have $b/N \leq \sigma/2$, and the total number of wavelengths required is

$$\begin{aligned} W_{tot}^{(1)} &= N\mu + N\sigma + \left[\frac{N(N-1)}{2(b+1)}\right]\sigma \\ &\approx N\mu + N\left[1+\frac{N^2}{2b}\right]\sigma \end{aligned}$$

In Regime 2, it follows from our assumptions that $b/N < \sigma$ and we have

$$\begin{aligned} W_{tot}^{(2)} &= N\mu + N\left[\frac{b+N}{\sqrt{(b+1)(b+N)}}\right]\sigma \\ &\approx N\mu + N\left[\sqrt{1+\frac{N}{b}}\right]\sigma \end{aligned}$$

Since $\frac{1}{2\sqrt{N}}$ approaches zero as N increases, for sufficiently large N we must have $b/N > \frac{1}{2\sqrt{N}}\sigma$ in Regime 3 and

$$W_{tot}^{(3)} = N\mu + N\sqrt{N}\sigma$$

By inspection, we observe that the minimum number of wavelengths is achieved in Regime 2:

$$W_{tot} \approx N\mu + N\left[\sqrt{1+\frac{N}{b}}\right]\sigma$$

with

$$W_s \approx \mu + \frac{b+2N}{2\sqrt{b(b+N)}}\,\sigma \quad , \quad W_b \approx \frac{N}{2\sqrt{b(b+N)}}\,\sigma$$

Comparing the total number of static wavelengths NW_s with the total number of dynamic wavelengths bW_b, we observe that

$$NW_s \approx N\mu + \frac{N^2}{\sqrt{b(b+N)}} + bW_b$$

which gives a sense of the optimal relative amounts of static and dynamic provisioning that is appropriate. Note that we can take advantage of the predictability of the traffic to provision significantly more wavelengths statically.

IV. PROVISIONING FOR SMALL NUMBERS OF USERS

The analysis in the preceding sections is asymptotic in the number of users. Recall that each "user" in our shared-link context corresponds to an input-output fiber pair. In practical networks, we rarely have an infinite number of such users. However, we also rarely need to have strictly non-blocking networks – typically, we will have a target overflow probability that we consider to be sufficiently low to deliver good service. We would therefore also like to know how to use the results from our asymptotic approach in the preceding sections to networks with a small number of users in which we may allow a target overflow probability. In this section, we discuss how to adapt our results to this scenario.

A. *Statistics of the Traffic Vector Length*

Consider the distance R of the traffic vector $\mathbf{c}$ from its mean point $\mu\mathbf{1}$:

$$R = \|\mathbf{c} - \mu\mathbf{1}\| = \sqrt{\sum_{n=1}^{N}(c_i-\mu)^2}$$

Define a new quantity $X = R^2$. For a fixed N, X is a random variable with mean and variance given by

$$E[X] = E\left[\sum_{n=1}^{N}(c_i-\mu)^2\right] = N\sigma^2$$
$$\mathsf{var}(X) = \mathsf{var}\left[\sum_{n=1}^{N}(c_i-\mu)^2\right] = 2N\sigma^4$$

where the central moments of a Gaussian random variable can be found in most probability texts (e.g. [11]). Note that X is the sum of N squared Gaussian random variables and by the Central Limit Theorem can itself be approximated by a Gaussian random variable with mean $N\sigma^2$ and standard deviation $\sqrt{2N}\sigma^2$. We can observe that the standard deviation of X as a fraction of the mean decreases with increasing N as

$$\frac{\sqrt{2N}\sigma^2}{N\sigma^2} = \sqrt{\frac{2}{N}}$$

B. Practical Network Provisioning

In practical network provisioning, it is often not necessary to achieve totally non-blocking operation – it suffices if the blocking probability is sufficiently low. Suppose Q is the target overflow probability for the network (i.e. we wish to design the network so that the probability of overflow is at most Q).

The probability that the traffic vector falls within a sphere of radius r centered at the mean is $P(R \leq r)$. We choose r to satisfy

$$P(R \leq r) = Q \quad \Rightarrow \quad P(X \leq r^2) = Q$$
$$P\left(\frac{X}{\sigma^2} \leq \frac{r^2}{\sigma^2}\right) = Q \quad \Rightarrow \quad P\left(Y \leq \frac{r^2}{\sigma^2}\right) = Q$$

where Y is defined as X/σ^2. Observe that since Y is the sum of N zero-mean unit variance Gaussian random variables, it is itself a chi-squared random variable with N degrees of freedom. Then:

$$r = \sigma\sqrt{\mathsf{chi2inv}(Q, N)}$$

where $\mathsf{chi2inv}(Q, N)$ is the inverse CDF of a chi-squared random variable with N degrees of freedom at the point Q.

Since r is now the radius of a traffic sphere within which the realized traffic vector will be found with desired probability Q, the wavelength provisioning should be chosen so that the minimum distance from the traffic mean to the nearest boundary constraint hyperplane F_{min} exceeds r. This will ensure that an overflow event happens with probability no greater than Q.

C. Numerical Example

Recall that Table I was calculated assuming that the mean point needed to be a minimum distance of $\sqrt{N}\sigma$ from all boundary hyperplanes, while for finite N we now require the minimum distance to be at least r. We can therefore conclude that the expressions for each regime in Table I hold for finite N also, as long as σ in the table is replaced with $r/\sqrt{N}$. Figure 7 shows the total number of wavelengths required for a link with $\mu = 100$, $\sigma = 10$, and 2 switches per user for both a 1% overflow probability and the asymptotic minimum W_{tot}. Note that even for a small number of users, the number of extra wavelengths required (compared to the asymptotic minimum) is not large, and diminishes rapidly as the number of users increases.

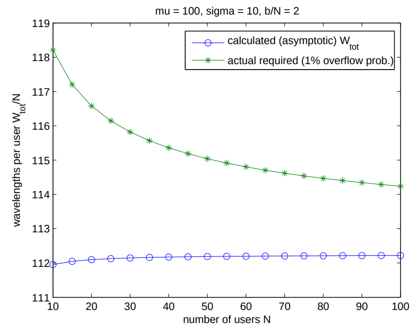

Fig. 7. A shared link with the mean statically provisioned targeting 1% overflow probability. The total number of wavelengths required for a link with $\mu = 100$, $\sigma = 10$, and 2 switches per user is shown both for 1% overflow probability and the asymptotic minimum W_{tot}.

V. CONCLUSION

We examined wavelength provisioning for a shared link in a backbone network, and considered networks with both static and dynamically provisioned wavelengths. Using a geometric argument, we obtained asymptotic results for the optimal wavelength provisioning on the shared link. We proved that the number of static wavelengths should be sufficient to support at least the traffic mean. We derived in closed form expressions for the optimal provisioning of the shared link given the mean μ and variance σ^2 of the traffic. We also showed that by allowing the dynamic wavelengths to be switched in bands of multiple wavelengths rather than individually, very efficient networks can be achieved while using a very small number of switches per user. We again derive the optimal static and dynamic provisioning as well as the optimal waveband size given the traffic characterization. Finally, we showed that the results could be adapted for networks with a small, finite number of users with a fixed target overflow probability.

REFERENCES

[1] R. Ramaswami and K. N. Sivarajan, *Optical Networks: A Practical Perspective*, Morgan Kaufmann, 1998.

[2] L. Li and A. K. Somani, "Dynamic wavelength routing using congestion and neighborhood information," *IEEE/ACM Trans. Networking*, vol. 7, pp. 779–786, October 1999.

[3] M. Kovacevic and A. Acampora, "Benefits of wavelength translation in all-optical clear-channel networks," *IEEE J. Select. Areas Commun.*, vol. 14, pp. 868–880, June 1996.

[4] O. Gerstel, G. Sasaki, S. Kutten, and R. Ramaswami, "Worst-case analysis of dyanmic wavelength allocation in optical networks," *IEEE/ACM Trans. Networking*, vol. 7, pp. 833–845, December 1999.

[5] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley-Interscience, 1991.

[6] R. Guerin and L. Y.-C. Lien, "Overflow analysis for finite waiting-room systems," *IEEE Trans. Commun.*, vol. 38, pp. 1569–1577, September 1990.

[7] R. B. Cooper, *Introduction to Queueing Theory, 2nd Ed.*, North Holland, New York, 1981.

[8] D. A. Garbin M. J. Fischer and G. W. Swinsky, "An enhanced extension to wilkinson's equivalent random technique with application to traffic engineering," *IEEE Trans. Commun.*, vol. 32, pp. 1–4, January 1984.

[9] A. A. Fredericks, "Congestion in blocking systems - a simple approximation technique," *Bell Syst. Tech. J.*, vol. 59, pp. 805–827, July-August 1980.

[10] L. Chen, *A Study on the Tradeoff Between Efficient Resource Allocation and Node Complexity in WDM Networks*, Ph.D. thesis, MIT, September 2002.

[11] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 2002.