Multiuser Capacity Analysis of WDM in Nonlinear Fiber Optics

Mohammad H. Taghavi N., George C. Papen, and Paul H. Siegel
Dept. of ECE, UCSD, La Jolla, CA 92093
Email: {mtaghavi, gpapen, psiegel}@ucsd.edu

Abstract

The capacity region of wavelength division multiplexing in nonlinear fiber optic channels with coherent communication is studied. We show that by optimally using the information from all the channels at the receiver, the loss in the capacity due to the mixing of channels is negligible in the weakly nonlinear regime. On the other hand, with single-channel detection, capacity saturates as the total interference power becomes comparable to the noise power. Similar results hold for the case of channels with memory. The transmitter can achieve the capacity without knowing the nonlinearity parameters. Some insight is provided on the structure of optimal/suboptimal receivers.

1 Introduction

In order to achieve higher data rates in long haul optical fiber systems, higher signal to noise ratios are required at the receiver. However, as the signal intensity increases, the nonlinearities in fiber affect the signal propagation. In a wavelength division multiplexing (WDM) system, these nonlinear effects cause the signal traveling in each frequency band to modulate the signals at all frequencies.

It appears that the first published effort to characterize the effect of nonlinearities on the throughput of wavelength division multiplexing (WDM) systems was the paper by Mitra and Stark [1], in which they modeled the nonlinear crosstalk as noise. They predicted that since the interference power grows faster than the signal power, the capacity will saturate when interference dominates. Ho and Kahn made a further step in [2] and used the fact that in certain regions of operation, the dominant crosstalk terms only depend on signal intensities. Therefore, by using phase modulation at the transmitters, crosstalk becomes predictable. However, this restriction on the modulation format reduces the capacity significantly. Xu and Brandt-Pearce in [3] showed that by using a multiuser detector (MUD) to simultaneously detect the symbols transmitted through all the sub-channels, the bit error probability for the practical, but restricted case of on-off keying (OOK) can be significantly improved.

In this paper, we investigate the capacity of the nonlinear fiber optic channel with WDM from a multiuser point of view. To model the channel, the Volterra series expansion of the input/output relation derived in [4] is used. We define the weakly nonlinear regime as the region where only the first nonlinear term in the Volterra series is significant, and the higher order terms can be neglected. With this approximation, the change in the capacity region due to the nonlinearity is shown to be negligible if optimal MUD is used at the receiver. However, if the receiver only uses the output from one sub-channel, the capacity experiences a large reduction from the linear case even in the weakly nonlinear regime, as indicated by earlier results. If fiber dispersion is strong, the combination of the relative "walk-off" of the different carriers with the nonlinear mixing causes the channel to have memory, which cannot be compensated for passively. We show that even with this memory, the capacity region with the optimal receiver is close to the linear case.

The rest of the paper is organized as follows. Section II defines channel model. In Section III, we derive the capacity region with the optimal detector, for both the weak and strong dispersion regimes. In Section IV, this capacity is compared to the capacity of the channel with conventional single-channel detection. Section V concludes the paper.

2 Channel Model

For a single-mode optical fiber with chromatic dispersion and Kerr nonlinearity, the slowly varying complex envelope or low-pass equivalent of the optical field, $A(t, z)$, at time t and distance z from the transmitter is described by the nonlinear Schrödinger equation [5]

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2}A + \frac{j}{2}\beta_2\frac{\partial^2 A}{\partial \tau^2} - j\gamma\left|A\right|^2 A, \tag{1}$$

where $\tau = t - \beta_1 z$ is the time in the reference frame of the moving pulse. In this equation, α is the fiber loss factor, γ is the nonlinearity coefficient, and β_1 and β_2 are respectively the linear and quadratic dispersion coefficients. Strong linear dispersion can result in a relative "walk-off" between the signals traveling on different carriers, because of the difference in their group velocities.

To obtain an analytical input-output relation for the channel, Volterra series is applied to derive a series expansion for the low-pass equivalent field in the frequency domain [4]. The Volterra series solution for the Fourier transform of the output low-pass equivalent field, $A\left(\omega, z\right)$, in terms of the input low-pass equivalent field, $A_0(\omega)$, up to the third order term can be written as [4]

$$\begin{aligned} A(\omega, z) =& H_1(\omega, z)A_0(\omega) \\ &+ \iint H_3(\omega_1, \omega_2, \omega - \omega_1 + \omega_2, z)A_0(\omega_1)A_0^*(\omega_2)A_0(\omega - \omega_1 + \omega_2)d\omega_1 d\omega_2. \end{aligned} \tag{2}$$

In the absence of nonlinearity, the signal experiences attenuation from absorption and scattering, as well as dispersion. Linear dispersion results in a relative walk-off between different carriers, and quadratic dispersion produces phase and amplitude distortion. At the receiver, these effects can be compensated for by using an optical amplifier for the attenuation and a dispersion compensating fiber (DCF) or an electronic equalizer for the dispersion. However, in the presence of both dispersion and nonlinearity, the DCF cannot compensate for their joint effect. The amplification introduces spontaneous emission noise that is well approximated for large amplifier gains as a signal-independent additive white Gaussian noise (AWGN) in the optical domain.

By examining the expression for the cubic and fifth order terms of the Volterra series in [4], we observe that the ratios of the magnitudes of these terms are, respectively, order $O(\delta)$ and order $O\left(\delta^2\right)$, where $\delta \triangleq \frac{\gamma|A|^2}{\alpha}$. Here, δ is a small and dimensionless number and is a measure of the total phase shift in the signal due to the nonlinearity. In this paper, we confine our analysis to the region $\delta \ll 1$, and neglect all terms smaller that or comparable to δ^2 in magnitude, leaving only the linear and cubic terms listed in (2).

We consider a WDM system with K users transmitting in different wavelength-channels with identical signal spectra. The receiver has access to the fiber output at all frequency bands, hence the channel can be modeled as multiple-access. We futher assume that frame-synchronism and symbol-synchronism among the users is achieved at the receiver, since the walk-off effect of linear dispersion can be compensated for. However, as a result of this walk-off along the fiber and the instantaneous nature of the nonlinearity, each symbol will be modulated not only by the other symbols transmitted during the same symbol period, but also by the previous and past symbols of other channels, which cannot be compensated for by a linear equalizer. Consequently, the nonlinear optical fiber has a memory that increases in length with dispersion. We will first study the weakly dispersive fiber, where the channel is memoryless, and later in Section III-C we will generalize the results to a channel with memory.

The frequency domain input to the fiber during the nth symbol period can be written as

$$A_n(\omega) = \sum_{k=1}^{K} g_k u_k(n) V(\omega - k\Delta\omega), \tag{3}$$

where $u_k(n), n = 1, \cdots, N$ is the channel-coded symbol sequence of the kth user, with N being the block length, g_k is a complex constant for normalizing the input power and phase bias of the kth transmitter, and $V(\omega)$ is the Fourier transform of the band-limited pulse shape, $v(t)$, normalized to have a unit total energy. To simplify the equations, we define $x_k(n) = g_k u_k(n)$ as the equivalent discrete-time channel inputs. The coded symbols satisfy a statistical and temporal average power constraint,

$$\frac{1}{N} \sum_{n=1}^{N} E\left[|u_k(n)|^2\right] \leq 1. \tag{4}$$

Hence, $|g_k|^2$ is the maximum allowed energy per symbol duration transmitted by the kth user. For brevity, we drop the index n in the analysis of the memoryless case.

To derive a discrete-time equivalent model for the channel, the received signal should be projected over a set of basis functions spanning the space of all possible received signals. In the absence of nonlinearity, it is optimal to choose the responses of the channel to the waveforms transmitted by the K users as the basis functions, i.e. to use a matched filter for each user followed by a sampler. Interestingly, this structure remains optimal in the weakly nonlinear regime (Fig. 1). This can be proved by finding a set of basis functions to span the remainder of the space not covered by the matched filters, and then showing that, given the outputs of the matched filters, the extra information contained in the other outputs is $O(\delta^2)$.

Assume that the linear dispersion is weak, so that the channel is memoryless. In this case, using (2) and (3) the complex output of the ith branch tuned to the ith user's signal can be written as

$$y_i = x_i + \sum_{k=1}^{K} \sum_{l=1}^{K} \sum_{m=1}^{K} \xi_{k,l,m}^{(i)} x_k x_l^* x_m + n_i, \tag{5}$$

where $\{n_i\}$ are the additive noise terms modeled as i.i.d. circularly symmetric complex Gaussian random variables with variance σ^2 in both the real and imaginary dimensions. The crosstalk coefficients $\xi_{k,l,m}^{(i)}$ are proportional to $\frac{\gamma}{\alpha}$, and can be calculated from (2) and (3). Since the frequency bands are non-overlapping, $\xi_{k,l,m}^{(i)}$ is only nonzero when $k - l + m = i$.

The effects that give rise to the interference terms in (5) are often classified into three categories. If $k = l = m = i$, then the corresponding term is caused by self-phase modulation

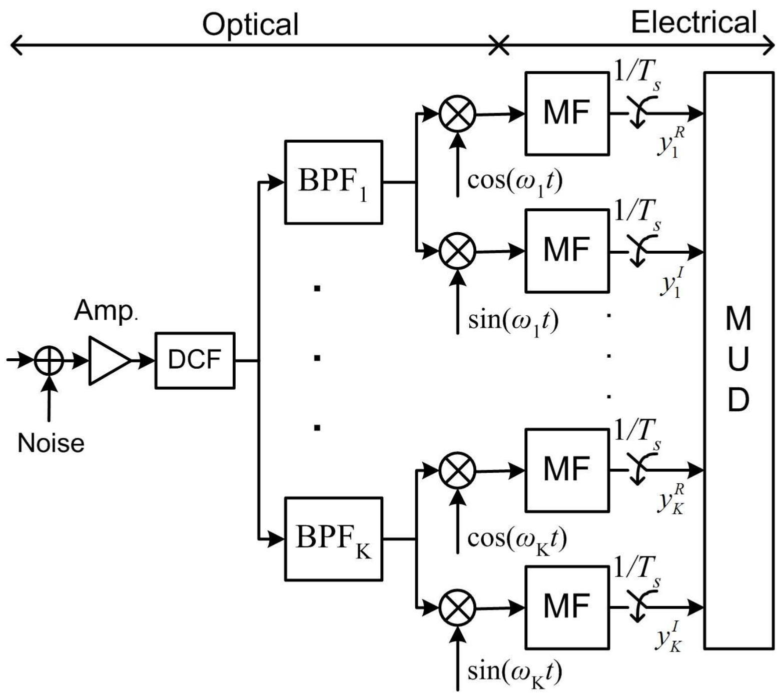

Figure 1: Equivalent Receiver Structure

(SPM). Cross-phase modulation (XPM) produces the terms for which $k = l$ and $m = i$, or $k = i$ and $l = m$, and the rest of the terms are four-wave mixing (FWM) terms. Of these classes, FWM is suppressed in a strongly dispersive fiber, since the signals at different frequencies travel at different group velocities and hence walk-off too rapidly to interact. Therefore, in a strongly dispersive fiber, FWM is much smaller than XPM, and it can be neglected.

Now we can rewrite (5) as

$$y_i = x_i + \kappa_i + \varphi_i + n_i, \tag{6}$$

where φ_i contains the FWM terms and κ_i contains the SPM and XPM terms. To the first order of approximation, κ_i can be written as

$$\kappa_i = x_i \sum_{k=1}^{K} j\rho_k^{(i)} \left| x_k \right|^2 \tag{7}$$

where $\rho_k^{(i)}$ is a real constant. This means that the effect of SPM and XPM on the signal magnitude is negligible, and they act as phase distortion.

Throughout the paper, we assume that channel parameters, i.e. all the gains and crosstalk coefficients, are known at both the transmitter and the receiver.

3 Capacity Region

In this section, we first study the capacity region of the memoryless system described by (5) up to the first order of approximation, i.e. by neglecting all terms of order $O(\delta^2)$. This result will be generalized for the channel with memory at the end of the section. Due to the space limitation, some of the proofs and details of the derivations are omitted, and the reader is referred to [6] for more details.

We group the inputs and outputs of the channel into two real vectors $X = \left[\underline{x}_1^T \; \cdots \; \underline{x}_K^T\right]^T$ and $Y = \left[\underline{y}_1^T \; \cdots \; \underline{y}_K^T\right]^T$ of size $2K \times 1$, where $\underline{x}_k = [x_k^R \; x_k^I]^T$ and $\underline{y}_k = [y_k^R \; y_k^I]^T$ are vector representations of complex samples, x_k and y_k, with superscripts R and I denoting the real and imaginary parts, respectively. Now, we can rewrite (5) in a vector form,

$$Y = X + \Theta + Z, \tag{8}$$

where Θ and Z contain the crosstalk and noise terms, respectively. The following proposition gives the capacity region of the channel when each user k transmits with an average energy per symbol duration upper-bounded by P_k, i.e. $|g_k|^2 = P_k$.

Proposition 1. *To the first order of approximation of the nonlinearity, the capacity region of the memoryless coherent WDM channel described by (5) is*

$$\left\{(R_1, \cdots, R_K) : \forall k \;\; 0 \le R_k \le \log\left(1 + P_k/2\sigma^2\right)\right\}. \tag{9}$$

Note that (9) is the same as the capacity region of the linear channel. In other words, to the first order of approximation, nonlinearity does not affect the channel capacity.

Proof. We will first prove that (9) is an outer bound on the capacity region, and then show its achievability.

3.1 Outer Bound on the Capacity Region

Since the gains $\{g_i\}$ are known at both the transmitters and the receivers, the $\{x_i\}$ are sufficient statistics for the coded symbols, $\{u_i\}$. Now, from the capacity region of the multiple-access channel [7, pp. 389-390], we have

$$\forall S \subset \{1, \cdots, K\} : 0 \le \sum_{i \in S} R_i \le I(X_S; Y | X_{S'}), \tag{10}$$

where S' is the complement of S with respect to the reference set $\{1, \cdots, K\}$, and X_A for any set A is a sub-vector of X obtained by keeping only the elements corresponding to users whose indices belong to A. For the right-hand side of (10), using the chain rule for mutual information, we can write

$$I(X_S; Y | X_{S'}) = I(X_S; Y_S | X_{S'}) + I(X_S; Y_{S'} | X_{S'}, Y_S). \tag{11}$$

This is the total amount of information that can be transferred if only the users in S are communicating through the channel. It can be shown that the second term is $O(\delta^2)$, and therefore can be neglected.

By expanding the first term on the right hand side of (11) we obtain

$$I(X_S; Y_S | X_{S'}) = h(Y_S | X_{S'}) - h(Y_S | X). \tag{12}$$

Since the interference terms are deterministic functions of $\{x_i\}$, the second term reduces to

$$h(Z_S) = |S| \log\left(2\pi e \sigma^2\right), \tag{13}$$

where we have used the fact that the noise terms are i.i.d. Gaussian random variables. This expression is independent of the distribution of $\{x_i\}$, hence for maximizing the mutual information it suffices to maximize the first term in (12) with respect to the probability distribution of X. This term satisfies

$$h(Y_S | X_{S'}) \le \sum_{i \in S} h(y_i | X_{S'}), \tag{14}$$

with equality if $\{y_i\}$ are independent. Furthermore, for each $i \in S$ we have [7, p. 234]

$$\begin{aligned} h(y_i|X_{S'}) &= E_\chi \left[h(y_i|X_{S'} = \chi)\right] \\ &\le E_\chi \left[\frac{1}{2}\log\left((2\pi e)^2 \cdot \det\left(\text{cov}\left[\underline{y}_i|X_{S'} = \chi\right]\right)\right)\right], \end{aligned} \tag{15}$$

with equality if $\underline{y}_i$ is a Gaussian 2-vector given $X_{S'}$.

To simplify the expressions, we introduce the notations

$$p_i^R := E[(x_i^R)^2],\ v_i^R := \left(\text{var}[x_i^R]\right)^{\frac{1}{2}},\ \mu_i^R := E[x_i^R]/v_i^R, \tag{16}$$

respectively for the power, standard deviation, and normalized mean of x_i^R, and similarly p_i^I, v_i^I, and μ_i^I for x_i^I. Also we define

$$q_i := \text{cov}[x_i^R, x_i^I]/v_i^R v_i^I, \tag{17}$$

to denote the correlation coefficient of x_i^R and x_i^I. By some algebra, the determinant in (15) can be expanded as

$$\begin{aligned} \det\left(\text{cov}\left[\underline{y}_i|X_{S'} = \chi\right]\right) &= \left(p_i^R + \sigma^2 - c_i\right)\left(p_i^I + \sigma^2 + c_i\right) - \left[q_i^2(v_i^R)^2(v_i^I)^2 + q_i O(P^2\delta)\right] \\ &\quad - \left[(\mu_i^R)^2(v_i^R)^2 p_i^I + \mu_i^R O(P^2\delta)\right] - \left[(\mu_i^I)^2(v_i^I)^2 p_i^R + \mu_i^I O(P^2\delta)\right], \end{aligned} \tag{18}$$

where

$$c_i \triangleq E\left[x_i^R x_i^I \rho_i^{(i)} |x_i|^2\right], \tag{19}$$

and $\rho_i^{(i)}$, as used in (7), is the SPM coefficient for the ith channel. Using the AM-GM inequality and the power constraint, P_i, for the first term we have

$$\begin{aligned} \left(p_i^R + \sigma^2 - c_i\right)\left(p_i^I + \sigma^2 + c_i\right) &\le \left[\frac{\left(p_i^R + \sigma^2 - c_i\right) + \left(p_i^I + \sigma^2 + c_i\right)}{2}\right]^2 \\ &\le \left(P_i/2 + \sigma^2\right)^2, \end{aligned} \tag{20}$$

with equality if $p_i^R = p_i^I = P_i/2$ and $c_i = 0$. Furthermore, all the three bracketed expressions on the right-hand side of (18) have the form

$$az^2 + O(a\delta)z,\ \ a > 0, \tag{21}$$

where z respectively denotes q_i, μ_i^R, and μ_i^I in these expressions. Expression (21) is positive for $|z| \gg \delta$, and equals zero for $|z| = 0$. Also, for $|z| = O(\delta)$, (21) becomes $O(a\delta^2)$, or equivalently, $O(P^2\delta^2)$, which is negligible compared to the largest term in (18), which is $p_i^R p_i^I$. Hence, to the first order of approximation, all the three bracketed expressions in (18) are lower-bounded by zero. This observation along with (20) yields

$$\det\left(\text{cov}\left[\underline{y}_i|X_{S'} = \chi\right]\right) \le \left(P_i/2 + \sigma^2\right)^2. \tag{22}$$

This result can be combined with (10)-(15) to conclude that

$$0 \le \sum_{i\in S} R_i \le \sum_{i \in S} \log\left(1 + \sigma^{-2}P_i/2\right)\ \ \forall S \subset \{1, \cdots, K\}, \tag{23}$$

which can be rewritten in the form of (9), as an outer bound on the capacity region.

3.2 Achievability of the Bound

A direct approach to prove the achievability would be to show that with a certain selection of probability densities for the source symbols, u_i, inequalities (14), (15), and (22) become equalities. Equality can be achieved in (22) by selecting $\{u_i\}$ to be i.i.d. circularly-symmetric complex Gaussian random variables with variance $1/2$ per complex dimension. Moreover, in the absence of FWM, this distribution makes the set $\{y_i\}$ independent and Gaussian; thus equality can be achieved in (14) and (15), as well. This results from the fact that each transmitted symbol has a uniformly distributed phase, hence it remains uniform and independent of other signals even with a phase distortion added by XPM and SPM. However, when FWM is present, even though (15) can be asymptotically satisfied using a central limit theorem approach, the joint dependence of $\{y_i\}$ cannot be neglected.

We show the achievability of (9) for the general case by using a simple and generally suboptimum interference cancellation scheme. Given the vector of received samples, Y, we use y_i as an estimate of x_i for each i, and use these estimates to cancel the crosstalk in the outputs. Then, we detect each x_i again from the corresponding sample, assuming that the other users' symbols are random.

To explain this method and its performance, let's assume that user i is the user of interest. Recall the expression (5) for the channel output samples. Now, we form the test statistic

$$z_i = y_i - \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K}\xi^{(i)}_{k,l,m} y_k y_l^* y_m. \tag{24}$$

and use it as the only reference for detecting x_i (and/or u_i), throwing away all the extra information in $\{y_k\}$. Expanding (24) using (5), and neglecting all the higher order terms, we obtain

$$\begin{aligned} z_i = x_i + n_i &- \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K}(\xi^{(i)}_{k,l,m} + \xi^{(i)}_{m,l,k})(n_k + \theta_k) x_l^* x_m \\ &- \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K}\xi^{(i)}_{k,l,m} x_k (n_l^* + \theta_l^*) x_m, \end{aligned} \tag{25}$$

where θ_k contains all the crosstalk terms on channel k. Finding the capacities of the channels with input-output pairs $(x_i, z_i),\ i = 1, \cdots, K$ gives a set of achievable rates for the original channel. The mutual information of (x_i, z_i) can be written as

$$I(x_i, z_i) = h(z_i) - h(z_i|x_i). \tag{26}$$

We assume that the symbols generated by each user, u_i, are i.i.d. Gaussian random variables.

Since the remainder of crosstalk is at least two orders smaller than x_i, we can neglect it when calculating $h(z_i)$. Hence, we we have

$$h(z_i) = \log\left(2\pi e(P_i/2 + \sigma^2)\right). \tag{27}$$

On the other hand, when evaluating $h(z_i|x_i)$, we can't neglect the crosstalk terms, since the crosstalk terms may not be negligible compared to noise, which is the dominant term now. As in Section III-A, we can upper bound this conditional entropy by

$$E_\chi\left\{\frac{1}{2}\log\left((2\pi e)^2\ \text{var}[z_i^R|x_i = \chi]\cdot\text{var}[z_i^I|x_i = \chi]\right)\right\}, \tag{28}$$

with equality if z_i^R and z_i^I are independent and Gaussian given x_i. We can also upper bound the product of variances in (28) by

$$E\left[(z_i^R)^2|x_i = \chi\right] \cdot E\left[(z_i^I)^2|x_i = \chi\right] \leq 1/4E\left[|z_i|^2\right]^2, \tag{29}$$

Using the properties of XPM and FWM, it can be shown that

$$E\left[|z_i|^2\right] = 2\sigma^2. \tag{30}$$

Finally, by combining (30) with (26)-(29), we obtain the lower bound

$$I(x_i, z_i) \geq \log(1 + \sigma^{-2}P_i/2). \tag{31}$$

This result implies that with a simple interference cancellation scheme the upper bound of Section III-A is achievable. Note that in this scheme, the transmitter doesn't use the crosstalk coefficients. Hence, to the first order of approximation, (9) is the capacity region of the channel, independent of whether the transmitter has the crosstalk parameters or not. □

3.3 Capacity Region of the Channel with Memory

In a channel with strong dispersion, the relative walk-off between signals traveling at different frequencies along with nonlinear mixing introduces (finite) memory to the channel. In this case, (5) should be rewritten to consider this effect by adding the dimension of time to the triple summation on the right-hand side. As an extension of Proposition 1, the following proposition determines the capacity region of this channel.

Proposition 2. *To the first order of approximation of the nonlinearity, the capacity region of the coherent multiuser WDM channel with memory is the region given by (9).*

Proof. Denote by X_S^N and Y_S^N the matrices containing respectively all the inputs and outputs of the channel at time-slots 1 to N and subchannels with indices in $S \subset \{1, \cdots, K\}$. Since the users are frame-synchronous and the channel has finite memory, we can use the following theorem to derive the capacity region.

Theorem 1. *(Verdú, [8]) The capacity region of a frame-synchronous multiple-access channel with finite memory M is given by*

$$C = \text{Closure}\left(\liminf_{N\to\infty} \frac{1}{N} C_N\right). \tag{32}$$

where C_N is given by

$$\bigcap_S \left\{(R_1, \cdots, R_K) : 0 \leq \sum_{k\in S} R_k \leq I(X_S^N; Y^N | X_{S'}^N)\right\}. \tag{33}$$

The channel model in the presence of memory can be visualized as a simple generalization of the memoryless channel, where vectors in (8) are replaced by matrices, i.e., we deal with both frequency and time indices. However, the properties of interference terms are preserved, hence we can use some of the results obtained in the previous section. Following the same approach as Section II, we can show that

$$I(X_S^N; Y^N | X_{S'}^N) \leq \sum_{i\in S} \sum_{n=1}^{N} \log\left(1 + \frac{\sigma^{-2}}{2} E\left[|x_i(n)|^2\right]\right). \tag{34}$$

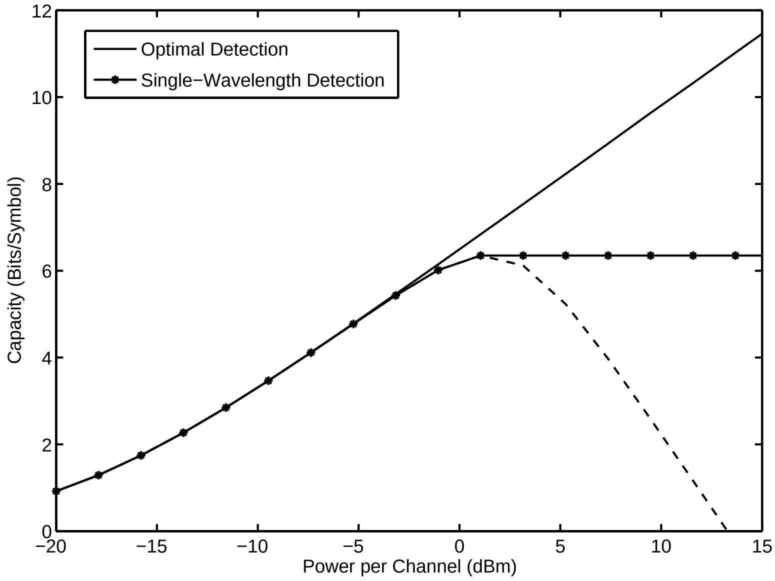

Figure 2: Capacity versus Transmit Power per Channel. Gaussian pulses with bandwidth 80GHz, $\gamma = 1$W^{-1}Km^{-1}, $\beta = -20$ps^2/Km, $\alpha = 0.25$dB/Km. The dashed line corresponds to the mutual information with single-user detection if the users are forced to transmit with the maximum power.

The inner summation will be maximized if all the symbols of each user have equal average energies. Then, we will have

$$I(X_S^N; Y^N | X_{S'}^N) \leq N \sum_{i \in S} \log\left(1 + P_i/2\sigma^2\right). \tag{35}$$

Similar to the previous subsection, this bound is achievable if the symbols transmitted in different time slots and sub-channels are all independent Gaussian random variables. Finally, plugging (35) into (33) and dividing by N completes the proof. □

4 Comparison and Discussion

To determine the capacity gain of multiuser detection, we need to calculate the capacity with a single-wavelength detector. In this scheme, all the symbols from the interfering users are treated as random sequences. Hence, as opposed to the case of optimal multi-wavelength detection, the crosstalk terms don't vanish in $h(y_i|x_i)$, making it difficult to compute. Fortunately, since the transmitted symbols are independent, the sum of crosstalk terms in (5) form a third order "weighted U-statistic," which obeys a central limit theorem for large numbers of users, although these term are not independent [9]. This enables us to model the interference as Gaussian noise to compute the entropy given a certain realization of x_i, which then should be averaged over x_i.

The capacity per user has been plotted versus the transmit power per user in Fig. 2 for a WDM channel with 32 users. It is assumed that users are transmitting with equal powers and symbol rates, and the channel is memoryless. It should be emphasized that all the results are valid as long as the fundamental assumption of weak nonlinearity holds. The uncertainty in the result due to this assumption can be estimated by computing the ratio of the total crosstalk power to the signal power, e.g. in (5). For the values plotted in Fig. 2, this uncertainty is

1% at $P = 5$ dBm and 5% at $P = 9$ dBm. It is observed that within the range of acceptable accuracy, the achievable rates with the single-wavelength detection scheme saturate as the crosstalk power becomes comparable to the noise power, while the optimal scheme can achieve the capacity of a linear fiber.

5 Conclusion

We derived the multiuser capacity region of WDM in a nonlinear fiber using a weak nonlinearity approximation. If the outputs of the fiber at all sub-channels are used for detection, the nonlinear crosstalk doesn't change the capacity. This result holds also if the channel has memory due to the walk-off between different carriers. Every point in the capacity region can be achieved if each user transmits Gaussian distributed channel-coded symbols, without knowing the nonlinearity parameters. On the other hand, if only the output of one sub-channel is used, the capacity will saturate when the crosstalk dominates. It is concluded that the crosstalk introduced by nonlinearity does not severely limit the capacity of optical fibers, as long as the weak nonlinearity assumption is not violated.

References

[1] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, 411, pp. 1027-1039, 2001.

[2] K. Ho and J. M. Kahn, "Channel capacity of WDM systems using constant-intensity modulation formats," *Proc. IEEE Optical Fiber Comm. (OFC'02)*, 2002.

[3] B. Xu and M. Brandt-Pearce, "Multiuser detection for WDM fiber-optic communication systems," *Proc. IEEE Int. Symp. Information Theory*, 2002.

[4] K. V. Peddanarappagari and M. Brandt-Pearce, "Volterra series transfer function of single-mode fibers," *Journal of Lightwave Technology*, vol. 15, pp. 2232-2241, Dec. 1997.

[5] G. P. Agrawal, *Nonlinear fiber optics*, New York: Academic Press, 1989.

[6] M. H. Taghavi, G. C. Papen, and P. H. Siegel, "On the multiuser capacity of WDM in nonlinear fiber optics: Coherent communication," *To be submitted to IEEE Trans. Inform. Theory*.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[8] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Inform. Theory*, vol. 35, pp. 605-619, May 1989.

[9] Y. Rinott and V. Rotar, "On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics," *The Annals of Applied Probability*, vol. 7, no. 4, pp. 1080-1105, 1997.