On the Multiuser Capacity of WDM in a Nonlinear Optical Fiber: Coherent Communication

Mohammad H. Taghavi, George C. Papen, *Member, IEEE*, and Paul H. Siegel, *Fellow, IEEE*

***Abstract*—Previous results suggest that the crosstalk produced by the fiber nonlinearity in a WDM system imposes a severe limit to the capacity of optical fiber channels, since the interference power increases faster than the signal power, thereby limiting the maximum achievable signal-to-interference-plus-noise ratio (SINR). We study this system in the weakly nonlinear regime as a multiple-access channel, and show that by optimally using the information from all the channels for detection, the change in the capacity region due to the nonlinear effect is minimal. On the other hand, if the receiver uses the output of only one wavelength channel, the capacity is significantly reduced due to the nonlinearity, and saturates as the interference power becomes comparable to the noise, which is consistent with earlier results. The results hold in channels with or without memory. Every point in the capacity region can be achieved without knowledge of the nonlinearity parameters at the transmitters. The structures of optimal/suboptimal receivers are briefly discussed.**

***Index Terms*—Capacity, multiple-access channel, multiuser detection, optical fiber nonlinearities, wavelength-division multiplexing (WDM).**

I. Introduction

In order to achieve higher data rates in long-haul optical fiber systems, higher signal-to-noise ratios (SNRs) are required at the receiver. However, as the signal intensity increases, the nonlinearities in the fiber affect the signal propagation. The dominant form of nonlinearity, known as the Kerr effect, is caused by the dependence of the index of refraction on the instantaneous signal intensity. In a wavelength-division multiplexing (WDM) system, this effect causes the signal centered at one frequency to modulate the signals at all frequencies by changing the index of refraction. The nonlinear effects can be separated into the cross-phase modulation (XPM) effect, where the phase of each signal is distorted as a function of the intensities of the signals centered at other frequencies, and the four-wave mixing (FWM) effect, which is a distortion on both the phase and the amplitude of the signals.

It appears that the first published effort to characterize the effect of nonlinearities on the throughput of WDM systems was the work by Mitra and Stark [1], which was later reproduced in more detail in [2]. In that work, they estimated the capacity for each user by modeling the crosstalk in each channel as a combination of multiplicative and additive noise. Their analysis predicted that since the interference power grows faster than the signal power, the mutual information between the input and output will start to decrease with power when the interference becomes comparable to the additive (optical amplifier) noise. Hence, they declared this effect as a fundamental limit to the capacity of fiber-based systems. A number of similar results were also obtained for different scenarios, e.g., in [3]. Ho and Kahn made a further step in [4] and used the fact that in certain regions of operation, the dominant crosstalk terms are caused by XPM, which only depends on the signal intensities. Therefore, by keeping the amplitudes constant and using phase modulation at the transmitters, the distortion caused by the nonlinearities becomes constant. However, this restriction on the modulation format imposes a large penalty on the capacity by taking away one degree of freedom. Moreover, this technique is ineffective against FWM, which in contrast to XPM, depends on both the intensity and the phase of the transmitted symbols. Among other works, Narimanov and Mitra in [6] developed a perturbation theory for evaluating the capacity of a single-wavelength nonlinear fiber.

In all the works mentioned above, detection in each subchannel was done independently from that in other channels. Thus, any interference from other channels had to be treated as random noise. A result by Xu and Brandt-Pearce in [5] was the first and, to the best of our knowledge, the only work that discussed the advantage of multiuser detection (MUD) for this channel. They studied the practical, but restricted case of on–off keying (OOK) modulation with square-law detection, and showed that, by using a multiuser detector to simultaneously detect the symbols transmitted through all the subchannels, the bit-error probability can be significantly improved.

In this paper, we investigate the capacity of the nonlinear optical fiber channel with WDM from a multiuser point of view. To model the channel, the Volterra series expansion of the input/output relation derived in [7] is used. We define the weakly nonlinear regime as the region where the first nonlinear term in the Volterra series is significant, and the higher order terms can be neglected. With this approximation, we show that the change in the capacity region due to the nonlinearity is negligible if the receiver optimally uses the outputs of every wavelength-channel, which is equivalent to MUD in correlated multiuser channels. However, if the receiver uses the output in only one wavelength-channel, the capacity experiences a large reduction due to the nonlinearity even in the weakly nonlinear regime, which is consistent with earlier results. Consequently, optimal multiwavelength detection (MWD) allows us to increase the

Manuscript received May 4, 2005; revised March 8, 2006. The material in this paper was presented in part at the 43rd Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2005.

The authors are with the Electrical and Computer Engineering Department, University of California, San Diego, La Jolla, CA 92093-0407 USA (e-mail: mtaghavi@ucsd.edu; gpapen@ucsd.edu; psiegel@ucsd.edu).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2006.883540

transmit power beyond the limit dictated by the result of [1], which translates to a higher capacity for an equal range of transmission, or a longer range with the same capacity. The assumption of weak nonlinearity introduces an uncertainty in the expression for the capacity, which can be estimated in terms of the transmit power. As demonstrated in Section VI, within a range of reasonable uncertainty, the multiuser capacity of the channel is much higher that the capacity predicted in [1], where the interference was modeled as noise.

In an optical fiber channel with chromatic dispersion, the group velocity (envelope velocity) is frequency (or wavelength) dependent. This causes any two narrowband signals centered at different wavelengths to lose their synchronism due to the unequal delays that they experience as they propagate in the fiber, a phenomenon known as the relative "walk-off" effect. When the dispersion is large, the combination of the walk-off of the carriers with the nonlinear mixing causes the channel to have memory, an effect which cannot be compensated passively. We show that even in the presence of memory, the capacity region of the optimal receiver is close to the linear case.

The rest of the paper is organized as follows. Section II defines the physical properties of the channel and our discrete-time model. In Section III, we derive the capacity region for the low-dispersion case, where the channel is memoryless. This result is generalized to the strongly dispersive case in Section IV. In Section V, the effect of nonlinearity on the capacity is investigated for the suboptimal receiver structure with single-wavelength detection. Some numerical comparisons are presented in Section VI. Section VII concludes the paper.

II. Channel Model

A. Nonlinear Optical Fiber

For a single-mode optical fiber with chromatic dispersion and a 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 [8]

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

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, β_1 is the inverse of the group velocity, β_2 is the group velocity dispersion (GVD) parameter, and γ is the nonlinearity coefficient. The last term on the right-hand side of the equation corresponds to the Kerr nonlinear effect. It is useful to introduce two parameters, the effective length and the dispersion length. The effective length $L_{\text{eff}} = \left(1 - e^{-\alpha L}\right)/\alpha$, where L is the physical length of the fiber, is a measure of the distance where the nonlinear effects become significant. The dispersion length $L_{\text{D}} = (2\pi\beta_2 B\Delta\nu)^{-1}$, where B is the channel bandwidth and $\Delta\nu = \frac{\Delta\omega}{2\pi}$ is the channel spacing, is a measure of the distance where the signals on different carriers start to "walk-off" as a result of chromatic dispersion.

Unfortunately, except for the cases where either dispersion or nonlinearity is negligible, no closed-form solution has been found for the nonlinear Schrödinger equation and approximate or numerical methods must be used to characterize the channel. For example, the Split-step Fourier method estimates the crosstalk caused by the nonlinearity by numerical techniques. However, in order to study the information-theoretic characteristics of the channel, analytic expressions for the input–output relation of the channel are needed. Peddanarappagari *et al.* used Volterra series in [7] to derive a series expansion to arbitrary order for the low-pass equivalent field in the frequency domain. This method converges for small signal intensities, where $\gamma\left|A\right|^2 L_{\text{eff}} \ll 1$, which is the region we are interested in. While this is a significant limitation, it is the only known technique that gives an analytical solution and we will use it in this work. Fortunately, major parts of our analysis do not depend on the exact modeling of the channel, and result from the fundamental properties of the transmission medium.

The Volterra series solution for the Fourier transform of the output low-pass equivalent field $A(\omega,z)$, in terms of the input low-pass equivalent field $A_0(\omega)$, up to the third-order term can be written as [7]

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

where H_1 and H_3, given in Appendix II, are the linear and cubic terms of the channel response, respectively. Although the Fourier transforms are originally performed with respect to τ, the solution in terms of the time t of the receiver reference frame has exactly the same form, since τ and t differ only by a constant delay. 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 by using an optical amplifier for the attenuation and dispersion compensation fiber (DCF) or electronic equalization 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. In practice, this effect dominates all other sources of noise in the channel, including signal-dependent shot noise, which is present in a standard Poisson channel. Fortunately, the equivalent low-pass linear frequency response of the channel has a constant amplitude over the frequency band that we consider, so that the equalization does not change the power spectral density of the signal and noise. Hence, we can assume that the additive optical noise remains white after the DCF.

By examining the expression for the cubic and fifth-order terms of the Volterra series in [7], we observe that the ratios of the magnitudes of these terms are, respectively, order $O(\delta)$ and order $O\left(\delta^2\right)$, where $\delta \triangleq \gamma|A|^2 L_{\text{eff}}$. The parameter 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 than or comparable to δ^2 in magnitude, leaving only the linear and cubic terms listed in (2). Consequently, the results will have a relative uncertainty of $O(\delta^2)$. We will estimate the uncertainty on the capacity due to this approximation in Section VI.

Remark 1: In this paper, the "big-O" notation implies a stronger meaning than its standard definition in asymptotic analysis. We write

$$f(z) = O\left(g(z)\right), \qquad z \in D \tag{3}$$

if there is a constant $\kappa > 0$ such that

$$|f(z)| \leq \kappa |g(z)|, \qquad z \in D \tag{4}$$

and

$$\kappa \not\gg 1. \tag{5}$$

In other words, $|f(z)|$ is smaller than or comparable to $|g(z)|$. In this paper, unless otherwise stated, the ordering arguments are in terms of δ, i.e., $z = \delta$, in the region $D = \{\delta \ll 1\}$.

B. Wavelength-Division Multiplexing

We consider a WDM system with K users transmitting in different subchannels with identical signal spectrum, and we study two cases for the receiver. In the first case, which corresponds to a multiple-access channel model, a single receiver has access to the fiber output in all frequency bands. In the second case, each user only has access to his own subchannel, thus making the system an interference channel. The multiuser model we use is also valid for current point-to-point (single-user) WDM systems where no cooperation or coding is performed between the subchannels. We further 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. 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 by a linear equalizer. Moreover, quadratic dispersion may produce linear intersymbol interference (ISI) within a single subchannel by broadening the transmitted pulses in time. Although the linear ISI can be compensated, the nonlinearity results in intermodulation between consecutive symbols of each channel. Consequently, the coupling between dispersion and nonlinearity produces memory in the channel. In order to gain insight, we first neglect the relative walk-off of the subchannels by assuming that the dispersion is weak, or there is sufficient guard time between the consecutive symbols of each user. In this case, the channel is memoryless, because during each symbol period only one symbol from each user contributes to the output signal. In Section IV, we will generalize the problem to the case with memory and show that channel memory does not reduce the capacity, although it makes the optimal receiver more complicated.

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

$$A_n(\omega) = \sum_{k=1}^{K} g_k u_k(n) V(\omega - k\Delta\omega) \tag{6}$$

where $u_k(n), n = 1, \ldots, N$ 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 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

$$\frac{1}{N} \sum_{n=1}^{N} E\left[|u_k(n)|^2\right] \leq 1. \tag{7}$$

In other words, the kth user transmits with an average energy per symbol duration upper-bounded by $|g_k|^2$. For brevity, we drop the index n in the analysis of the memoryless case.

As a result of nonlinear effects, each subchannel experiences spectral broadening. A practical assumption that simplifies the analysis is that the spacing between the carrier frequencies $\Delta\nu$ is large enough with respect to the bandwidth of the signals B, so that the spectrum of each subchannel after nonlinear mixing does not overlap that of any other subchannel. With the current technology, channel spacing is several times the bandwidth of each channel because of practical limitations in the frequency stability and bandwidth of optical filters. On the other hand, from (2), we see that the nonlinear term in the output is a triple convolution of transmitted signals in the frequency domain, so its bandwidth is strictly less than three times the bandwidth of the linear term. In fact, the effective bandwidth is much smaller, e.g., if the pulses are Gaussian, the effective bandwidth of the crosstalk will be $\sqrt{3}$ times that of the original signal. Hence, for current systems, the assumption about carrier spacing does not impose any constraint on the system. Moreover, for future more spectrally efficient systems, the capacity-achieving interference cancellation scheme that we study in Section III-C can perform well even if the spectra of the subchannels overlap due to nonlinearity. Therefore, we believe that the overlap of subchannels, although making the equations much more complicated, does not change the results significantly.

The first step in analyzing the capacity of the system is to derive a discrete-time equivalent model for the channel. This can be done by finding a complete set of orthonormal basis functions for the space containing all the possible received signals, and then projecting the received signal along each of the basis functions. In the absence of nonlinearity, it is optimal to choose as the basis functions the responses of the channel (including the possible dispersion compensator) to the waveforms transmitted by the K users, i.e., to use a bank of K electrical matched filters, each followed by a sampler. An equivalent diagram of this structure is shown in Fig. 1. The down-conversion can be either homodyne or heterodyne and is done by adding a locally generated tone to the received signal and passing it through a square-law detector. If a balanced receiver is used and the local oscillator is strong enough, this process results in a frequency shift in the received signal, without changing the characteristics of the signal and the noise.

For the nonlinear optical fiber channel, the number of filters required for spanning the space of the received signal is more

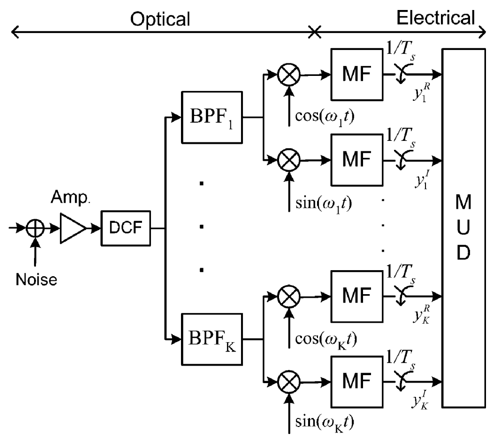

Fig. 1. Equivalent receiver structure (MF denotes the electronic matched filter).

than the number of subchannels, K. However, since we are assuming that the crosstalk is much smaller than the signal term, almost all the useful information can be provided by the outputs of the K matched filters. In Appendix I, the optimality of this structure to the first order of approximation is shown for both the memoryless case and the case with channel memory. Therefore, we use the structure of Fig. 1 for the (weakly) nonlinear channel.

We initially assume that the first-order dispersion is weak, so that the channel can be considered memoryless. In this case, using (2) and (6), the output of the ith branch tuned to the ith user's signal can be written as

$$y_i = x_i + \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K} \xi_{k,l,m}^{(i)} x_k x_l^* x_m + n_i \tag{8}$$

where $\{n_i\}$ are the additive noise terms modeled as independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian random variables with zero mean and variance σ^2 in both the real and imaginary dimensions. In this equation, the triple summation corresponds to the crosstalk terms coming from the Kerr nonlinearity. The crosstalk coefficients $\xi_{k,l,m}^{(i)}$ are proportional to γL_{eff}, and can be calculated from (2) and (6). Since the frequency bands are nonoverlapping, $\xi_{k,l,m}^{(i)}$ is only nonzero when $k-l+m=i$. Details on the derivation of this model are presented in Appendix II. In Section IV, where we study the effect of memory on the capacity, we will discuss how this model can be generalized.

The effects that give rise to the interference terms in (8) are often classified in the literature into three categories. If $k=l=m=i$, then the corresponding term is caused by self-phase modulation (SPM). XPM produces the terms for which $k=l$ and $m=i$, or $k=i$ and $l=m$, and the rest of the terms are 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, with a strongly dispersive fiber, i.e., where $L_{\text{D}} < L_{\text{eff}}$, as is the case for practical systems, FWM is much smaller than XPM, and it can be neglected. However, as mentioned before, strong dispersion also introduces memory into the channel.

Now we can rewrite (8) as

$$y_i = x_i + \kappa_i + \varphi_i + n_i \tag{9}$$

where

$$\varphi_i = \sum\sum\sum_{\substack{k-l+m=i \\ k\neq i\neq m}} \xi_{k,l,m}^{(i)} x_k x_l^* x_m \tag{10}$$

contains the FWM terms and

$$\kappa_i = \sum_{k=1}^{K} \left(\xi_{k,k,i}^{(i)} + \xi_{i,k,k}^{(i)}\right) |x_k|^2 x_i \tag{11}$$

contains the SPM and XPM terms.

By computing the coefficients $\xi_{k,k,i}^{(i)}$ and $\xi_{i,k,k}^{(i)}$ to obtain $\{\kappa_i\}$, it can be observed that both have negligible real parts for any i and k. Hence, (11) can be rewritten as

$$\kappa_i = x_i \sum_{k=1}^{K} j\rho_k^{(i)} |x_k|^2 \tag{12}$$

where

$$\rho_k^{(i)} \triangleq -j\left(\xi_{k,k,i}^{(i)} + \xi_{i,k,k}^{(i)}\right)$$

is a real constant. This means that in the signal space, SPM and XPM are orthogonal to the signal term and therefore they act as phase distortions on the signal to the first order of approximation. This directly results from the symmetry between any two users and the fact that the energy loss in the fiber is only a function of α and the fiber length, hence the total energy at the receiver is not affected by dispersion or fiber nonlinearity.

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.

III. Capacity Region of the Memoryless Channel

In this section, we study the capacity region of the memoryless system described by (8) up to the first order of approximation, i.e., by neglecting all terms of order $O(\delta^2)$. Particularly, all the terms resulting from the square of crosstalk are negligible compared to the square of the linear (signal) term, but are not negligible with respect to the square of noise in the absence of the linear term. This is a valid approximation, since the nonlinear terms become significant at high signal-to-noise values, where the magnitude of the crosstalk, even though much smaller than the linear term, may be comparable to or even larger than the additive optical noise.

We group the inputs and outputs of the channel into two real $2K\times 1$ vectors $X = \left[\underline{x}_1^T \cdots \underline{x}_K^T\right]^T$ and $Y = \left[\underline{y}_1^T \cdots \underline{y}_K^T\right]^T$, 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 (8) in a vector form

$$Y = X + \Theta + Z \tag{13}$$

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 (8) is

$$\{(R_1,\ldots,R_K) : \forall k=1,\ldots,K, \quad 0 \le R_k \le \log\left(1+P_k/2\sigma^2\right)\}. \quad (14)$$

Remark 2: Note that (14) is the same as the capacity region of the linear channel. In other words, Proposition 1 states that, to the first order of approximation, nonlinearity does not affect the capacity region of the channel.

Proof: We will prove the proposition in two steps. We will first show that (14) is an outer bound on the capacity region, and then show that this bound is achievable.

A. 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 [10, pp. 389–390], we have

$$\forall S \subset \{1,\ldots,K\} : 0 \le \sum_{i\in S} R_i \le I(X_S; Y|X_{S'}) \quad (15)$$

where S' is the complement of S with respect to the reference set $\{1,\ldots,K\}$, and V_A for any vector V and set A is a subvector of V obtained by keeping only the elements corresponding to users whose indices belong to A. When evaluating $I(X_S;Y|X_{S'})$, we refer to users with indices in S' as *inactive* users, as their rate of transmission is not a concern. For the right-hand side of (15), using the chain rule for mutual information, we can write

$$I(X_S;Y|X_{S'}) = I(X_S;Y_S|X_{S'}) + I(X_S;Y_{S'}|X_{S'},Y_S). \quad (16)$$

This is the total amount of information that can be transferred if only the users in S are communicating through the channel. We characterize the first term on the right-hand side, and in Appendix III we show that the second term is $O(\delta^2)$, and therefore can be neglected. This means that looking at the outputs of channels corresponding to inactive users does not provide any extra information, even if the inactive users collaborate by transmitting deterministic signals.

By expanding the first term on the right-hand side of (16) we obtain

$$I(X_S;Y_S|X_{S'}) = h(Y_S|X_{S'}) - h(Y_S|X) \quad (17)$$

where h denotes the differential entropy. Since the interference terms are deterministic functions of $\{x_i\}$, the second term reduces to

$$h(Z_S) = \frac{1}{2}\log\left((2\pi e)^{2|S|}\left|\det\left(E\left[Z_S Z_S{}^T\right]\right)\right|\right) = |S|\log\left(2\pi e\sigma^2\right) \quad (18)$$

where we have used the fact that the elements of Z, i.e., noise terms, are Gaussian i.i.d. random variables. This expression is independent of the distribution of $\{x_i\}$, hence to maximize the mutual information it suffices to maximize the first term in (17) with respect to the probability distribution of X. This term satisfies

$$h(Y_S|X_{S'}) \le \sum_{i\in S} h(y_i|X_{S'}) \quad (19)$$

with equality if $\{y_i\}$ are independent. Furthermore, for each $i \in S$ we have [10, p. 234]

$$h(y_i|X_{S'}) = E_\chi\left[h(y_i|X_{S'}=\chi)\right] \le E_\chi\left[\frac{1}{2}\log\left((2\pi e)^2\det\left(\operatorname{cov}\left[\underline{y}_i|X_{S'}=\chi\right]\right)\right)\right] \quad (20)$$

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

To simplify the expressions, we introduce the notations

$$p_i^R \triangleq E[(x_i^R)^2],\ v_i^R \triangleq \left(\operatorname{var}[x_i^R]\right)^{\frac{1}{2}},\ \mu_i^R \triangleq E[x_i^R]/v_i^R, \quad (21)$$

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

$$q_i \triangleq \operatorname{cov}[x_i^R, x_i^I]/v_i^R v_i^I, \quad (22)$$

to denote the correlation coefficient of x_i^R and x_i^I. Due to the power constraint, we have

$$p_i^R + p_i^I \le P_i. \quad (23)$$

Following the steps in Appendix IV, the determinant in (20) can be expanded as

$$\begin{aligned}\det\left(\operatorname{cov}\left[\underline{y}_i|X_{S'}=\chi\right]\right) &= \left(p_i^R+\sigma^2-c_i\right)\left(p_i^I+\sigma^2+c_i\right)\\ &\quad - \left[(v_i^R)^2(v_i^I)^2q_i^2 + q_iO(P^2\delta)\right]\\ &\quad - \left[(\mu_i^R)^2(v_i^R)^2p_i^I + \mu_i^RO(P^2\delta)\right]\\ &\quad - \left[(\mu_i^I)^2(v_i^I)^2p_i^R + \mu_i^IO(P^2\delta)\right]\end{aligned} \quad (24)$$

where

$$c_i \triangleq E\left[x_i^R x_i^I \rho_i^{(i)}|x_i|^2\right], \quad (25)$$

and $\rho_i^{(i)}$, as used in (12), is the SPM coefficient for the ith channel. Using the inequality of arithmetic and geometric means (AM-GM inequality) and (23), the first term on the right-hand side of (24) can be upper-bounded by

$$\left[\frac{\left(p_i^R+\sigma^2-c_i\right)+\left(p_i^I+\sigma^2+c_i\right)}{2}\right]^2 \le \left(P_i/2+\sigma^2\right)^2 \quad (26)$$

with equality if $p_i^R = p_i^I = P_i/2$, and $c_i = 0$. The next step is to lower-bound the three bracketed expressions on the right-hand side of (24). All these expressions are of the form

$$az^2 + O(a\delta)z, \qquad a > 0 \quad (27)$$

where z, respectively, denotes q_i, μ_i^R, and μ_i^I in these expressions. Expression (27) is positive for $|z| \gg \delta$, and equals zero for $|z| = 0$. Also, for $|z| = O(\delta)$, (27) becomes $O(a\delta^2)$, or equivalently, $O(P^2\delta^2)$, which is negligible compared to the largest term in (24), which is $p_i^R p_i^I$. Hence, to the first order of approximation, all the three bracketed expressions in (24) are lower-bounded by zero. This observation along with (26) yields

$$\det\left(\operatorname{cov}\left[\underline{y}_i|X_{S'}=\chi\right]\right)\le\left(P_i/2+\sigma^2\right)^2. \tag{28}$$

Substituting this in (20) gives

$$h(y_i|X_{S'})\le\log\left(2\pi e\left(P_i/2+\sigma^2\right)\right). \tag{29}$$

This result can be combined with (15)–(19) to conclude that

$$0\le\sum_{i\in S}R_i\le\sum_{i\in S}\log\left(1+P_i/2\sigma^2\right),\quad\forall S\subset\{1,\ldots,K\}, \tag{30}$$

which can be rewritten in the form of (14), as an outer bound on the capacity region.

This bound implies that the introduction of crosstalk cannot improve the achievable rates of any user even if the other users transmit deterministic signals. This is due to the fact that making the real and imaginary parts of the symbols independent and zero-mean to maximize the entropy also makes all the FWM terms uncorrelated with the linear term, so they cannot change the mean signal power. The other essential property used in this analysis is that all the SPM and XPM coefficients are imaginary; hence they do not affect the signal magnitude.

B. Achievability of the Bound

To show the achievability of the bound, we first look at the general case where SPM, XPM, and FWM are present. Starting from (15), we select the sequence of the channel coded symbols generated by each user $i=1,\ldots,K$, i.e., $\{u_i\}$, to be i.i.d. circularly symmetric complex Gaussian random variables with zero mean and variance $1/2$ per complex dimension. Multiplying by a constant g_i only changes the variance of the symbols, such that x_i^R and x_i^I will become i.i.d. and Gaussian with variance $P_i/2$ and vanishing odd moments. Hence, we have

$$\mu_i^R=\mu_i^I=q_i=0 \tag{31}$$

and also

$$E\left[x_i^Rx_i^I\rho_i^{(i)}|x_i|^2\right]=\rho_i^{(i)}\left(E\left[(x_i^R)^3x_i^I\right]+E\left[x_i^R(x_i^I)^3\right]\right)=0. \tag{32}$$

Using (31) and (32) in (24) makes (28) an equality.

Now to achieve equality in (19) and (20), we must show that the received samples $y_i,\ i\in S$ are independent complex Gaussian random variables, given $\{x_j,\ j\in S'\}$.

1) Negligible FWM: Let us first assume that SPM and XPM are the dominant crosstalk terms. While this assumption is not realistic for the memoryless channel, the result obtained for this scenario can be directly applied to derive the capacity of the more realistic strongly dispersive channel, where FWM is in fact negligible.

Neglecting the FWM terms, for an arbitrary channel i we have

$$y_i=v_i+n_i \tag{33}$$

where v_i, defined as

$$v_i\triangleq x_i+\sum_{k=1}^{K}j\rho_k^{(i)}|x_k|^2x_i = x_i\left(1+j\sum_{k=1}^{K}\rho_k^{(i)}|x_k|^2\right), \tag{34}$$

contains the signal and crosstalk terms, and is independent of the Gaussian noise n_i. Since $X_{S'}$ is given, we can treat $\{x_j,\ j\in S'\}$ as constants. We first prove that v_i has a complex Gaussian distribution, which makes y_i Gaussian. Since $\{\rho_k^{(i)}\}$ are real, the second term inside the parentheses in (34) becomes imaginary. Hence, the magnitude of v_i is equal to

$$|v_i|=|x_i|\left(1+\left|\sum_{k=1}^{K}\rho_k^{(i)}|x_k|^2\right|^2\right)^{1/2} = |x_i| \tag{35}$$

where we have neglected the square of the crosstalk, which is second order. Hence, $|v_i|$ has the same distribution as $|x_i|$, i.e., Rayleigh. Let $\arg(z)\in[0,2\pi)$ be the phase of a complex variable z. From (34)

$$\arg(v_i)=\arg(x_i)\oplus\arg\left(1+j\sum_{k=1}^{K}\rho_k^{(i)}|x_k|^2\right) \tag{36}$$

where $\oplus$ denotes addition modulo 2π. Since x_i has a uniform phase distribution, independent of its absolute value, the same is true for v_i, as well. This follows from the fact that the first term in (36) is independent of the second term, and has the uniform distribution on $[0,2\pi)$. As a result, since the magnitude and phase of v_i are independent, and, respectively, have Rayleigh and uniform distributions, v_i is Gaussian.

To prove that $\{v_i\}$ are mutually independent, it is sufficient to show that the phases and magnitudes of $\{v_i\}$ are all mutually independent. Similar to the previous argument, we see from (35) and (36) that each $|v_i|$ only depends upon $|x_i|$, and $\arg(v_i)$ is independent of $|x_k|$ for every k and $\arg(x_k)$ for $k\neq i$. This, along with the mutual independence of phases and absolute values of $\{x_i\}$, verifies the mutual independence of $\{v_i\}$, and since each v_i has a Gaussian distribution, they are jointly Gaussian. Finally, $\{y_i\}$ are jointly Gaussian, as the noise terms are jointly Gaussian and independent of $\{v_i\}$. This completes the proof of achievability for this case.

To observe the effect of FWM, we assume for simplicity that FWM is the dominant term, and also that the dependent FWM terms (defined in Appendix IV) are negligible, so that FWM in each y_i becomes independent of the linear term, x_i. In this case, we can write

$$y_i=x_i+\varphi_i+n_i \tag{37}$$

where φ_i, as defined in (10), contains the FWM terms. The three terms on the right-hand side are independent, and the first and the third are Gaussian. Hence, we need φ_i to be Gaussian, as well, in order for y_i to become Gaussian. While this is not true in general, it will be shown in Section V that the distribution of φ_i is asymptotically Gaussian as the number of users increases.

The next step is to show that $\{y_i\}$ are mutually independent. Since all $\{x_i\}$ and $\{n_i\}$ are mutually independent, we also need each φ_i to be mutually independent of $\{x_j,\ j \neq i\}$. However, this is not true, since φ_i is a deterministic function of the set $\{x_j,\ j \neq i\}$. This means that $\{y_i\}$ are not independent, and hence the outer bound cannot be achieved with the previous approach. In the next subsection, we present a more general analysis to achieve the bound, which applies to the case where SPM, XPM, and FWM all exist.

2) General Case: Now, we show the achievability of (14) in 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 us assume that user i is the user of interest. Recall the expression (8) for the channel output samples. Now, we form the test statistic

$$z_i = y_i - \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K} \xi_{k,l,m}^{(i)} y_k y_l^* y_m. \tag{38}$$

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 (38) using (8), and neglecting all the higher order terms, we obtain

$$\begin{aligned} z_i = x_i + n_i &- \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K} \xi_{k,l,m}^{(i)} (n_k+\theta_k) x_l^* x_m \\ &- \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K} \xi_{k,l,m}^{(i)} x_k (n_l^*+\theta_l^*) x_m \\ &- \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{m=1}^{K} \xi_{k,l,m}^{(i)} x_k x_l^* (n_m+\theta_m) \end{aligned} \tag{39}$$

where θ_k contains all the crosstalk terms on channel k. Finding the capacities of the channels with input–output pairs $(u_i, z_i),\ i = 1, \ldots, K$ gives a set of achievable rates for the original channel. As mentioned before, $\{x_i\}$ form sufficient statistics for the channel-coded symbols $\{u_i\}$, hence the rate of information communicated by any (u_i, z_i) pair can be written as

$$I(x_i; z_i) = h(z_i) - h(z_i|x_i). \tag{40}$$

Similar to the previous subsection, we assume that the channel-coded symbols generated by each user are i.i.d. Gaussian random variables.

In calculating $h(z_i)$, the only dominant terms are the signal and noise, since the residual crosstalk after cancellation is at least two orders of magnitude smaller than the linear term x_i. Hence, we have

$$h(z_i) = \log\left(2\pi e (P_i/2 + \sigma^2)\right). \tag{41}$$

However, $h(z_i|x_i)$ should be computed more carefully, since, unlike the previous cases, the residual crosstalk is not a deterministic function of x_i, and hence can contribute to $h(z_i|x_i)$. Also, crosstalk terms may not be negligible compared to the noise term, which, in the absence of the signal, is the dominant term. Using the Gaussian bound again, we have

$$\begin{aligned} h(z_i|x_i) &\le h(z_i^R|x_i) + h(z_i^I|x_i) \\ &\le E_\chi\left\{\frac{1}{2}\log\left((2\pi e)^2\ \mathrm{var}[z_i^R|x_i=\chi]\mathrm{var}[z_i^I|x_i=\chi]\right)\right\} \end{aligned} \tag{42}$$

with equality if z_i^R and z_i^I are independent Gaussian random variables given x_i. Now, to compute $\mathrm{var}[z_i^R|x_i]$ and $\mathrm{var}[z_i^I|x_i]$, observe that the variance of crosstalk in (39) is $O\left((\sigma^2 + P\delta^2)\delta^2\right)$ which is negligible compared to the variance of the noise, σ^2. Hence, the dominant terms are the variance of the noise, and the covariance of the noise with the residual crosstalk. The only crosstalk terms that can have nonzero correlation with the noise are those which contain n_i or n_i^*, with the rest being independent of n_i, and hence not contributing to the covariance.

We define ζ_i as the collection of terms in z_i that contribute to $h(z_i|x_i)$, and factor these terms by n_i and n_i^* to obtain

$$\begin{aligned} \zeta_i &= n_i + n_i \sum_{k=1}^{K} j\rho_k^{(i)} |x_k|^2 + n_i^* \sum\sum_{k+m=2i} \xi_{k,i,m}^{(i)} x_k x_m \\ &= n_i\left(1 + \sum_{k=1}^{K} j\rho_k^{(i)}|x_k|^2\right) + n_i^* \sum\sum_{k+m=2i} \xi_{k,i,m}^{(i)} x_k x_m. \end{aligned} \tag{43}$$

Now, to first order, we can write

$$\begin{aligned} \mathrm{var}[z_i^R|x_i=\chi]&\mathrm{var}[z_i^I|x_i=\chi] \\ &= \mathrm{var}[\zeta_i^R|x_i=\chi]\mathrm{var}[\zeta_i^I|x_i=\chi] \\ &\le E\left[(\zeta_i^R)^2|x_i=\chi\right] E\left[(\zeta_i^I)^2|x_i=\chi\right] \\ &\le 1/4 E\left[|\zeta_i|^2\right]^2. \end{aligned} \tag{44}$$

Using the fact that the absolute value of the term inside the parentheses in (43) is equal to one (cf. (35)), we have

$$\begin{aligned} E\left[|\zeta_i|^2\right] &= E\left[|n_i|^2\right] + E\left[n_i^2\right] E\left[\left(\sum\sum_{k+m=2i} \xi_{k,i,m}^{(i)} x_k x_m\right)^2\right] \\ &= 2\sigma^2 \end{aligned} \tag{45}$$

where we have neglected the square of the second term in (43). Hence, combining (45) with (44) and (42), we obtain

$$h(z_i|x_i) \le \log(2\pi e \sigma^2). \tag{46}$$

Finally, by substituting (46) into (40), we conclude that the rate

$$I(x_i; z_i) = \log(1 + P_i/2\sigma^2) \tag{47}$$

is achievable.

This result implies that, even with a simple interference cancellation scheme, the outer bound of Section III-A is achievable. Hence, to the first order of approximation, (14) is the capacity region of the channel. □

Remark 3: In this section, we proved that every point in the capacity region can be achieved by mutually independent Gaussian inputs, independent of the channel parameters. This means that, although for the capacity analysis we assumed that the transmitters know the gains and crosstalk coefficients, they can achieve the capacity without using this information by only knowing their corresponding SNRs at the receiver.

IV. Capacity Region of the Channel With Memory

In a channel with strong dispersion, the relative walk-off between the signals traveling at different frequencies along with nonlinear mixing introduces (finite) memory to the channel. In this case, (8) must be rewritten by adding time to the triple summation on the right-hand side. As mentioned in Section II, another effect of strong dispersion is to suppress the FWM, such that in many practical cases FWM can be neglected. Specifically, it can be shown that the dominant effect in this regime is a generalized form of XPM over time and frequency, while the mixing coefficients are imaginary, as before. More precisely, we can write

$$y_i(n)=x_i(n)+\sum_{p=-M}^{M}\sum_{k=1}^{K} j\rho_k^{(i)}(p)|x_k(n-p)|^2 x_i(n)+n_i(n) \quad (48)$$

where M is the two-sided memory of the channel, $n_i(n)$ for $i=1,\dots,K$ and $n=1,\dots,N$ are i.i.d. complex Gaussian noise samples with variance σ^2 per complex dimension, and $\{\rho_k^{(i)}(p)\}$ are real constants. If the dispersion is not strong enough to suppress FWM completely, then the expression reflecting the presence of FWM can also be derived from (8) in a similar way. We assume that users are frame-synchronous and there is a guard time between every two blocks of coded symbols, so that there is no interblock interference. As the block length N becomes large, the overhead due to this guard time becomes negligible.

The number of symbols from channel k that affect a symbol in channel i grows with the difference between the carrier frequencies of these channels. The reason is that increasing the spacing between two carriers also increases the difference in the propagation speed of signals in those channels, and hence each symbol in one channel will be affected by more symbols from the other channel.

To derive the capacity region of this channel, we use the following theorem.

Theorem 1 (Verdú, [11]): The capacity region of a frame-synchronous multiple-access channel with finite memory is given by

$$C=\text{Closure}\left(\liminf_{N\to\infty}\frac{1}{N}C_N\right). \quad (49)$$

where

$$C_N=\bigcap_{S\subset\{1,\dots,K\}}\Big\{(R_1,\dots,R_K):\; 0\le\sum_{k\in S}R_k\le I(X_S^N;Y^N|X_{S'}^N)\Big\}. \quad (50)$$

The following proposition determines the capacity region of the channel with memory.

Proposition 2: To the first order of approximation on the nonlinearity, the capacity region of the coherent WDM channel with memory is the region given by (14).

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 $S\subset\{1,\dots,K\}$. As before, we drop the subscript if $S=\{1,\dots,K\}$. Since the users are frame-synchronous and the channel has a finite memory, we can use Theorem 1 to derive the capacity region.

The channel model in the presence of memory can be visualized as a simple generalization of the memoryless case, where the vectors are replaced by matrices in (13), 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 in Section II, we can write

$$\begin{aligned}I(X_S^N;Y^N|X_{S'}^N)&=I(X_S^N;Y_S^N|X_{S'}^N)+I(X_S^N;Y_{S'}^N|X_{S'}^N,Y_S^N)\\&\approx I(X_S^N;Y_S^N|X_{S'}^N).\end{aligned} \quad (51)$$

Furthermore, similar to (30), we have

$$I(X_S^N;Y_S^N|X_{S'}^N)\le\sum_{n=1}^{N}\sum_{i\in S}\log\left(1+\frac{1}{2}\sigma^{-2}E\left[|x_i(n)|^2\right]\right), \quad (52)$$

or in another form

$$I(X_S^N;Y_S^N|X_{S'}^N)\le\sum_{i\in S}\log\left(\prod_{n=1}^{N}\left(1+\frac{1}{2}\sigma^{-2}E\left[|x_i(n)|^2\right]\right)\right). \quad (53)$$

Inside the logarithm, we are multiplying N terms whose sum is upper-bounded due to the power constraint (7). Hence, this product will be maximized if all the symbols of each user have equal powers. Then, we will have

$$I(X_S^N;Y_S^N|X_{S'}^N)\le\sum_{i\in S}\log\left(\left(1+P_i/2\sigma^2\right)^N\right). \quad (54)$$

As in Section III, it can be shown that this bound is achievable if the symbols transmitted in different time slots and subchannels are all independent and circularly symmetric complex Gaussian random variables. Finally, C_N is given by

$$C_N=\bigcap_{S\subset\{1,\dots,K\}}\Big\{(R_1,\dots,R_K):\; 0\le\sum_{k\in S}R_k\le N\sum_{k\in S}\log\left(1+P_k/2\sigma^2\right)\Big\}. \quad (55)$$

This, along with (49), yields

$$C=\bigcap_{S\subset\{1,\dots,K\}}\Big\{(R_1,\dots,R_K):\; 0\le\sum_{k\in S}R_k\le\sum_{k\in S}\log\left(1+P_k/2\sigma^2\right)\Big\} \quad (56)$$

which is equivalent to (14). □

Remark 4: Proposition 2 states that even the channel memory does not affect the first-order capacity region of the nonlinear WDM channel. However, including memory does increase the complexity of the optimum detection scheme.

V. Capacity With Single-Wavelength Detection

We showed that with optimal detection in a multiple-access model of the fiber channel, first-order nonlinearity does not limit the achievable rates of transmission. Now, we look at the capacity region when the receiver only looks at one of the WDM subchannels to detect the signal of interest. This problem is an example of an interference channel. In this type of multiuser channel, each user, by looking only at his own wavelength channel, should in principle be able to collect some information about the interfering signals, knowing that they are coded data sequences, and use this information to improve the detection. Unfortunately, no general solution is known for the capacity of interference channels, and the capacity region is characterized by bounds. In the case of Gaussian channels with weak interference, the best known estimate of the capacity region is the inner bound achieved by treating the interfering signals as noise [12]. This bound is asymptotically tight because it is not possible to make a good estimate of the other users' signals from the weak interference.

In our problem, since the interference is assumed to be weak, the same argument can be made. Therefore, we define single-wavelength detection (SWD) as the strategy whereby the information transmitted by user k is decoded by only using the output of the kth channel and treating the crosstalk from other channels as random interference, without any knowledge of the codebooks of other users. Although each user detects its signal independently without attempting to decode the interference, the notion of multiuser capacity region should still be considered, because the distribution of the coded symbols transmitted by each user affects the permissible rates of other users by changing the interference distribution. In order to simplify the analysis, we only characterize the sum-rate capacity of the channel, and assume that the channel is memoryless. However, the results can be easily generalized to the channel with memory, by reasoning similar to that applied in Section IV.

Using SWD, the sum-rate capacity of the channel is equal to

$$C_{\text{sum}} = \sum_{i=1}^{K} I(x_i; y_i). \tag{57}$$

Now, $I(x_i; y_i)$ can be expanded as

$$I(x_i; y_i) = h(y_i) - h(y_i|x_i). \tag{58}$$

Here, unlike in the multiple-access case, despite knowing the desired input, we cannot completely estimate and cancel the interference. Hence, the second term in (58) is not a constant, in contrast to (18). Specifically, using (8) we have

$$\begin{aligned} h(y_i|x_i) &= E_u\left[h(y_i|x_i = u)\right] \\ &= E_u\left[h\Big(\sum\sum_{k-l+m=i}\sum \xi_{k,l,m}^{(i)} x_k x_l^* x_m + n_i\Big)\Big| x_i = u\right] \end{aligned} \tag{59}$$

which depends on the distribution of all users.

Finding the capacity region (57) is complicated, because we need to search over all probability distributions of sources to find the admissible rates. Note that even maximizing (58), which is itself difficult, does not necessarily result in the actual capacity region. Therefore, in order to derive analytical estimates, we confine our analysis to the two regimes where either FWM or XPM is dominant.

We first study the case in which FWM is the dominant nonlinear effect. In this case, the total interference power in each wavelength is evenly distributed among the roughly N^2 nonzero crosstalk terms in (8), in the sense that the sum over any group of $O(N)$ crosstalk terms is small compared to the total sum. Having this property, we show that the distribution of the total crosstalk approaches a complex Gaussian random variable, with the help of the properties of weighted U-statistics. Given a sequence of n real and i.i.d. random variables $\{X_i\}$

$$U_n^p = \sum_J w(J)\psi(X_{j_1}, \dots, X_{j_p}) \tag{60}$$

is a weighted U-statistic of order p, where $p \le n$ is the number of symbols contributing to each term, and the summation is over all the choices $J = (j_1, \dots, j_p)$ of p indices from $1, \dots, n$ such that $j_1 \ne \cdots \ne j_p$. Here, ψ is a given function which is symmetric under permutations of its arguments, and the weights w are functions of the summation indices. It is shown in [13] that under certain conditions for ensuring a sufficiently weak dependence between the summands in (60), U_n^p for fixed p approaches a Gaussian random variable as n, the number of random variables, becomes large.

To apply this result to the crosstalk terms in (59), we assume that all the channel-coded symbols $\{u_k\}$ have the same distribution, and that the real and imaginary parts of each u_k are i.i.d. and zero-mean. The difference in the transmission powers is reflected in $\{g_k\}$. The only exception to this i.i.d. assumption is the ith user's symbol, since in (59) we are conditioning on x_i, or equivalently, u_i. However, there are only $O(N)$ terms containing this symbol, and hence they can be neglected compared to the total sum of the crosstalk terms. Now, we can consider the real and imaginary parts of the nonlinear crosstalk in channel i as two U-statistics, where $\{X_k\}$ in (60) are the real and imaginary parts of the channel-coded symbols $\{u_k^R, u_k^I\}$. In this analogy, $n = 2K$, $p = 3$, $\psi(X_k, X_l, X_m) = X_k X_l X_m$, and $w(X_k, X_l, X_m) = 0$ if $k - l + m \ne i$. We conclude that the real and imaginary parts of the crosstalk each approach a Gaussian as the number of channels K increases. This also holds for any linear combination of them, thus making them jointly Gaussian. Furthermore, due to the circular symmetry of the distribution of $\{u_k\}$, and the form of the crosstalk terms, it is not difficult to check that the real and imaginary parts of the crosstalk are uncorrelated and have equal variances. As a result, the crosstalk sum in (59) approaches a circularly symmetric complex Gaussian random variable.

Using this result, for large K we can compute the conditional entropy (59) by evaluating the variance of the crosstalk and noise, which can be simplified as

$$\begin{aligned} h(y_i|x_i) &= \log(2\pi e\sigma^2) \\ &\quad + \log\left(1 + \frac{1}{2\sigma^2}\sum\sum_{k-l+m=i}\sum |\xi_{k,l,m}^{(i)}|^2 p_k p_l^* p_m\right) \end{aligned} \tag{61}$$

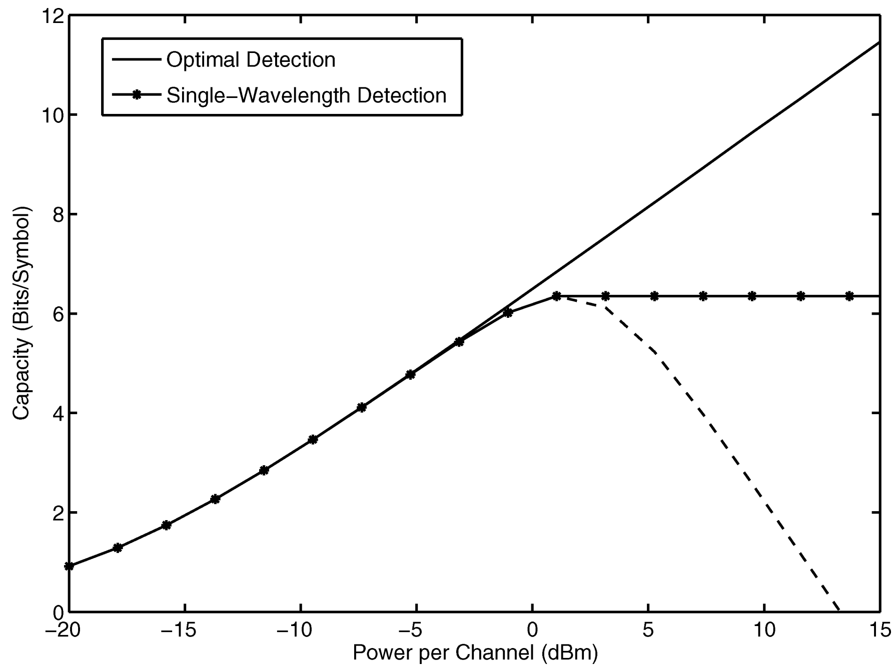

Fig. 2. Capacity versus transmit power per channel. Gaussian pulses with bandwidth 80 GHz, $\gamma = 1\ \mathrm{W}^{-1}\mathrm{km}^{-1}$, $\beta = -20\ \mathrm{ps}^2/\mathrm{km}$, $\alpha = 0.25\ \mathrm{dB/km}$. The dashed line corresponds to the mutual information with single-wavelength detection if the users are forced to transmit with the maximum power.

where p_k is the transmitted energy per symbol of user k, upper-bounded by P_k. Since this expression is independent of the signal distributions, it is enough to maximize the first term on the right-hand side of (58). As in the previous sections, this maximum occurs when the symbols are i.i.d. complex Gaussian random variables, and is equal to

$$h(y_i) = \log\left(2\pi e(p_i/2 + \sigma^2)\right). \quad (62)$$

Therefore, the maximum mutual information in (58) can be written as

$$I(x_i; y_i) = \log(1 + p_i/2\sigma^2) - \log\left(1 + \frac{1}{2\sigma^2}\sum\sum_{k-l+m=i}\sum |\xi_{k,l,m}^{(i)}|^2 p_k p_l^* p_m\right) \quad (63)$$

which, along with (57), gives the set of permissible sum rates in terms of $\{p_k\}$.

In the second case, in which XPM is the dominant nonlinear effect, the conditional entropy (59) can be rewritten as

$$h(y_i|x_i) = E_u\left[h\left(jx_i \sum_{k\in S}\rho_k^{(i)}|x_k|^2 + n_i\right)\middle| x_i = u\right]. \quad (64)$$

We can use the standard Central Limit Theorem for independent summands to approximate the summation on the right-hand side as a real Gaussian random variable for large K. Then, by rotating the complex coordinate axes for fixed u so that the interference term only appears in the real dimension, it is easy to show that (59) reduces to

$$h(y_i|x_i) = E_u\left[\frac{1}{2}\log\left(2\pi e\left[\sigma^2 + |u|^2\sum_{k\in S}(\rho_k^{(i)})^2 \mathrm{var}\left[|x_k|^2\right]\right]\right)\right] + \frac{1}{2}\log\left(2\pi e\sigma^2\right). \quad (65)$$

To maximize the mutual information in (58), a complicated optimization over the distribution of all users is still needed. A reasonable lower bound can be found by assuming that, as in the previous case, each x_i has a zero-mean complex Gaussian distribution with variance $p_i/2 \le 1/2$ per dimension. Then, from (65) we can write

$$h(y_i|x_i) = \log\left(2\pi e\sigma^2\right) + \int_0^\infty \frac{1}{2}\log\left(1 + \sigma^{-2}|u|^2\sum_{k\in S}(\rho_k^{(i)})^2 p_k^2\right) \cdot e^{-\frac{|u|^2}{p_i}}\,\frac{d|u|^2}{p_i} \quad (66)$$

where we used the fact that $|x_i|^2$ has a negative exponential distribution, and also that $E\left[|x_k|^4\right] = 2E\left[|x_k|^2\right]^2$ for each k. Finally, substituting (66) and (62) into (58) gives the expression for mutual information in terms of the transmission powers.

It should be emphasized that in both FWM- and XPM-limited cases, increasing each p_i does not necessarily improve the capacity region, as the effect of increased interference in other channels may dominate the improvement in the rate in channel i. Therefore, we need to optimize over the transmission powers to obtain the actual sum-rate capacity.

VI. Comparison and Discussion

To demonstrate the capacity gain of multiwavelength detection, we compare the average maximum rate per user for single-wavelength and multiwavelength detection using the results of Sections III and V. 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 P and equal 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 total crosstalk power to the signal power. For the values plotted in Fig. 2, the uncertainty is 1% at $P = 5$ dBm and 5% at $P = 9$ dBm.

It is observed that within the range of acceptable accuracy, i.e., where the uncertainty is much smaller than the difference between the results, the maximum achievable rate with the single-wavelength detection scheme saturates as the crosstalk power becomes comparable to the noise power. This effect results from the fact that the interference power in each channel increases as P^3. Thus, at high signal powers where the interference dominates the noise, the effective signal-to-interference-plus-noise ratio (SINR) behaves as P^{-2}; i.e., increasing the power does not improve the capacity. On the other hand, with the optimal scheme interference is totally canceled, and we can achieve the capacity of a linear fiber.

VII. Conclusion and Future Work

We derived the multiuser capacity region of WDM in a nonlinear fiber channel using a weak nonlinearity assumption. If the outputs of the fiber at all subchannels are used for detection, the channel with crosstalk caused by Kerr nonlinearity in the fiber will have the same capacity as a linear fiber channel. 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 subchannel is used, the capacity will saturate when the crosstalk dominates. We conclude that the crosstalk introduced by the Kerr nonlinearity does not severely limit the capacity of optical fibers, as long as the weak nonlinearity assumption is not violated.

Due to the nonlinear structure of the channel, optimal multi-wavelength detection for this channel is intractable, especially since the power-dependent nonlinear effects are important mostly at very high aggregate data rates, i.e., several gigabits per second. Applying iterative multiuser detection schemes seems to be the best solution to this problem, and studying the complexity–performance tradeoff of these schemes is a potential area for future work.

The focus of this work has been on optical communications with coherent transmission/reception. Although there is an increasing interest in coherent optical communications, at present, for practical reasons, noncoherent detection is more widely used in optical fibers. Hence, an important problem still to be studied is the question of how the results of this paper can be extended to noncoherent receivers. Furthermore, more accurate analytical models for the nonlinear channel are needed in order to design and analyze schemes that can perform well in real systems.

A simplifying assumption made for the analysis was that the subchannels remain nonoverlapping in the frequency domain, even though they experience spectral broadening. While this assumption is not restrictive in the weakly nonlinear regime, as the total transmission power continues to increase, not only the subchannels overlap, but also the total spectrum of the composite signal increases. Hence, the spectral efficiency is expected to experience a limit at very high power levels. Moreover, the impact of other types of nonlinearity, such as Raman scattering, should also be considered in this regime.

Appendix I
Optimality of Matched Filtering

In this appendix, we prove that the sampled outputs of the K filters, each matched to the response of the linear channel to the waveform transmitted in one of the K subchannels, provide all the useful information for decoding the symbols. In other words, the information that any other filter provides is redundant given the outputs of the matched filters.

In order to provide sufficient statistics for the received signal, we need to select a complete set of orthogonal basis functions for the space of the received signals. We will first discuss the memoryless case, and then comment on the case where the channel has memory. Moreover, for simplicity, we assume that XPM is the dominant nonlinear effect. As outlined at the end of this appendix, it can be shown with a similar, but more tedious, derivation that the same result holds also with the existence of FWM.

We adopt the K matched filters used for the linear channel response (which are orthogonal) as the first K basis functions and then find other orthonormal filters to span the complete space along with these filters. Clearly, the linear (zero-order) term of the output signal corresponding to the kth user is orthogonal to, and hence canceled by, all the filters except the kth matched filter. However, since the crosstalk term in the kth channel is nonlinear, it is not matched to the kth matched filter, and may contribute to the output of the other basis functions. Consequently, $\{y_k\}$, the sampled outputs of the matched filters, will each contain one linear signal term (if the kth user is transmitting anything), a number of third-order crosstalk terms, and an additive noise term. On the other hand, $\{w_i\}$, the sampled outputs of the rest of the basis functions, i.e., those not among the K matched filters, will only contain noise, and possibly some third-order crosstalk terms. For any w_i, we can write

$$w_i = x_{m(i)} \sum_{k=1}^{K} \lambda_k^{(m(i))} |x_k|^2 + \eta_i \tag{67}$$

where, for simplicity, we have assumed that the crosstalk terms in w_i are contributed by the XPM and SPM produced on only one subchannel, whose index is denoted by $m(i)$. A more general form for w_i can be obtained by including all the third-order combinations of $\{x_i\}$. Using a similar, but more tedious, analysis, it can be shown that the result derived here for (67) also holds for the general case.

We denote by W the vector containing all $\{w_i\}$, and suppose that we are interested in finding the sum-rate capacity of the channel. For the mutual information between the inputs and outputs of the channel, we have

$$I(X;Y,W) = I(X;Y) + I(X;W|Y), \tag{68}$$

and the second term in the right-hand side can be written as

$$\begin{aligned} h(W|Y) - h(W|X,Y) &= h(W|Y) - \sum_i h(\eta_i) \\ &\le \sum_i [h(w_i|Y) - h(\eta_i)] \end{aligned} \tag{69}$$

where we used the fact that $\{\eta_i\}$ are mutually independent. Each term inside the summation in (69) can be upper-bounded as

$$h(w_i|Y) - h(\eta_i) \leq \left[h(w_i^R|Y) - h(\eta_i^R)\right] + \left[h(w_i^I|Y) - h(\eta_i^I)\right] \quad (70)$$

since η_i^R and η_i^I are independent. Using a bound similar to (20), the first bracketed expression on the right-hand side can be written as

$$E_\Upsilon\left[h(w_i^R|Y=\Upsilon)-h(\eta_i^R)\right] \leq E_\Upsilon\left[\frac{1}{2}\log\left(\frac{\mathrm{var}\left[w_i^R|Y=\Upsilon\right]}{\sigma^2}\right)\right] \quad (71)$$

where $\Upsilon = [\upsilon_1, \ldots, \upsilon_K]^T$ is a realization of Y. In order to estimate the variance on the right-hand side, using (67) we can write

$$w_i^R = x_{m(i)}^R \theta^R - x_{m(i)}^I \theta^I + \eta_i^R \quad (72)$$

where θ^R and θ^I are, respectively, the real and imaginary parts of the summation in (67), and both are $O(\delta)$. Since the noise is independent of the other terms, we can write

$$\mathrm{var}\left[w_i^R|Y=\Upsilon\right] = \mathrm{var}\left[x_{m(i)}^R\theta^R - x_{m(i)}^I\theta^I|Y=\Upsilon\right] + \sigma^2. \quad (73)$$

Furthermore, for the first term on the right-hand side, we have

$$\mathrm{var}\left[x_{m(i)}^R\theta^R - x_{m(i)}^I\theta^I|Y=\Upsilon\right] \leq 2\mathrm{var}\left[x_{m(i)}^R\theta^R|Y=\Upsilon\right] + 2\mathrm{var}\left[x_{m(i)}^I\theta^I|Y=\Upsilon\right]. \quad (74)$$

From (9), by neglecting the FWM, we can write

$$x_{m(i)}^R = y_{m(i)}^R - \kappa_{m(i)}^R - n_{m(i)}^R. \quad (75)$$

Hence, the first term on the right-hand side of (74) can be written as

$$\begin{aligned}\mathrm{var}&\left[y_{m(i)}^R\theta^R - \kappa_{m(i)}^R\theta^R - n_{m(i)}^R\theta^R|Y=\Upsilon\right]\\ &\leq 3\,\mathrm{var}\left[y_{m(i)}^R\theta^R|Y=\Upsilon\right] + 3\,\mathrm{var}\left[\kappa_{m(i)}^R\theta^R|Y=\Upsilon\right]\\ &\quad + 3\,\mathrm{var}\left[n_{m(i)}^R\theta^R|Y=\Upsilon\right]. \end{aligned} \quad (76)$$

The second and third terms on the right-hand side of this inequality are, respectively, $O(P\delta^4)$ and $O(\sigma^2\delta^2)$, which are both negligible with respect to σ^2 in (73). Also, since $\{y_k\}$ are given, the first term is equal to

$$3\mathrm{var}\left[y_{m(i)}^R\theta^R|Y=\Upsilon\right] = 3(y_{m(i)}^R)^2\mathrm{var}\left[\theta^R|Y=\Upsilon\right]. \quad (77)$$

For $\mathrm{var}\left[\theta^R|Y=\Upsilon\right]$, we have

$$\mathrm{var}\left[\theta^R|Y=\Upsilon\right] = O\left(\sum_{k=1}^{K}\left|\lambda_k^{(m(i))}\right|^2\mathrm{var}\left[|x_k|^2|Y=\Upsilon\right]\right). \quad (78)$$

Recall from (35) that

$$\begin{aligned}|x_k|^2 &= |y_k - n_k|^2\\ &= |y_k|^2 - 2\mathrm{Re}\{y_k n_k^*\} + |n_k|^2.\end{aligned} \quad (79)$$

Assuming that the system is operating at high SNR, for the summand in (78) we have

$$\begin{aligned}&\left|\lambda_k^{(m(i))}\right|^2\mathrm{var}\left[|x_k|^2|Y=\Upsilon\right]\\ &\approx \left|\lambda_k^{(m(i))}\right|^2\mathrm{var}\left[2\mathrm{Re}\{y_k n_k^*\}|Y=\Upsilon\right]\\ &= O\left(\left|\lambda_k^{(m(i))}\right|^2|y_k|^2\sigma^2\right).\end{aligned} \quad (80)$$

Hence (78) and (80) yield

$$\mathrm{var}\left[\theta^R|Y=\Upsilon\right] = O\left((\gamma L_{\text{eff}})^2|y_k|^2\sigma^2\right) \quad (81)$$

where, as defined in Section II, γ and L_{eff}, respectively, denote the nonlinearity parameter and the effective length of the fiber, and we used the fact that the crosstalk coefficients λ_k^m are $O(\gamma L_{\text{eff}})$. Combining (81) with (77) yields

$$\mathrm{var}\left[y_{m(i)}^R\theta^R|Y=\Upsilon\right] = O(\delta^2\sigma^2) \quad (82)$$

where we used the fact that

$$O\left(\gamma L_{\text{eff}}|y_k|^2\right) = O\left(\gamma L_{\text{eff}}(y_{m(i)}^R)^2\right) = O\left(\delta\right). \quad (83)$$

By substituting (82) in (76) we obtain

$$\mathrm{var}\left[x_{m(i)}^R\theta^R|Y=\Upsilon\right] = O(\delta^2\sigma^2) + O(P\delta^4). \quad (84)$$

A similar equation can be derived for $x_{m(i)}^I\theta^I$. Hence, using (73) and (74), we can write

$$\mathrm{var}\left[w_i^R|Y=\Upsilon\right] = O(\delta^2\sigma^2) + O(P\delta^4) + \sigma^2 \approx \sigma^2 \quad (85)$$

for small δ. Hence, we can write the logarithm on the right-hand side of (71) as

$$\log\left(\frac{\mathrm{var}\left[w_i^R|Y=\Upsilon\right]}{\sigma^2}\right) \approx \log(1) = 0, \quad (86)$$

to the first order of approximation. Therefore, we have

$$h(w_i^R|Y) - h(\eta_i^R) \approx 0. \quad (87)$$

A similar result can be derived for w_i^I. Finally, using (70), each term inside the summation on the right-hand side of (69) is negligible, which implies that, to the first order of approximation

$$I(X;W|Y) = 0. \quad (88)$$

The result in (88) proves that, having collected all the information from the K matched filters, the outputs of other filters do not contain significant additional information for the decoder. While this analysis was performed for the sum-rate capacity, it is straightforward to do the same analysis for all the other terms in the capacity region of the channel, and the only difference with the above case will be that the computations will now be done conditioned on a subset of users' signals. As a result, we can show that

$$I(X_S;Y,W\,|X_{S'}) \approx I(X_S;Y|X_{S'}),\ \forall S\subset\{1,\dots,K\} \quad (89)$$

where X_S and $X_{S'}$ are defined in Section III. Consequently, the outputs of the K matched filters provide sufficient statistics for detection of the $\{x_i\}$.

To extend the analysis to the general case, where FWM is important and w_i contains arbitrary combinations of $\{x_i\}$, we write x_k in terms of y_k for each k, in a similar manner to (79). In this case, the variance of interference will also appear in (80); however, since the interference is weak, we will still be able to derive an equation similar to (81).

Now, assume that the channel has memory. As discussed in Section IV, memory adds a new dimension to the space of the received signals. For this space, we adopt as the first KN basis functions the normalized filters matched to the linear response of the channel to the pulses transmitted in subchannels $1,\dots,K$ and in time slots $1,\dots,N$, where N is the block length. Assuming that the quadratic dispersion is compensated, these basis functions are orthogonal, as each pulse is orthogonal to other pulses transmitted in different time slots and/or frequency bands. The linear term of the output signal corresponding to the kth user and nth time slot appears only in the output of one matched filter, and is orthogonal to all other basis functions. Hence, as in the memoryless case, the sampled output of the matched filters $\{y_k(n)\}$ for subchannels $k=1,\dots,K$ and time slots $n=1,\dots,N$ will contain one linear signal term (if the kth user is transmitting anything at that time), but the sampled outputs of the rest of the filters, $\{w_i\}$, will only contain noise, and possibly some third-order crosstalk terms. By an analysis similar to that in the memoryless case, it can be shown that, having observed the outputs of matched filters, the remaining samples w_i do not contain useful additional information for detecting the symbols. The only difference with the memoryless case is that there will be more crosstalk terms, due to the addition of the time dimension.

Remark 5: The discussion presented in this appendix does not imply that only the outputs of the matched filters contain useful information for the detector. The point made here is that the outputs of the other filters contain redundant information if we have already observed the output samples of the matched filters. In fact, it is not difficult to show that, although $I(X;W|Y)$ is $O(\delta^2)$ and negligible, $I(X;W)$ is not.

APPENDIX II
MODEL DETAILS

In this appendix, we derive the discrete-time model of (8). From [9], the low-pass equivalent linear transfer function of the channel in (2) is given by

$$H_1(\omega,z) = \exp\left(-\frac{\alpha z}{2}\right)\exp\left(-\frac{j\beta_2\omega^2 z}{2}\right). \quad (90)$$

With respect to the third-order term in the Volterra series expansion (2), we have

$$H_3(\omega_1,\omega_2,\omega,z) = H_1(\omega,z)H_3'(\omega_1,\omega_2,\omega,z), \quad (91)$$

where

$$H_3'(\omega_1,\omega_2,\omega,z) = -j\frac{\gamma}{4\pi^2}\frac{1-\exp\left(-\alpha z - j\beta_2(\omega_1-\omega)(\omega_1-\omega_2)z\right)}{\alpha + j\beta_2(\omega_1-\omega)(\omega_1-\omega_2)}. \quad (92)$$

and we have neglected the frequency dependence of the fiber loss factor α. In the expression for $H_1(\omega,z)$, the first factor is the attenuation from absorption and scattering in the fiber, and the second factor is the quadratic phase distortion caused by dispersion, which produces both phase and amplitude distortion in the time domain.

Now, if we substitute (6) in (2), the third-order term will contain a triple summation over channel indices. Each of the matched filters have frequency response proportional to $H_1^*(\omega,L)V^*(\omega)$. As the linear dispersion term (90) has a constant magnitude frequency response, matched filtering will perfectly equalize the linear term in (2). Therefore, assuming that the fiber attenuation has been compensated by the optical amplifier, the output of the matched filter corresponding to wavelength (user) i can be written as (8), where

$$\xi_{k,l,m}^{(i)} = \int d\omega V^*(\omega)\iint H_3'(\omega_1+k\Delta\omega,\omega_2+l\Delta\omega,\omega+i\Delta\omega,z)\, V\left(\omega-\omega_1+\omega_2+(i-k+l-m)\Delta\omega\right) V(\omega_1)V^*(\omega_2)d\omega_1 d\omega_2. \quad (93)$$

APPENDIX III
NEGLIGIBILITY OF $I(X_S;Y_{S'}|X_{S'},Y_S)$

For simplicity, here we show the negligibility of $I(X_S;Y_{S'}|X_{S'},Y_S)$ for the case where FWM is small compared to XPM. Extending the analysis to the general case with the FWM effect, although more tedious, follows the same steps as the XPM-limited case.

By definition, for $I(X_S;Y_{S'}|X_{S'},Y_S)$ we have

$$I(X_S;Y_{S'}|X_{S'},Y_S) = h(Y_{S'}|X_{S'},Y_S) - h(Y_{S'}|X,Y_S). \quad (94)$$

For the second term on the right-hand side we have

$$h(Y_{S'}|X,Y_S) = h(Z_{S'}) = |S'|\log(2\pi e\sigma^2), \quad (95)$$

which is a constant, and for the first term we can write

$$h(Y_{S'}|X_{S'},Y_S) \le \sum_{i\in S'}\left[h(y_i^R|X_{S'},Y_S) + h(y_i^I|X_{S'},Y_S)\right]. \quad (96)$$

We need to upper-bound each term inside the summation. From (9), neglecting the FWM, $h(y_i^R|X_{S'},Y_S)$ can be upper-bounded for any $i\in S'$ by

$$h\left(\kappa_i^R + n_i^R \mid x_i^I, Y_S\right) \le E_{\chi,\Upsilon_S}\left\{\frac{1}{2}\log\left(2\pi e\,\mathrm{var}\left[\kappa_i^R + n_i^R \,\middle|\, x_i^I=\chi, Y_S=\Upsilon_S\right]\right)\right\} \quad (97)$$

where

$$\kappa_i^R = -x_i^I \sum_{k\in S} \rho_k^{(i)} |x_k|^2. \tag{98}$$

Moreover, we can estimate the variance inside the logarithm as

$$\begin{aligned}\text{var}\left[\kappa_i^R + n_i^R \middle| x_i^I = \chi, Y_S = \Upsilon_S\right] \\ = O\left(|\chi|^2 \sum_{k\in S} (\rho_k^{(i)})^2 \text{var}\left[|x_k|^2 \middle| Y_S = \Upsilon_S\right]\right) + \sigma^2. \end{aligned}\tag{99}$$

By an analysis similar to (80), we can write

$$\begin{aligned}|\chi|^2(\rho_k^{(i)})^2 \text{var}\left[|x_k|^2 \middle| Y_S = \Upsilon_S\right] &= |\chi|^2 O\left((\rho_k^{(i)})^2 |y_k|^2 \sigma^2\right) \\ &= O(\sigma^2\delta^2), \ \ k \in S. \end{aligned}\tag{100}$$

Now, by combining (99) and (100), we obtain

$$\begin{aligned}\text{var}\left[\kappa_i^R + n_i^R \middle| x_i^I = \chi, Y_S = \Upsilon_S\right] &= \sigma^2 + O\left(\delta^2\sigma^2\right) \\ &\approx \sigma^2. \end{aligned}\tag{101}$$

Substituting this equation in (97) yields

$$h(y_i^R | X_{S'}, Y_S) \le \frac{1}{2}\log(2\pi e \sigma^2), \qquad i \in S'. \tag{102}$$

Using the same line of reasoning, we can derive a similar equation for $h(y_i^I | X_{S'}, Y_S)$. Hence, from (96) we obtain

$$h(Y_{S'} | X_{S'}, Y_S) \le |S'| \log(2\pi e \sigma^2). \tag{103}$$

Finally, by combining (94), (95), and (103), and using the nonnegativity of mutual information, we conclude that

$$I(X_S; Y_{S'} | X_{S'}, Y_S) \approx 0, \tag{104}$$

to the first order of approximation.

This result shows that if some of the users are not using the channel, even if they do not turn their transmitters off, the crosstalk produced on their channels cannot be used to improve the detection of the other users' signals. Therefore, without loss of generality, we can assume that the detector has only access to the output of the subchannels corresponding to active users.

APPENDIX IV
DERIVATION OF (24)

To compute the determinant in (20), we neglect all the terms that are at least two orders of δ (i.e., $O(\delta^2)$) smaller than the largest term, which is of the order of the square of the signal power. Specifically, in each element of the covariance matrix, we only keep the signal covariance, noise covariance, and the cross-terms of signal with crosstalk. Considering the properties of FWM, it can be observed that among the FWM terms in channel i, all but the terms of the form $x_k x_i^* x_m$ are independent of the signal term x_i, and hence do not have any cross-term with the signal in the covariance. Thus, we can neglect all of these terms and only keep those of the form $x_k x_i^* x_m$, which we call *dependent FWM*.

Neglecting the independent FWM terms, we can rewrite y_i as

$$y_i = x_i + j x_i \rho_i^{(i)} |x_i|^2 + j x_i \alpha_i + x_i^* \beta_i + n_i \tag{105}$$

where the terms on the right-hand side are, respectively, the signal, SPM, XPM, dependent FWM, and noise terms, and

$$\alpha_i = \sum_{k\neq i} \rho_k^{(i)} \, |x_k|^2 \tag{106}$$

and

$$\beta_i = \sum_{\substack{k-l+m=i \\ l=i}}\sum\sum \xi_{k,l,m}^{(i)} \, x_k x_m \tag{107}$$

are dimensionless $O(\delta)$ random variables independent of x_i. By separating the real and imaginary parts of (105), we obtain

$$\begin{cases} y_i^R = x_i^R - x_i^I \rho_i^{(i)} |x_i|^2 - x_i^I \alpha_i + x_i^R \beta_i^R + x_i^I \beta_i^I + n_i^R \\ y_i^I = x_i^I + x_i^R \rho_i^{(i)} |x_i|^2 + x_i^R \alpha_i + x_i^R \beta_i^I - x_i^I \beta_i^R + n_i^I. \end{cases} \tag{108}$$

We can calculate the determinant in (20) as

$$\det\left(\text{cov}\left[\underline{y}_i \middle| X_{S'}\right]\right) = C^R C^I - (C^{RI})^2 \tag{109}$$

where

$$C^R \triangleq \text{var}[y_i^R | X_{S'} = \chi], \tag{110}$$

$$C^I \triangleq \text{var}[y_i^I | X_{S'} = \chi], \tag{111}$$

and

$$C^{RI} \triangleq \text{cov}[y_i^R, y_i^I | X_{S'} = \chi]. \tag{112}$$

Using (108) and the notations defined in (21) and (22), (110) can be expanded as

$$\begin{aligned}C^R = (v_i^R)^2 - \text{cov}\left[x_i^R, x_i^I \rho_i^{(i)} |x_i|^2\right] - q_i v_i^R v_i^I a_i \\ + (v_i^R)^2 b_i^R + q_i v_i^R v_i^I b_i^I + \sigma^2 \end{aligned}\tag{113}$$

where b_i^R, b_i^I, and a_i are, respectively, the expected values of β_i^R, β_i^I, and α_i conditioned on $X_{S'} = \chi$, and we have neglected all $O(\delta^2)$ terms. For the second term on the right-hand side of (113), we have

$$\begin{aligned}\text{cov}\left[x_i^R, x_i^I \rho_i^{(i)} |x_i|^2\right] \\ = E\left[x_i^R x_i^I \rho_i^{(i)} |x_i|^2\right] - \mu_i^R v_i^R E\left[x_i^I \rho_i^{(i)} |x_i|^2\right]. \end{aligned}\tag{114}$$

Now, combining (114) and (113), and with some reordering of terms, we have

$$\begin{aligned}C^R = p_i^R - (\mu_i^R)^2 (v_i^R)^2 - E\left[x_i^R x_i^I \rho_i^{(i)} |x_i|^2\right] \\ + \mu_i^R O(P\delta) + q_i O(P\delta) + p_i^R b_i^R + \sigma^2. \end{aligned}\tag{115}$$

Similarly, we can write

$$C^I = p_i^I - (\mu_i^I)^2 (v_i^I)^2 + E\left[x_i^R x_i^I \rho_i^{(i)} |x_i|^2\right] + \mu_i^I O(P\delta) + q_i O(P\delta) - p_i^I b_i^R + \sigma^2 \quad (116)$$

and

$$C^{RI} = q_i v_i^R v_i^I + O(P\delta). \quad (117)$$

Substituting (115)–(117) in (109) results in (24).

Acknowledgment

The authors would like to thank Prof. Patrick Fitzsimmons for providing [13], and acknowledge helpful discussions with Nikola Alíc.

References

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

[2] L. G. L. Wegener, M. L. Povinelli, A. G. Green, P. P. Mitra, J. B. Stark, and P. B. Littlewood, "The effect of propagation nonlinearities on the information capacity of WDM optical fiber systems: Cross-phase modulation and four-wave mixing," *Phys. D*, vol. 189, pp. 81–99, Feb. 2004.

[3] J. Tang, "The multispan effects of Kerr nonlinearity and amplifier noises on Shannon channel capacity of a dispersion-free nonlinear optical fiber," *J. Lightw. Technol.*, vol. 19, no. 8, pp. 1110–1115, Aug. 2001.

[4] K. Ho and J. M. Kahn, "Channel capacity of WDM systems using constant-intensity modulation formats," in *Proc. IEEE Optical Fiber Communications (OFC'02)*, Anaheim, CA, Mar. 2002, pp. 731–733.

[5] B. Xu and M. Brandt-Pearce, "Multiuser detection for WDM fiber-optic communication systems," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 474.

[6] E. E. Narimanov and P. Mitra, "The channel capacity of a fiber optics communication system: Perturbation theory," *J. Lightw. Technol.*, vol. 20, no. 3, pp. 530–537, Mar. 2002.

[7] K. V. Peddanarappagari and M. Brandt-Pearce, "Volterra series transfer function of single-mode fibers," *J. Lightw. Technol.*, vol. 15, no. 12, pp. 2232–2241, Dec. 1997.

[8] G. P. Agrawal, *Nonlinear Fiber Optics*. San Diego, CA: Academic, 1989.

[9] B. Xu and M. Brandt-Pearce, "Comparison of FWM- and XPM-induced crosstalk using the Volterra series transfer function method," *J. Lightw. Technol.* , vol. 21, no. 1, pp. 40–53, Jan. 2003.

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

[11] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Inf. Theory*, vol. 35, no. 3, pp. 605–619, May 1989.

[12] M. H. M. Costa, "On the Gaussian interference channel," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 5, pp. 607–615, Sep. 1985.

[13] 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," *Ann. Appl. Probab.*, vol. 7, no. 4, pp. 1080–1105, 1997.