

On statistical effects on stimulated Raman cross-talk

J Villarroel¹ and A G Grandpierre²

¹ Univ. de Salamanca, Fac. de Ciencias, Plaza Merced s/n, Salamanca 37008, Spain

² Infineon Technologies Dresden GmbH & Co. OHG, Knigsbrcker Strasse 180, 01099 Dresden, Germany

E-mail: javier@gugu.usal.es and Alexandra.Grandpierre@infineon.com

Received 22 March 2005, in final form 15 June 2005

Published 13 July 2005

Online at stacks.iop.org/JPhysB/38/2601

Abstract

We provide a statistical model that accounts for the effects of dispersion on stimulated Raman scattering cross-talk in wavelength division multiplexed systems with random initial binary data. When either the input power is not too large or the number of channels involved is moderate or the transmission distance is small, channels are found to follow a Binomial probability distribution that is the same for channels symmetric with respect to the middle one. This distribution can be approximated by a Gaussian when the number of channels is not too small. Our model works for any fibre dispersion or number of channels.

1. Introduction

In recent years, stimulated Raman scattering (SRS) has attracted much attention, as this mechanism is responsible for power transfer from the lowest to the highest wavelengths in wavelength division multiplexed systems (WDM) [1]. This phenomenon is noticeable when a large number of wavelengths or channels are involved. In fact, the lowest wavelength signals may even become depleted to powers below the minimum receiver threshold levels. Recently, the exact solution for SRS power exchange in WDM systems was found in the continuous wave (CW) regime [2]. It was shown that for an initially fully loaded WDM system, SRS can lead to an exponential-like power distribution among channels, which increases as a function of distance [3, 4]. More recently, a model was proposed to explain the SRS power exchange in high-speed systems [4]. This work was based on the fact that if channels have different group velocities, for high enough chromatic dispersion WDM channels are walking off from each other, and as a result every channel experiences only the average power of all its neighbours. It is assumed that the initial signal in every channel is random binary data, i.e., it involves a random sequence of '1' and '0' bits with pulse duration τ_0 . Since the probability of having bits of '1' or '0' is usually taken the same, the problem can be reduced to a CW model for which the power of channel m seen by channel k is half of the actual peak power carried

by $\#m$. In addition, it was supposed that statistical quantities can be approximated by their statistical average. This assumption corresponds to a situation where the dispersion is high enough. (See [4, 5]. See also [6] for interesting preliminary results.)

When the former assumption is dropped, the question remains as to how to account for the statistical nature of bit power (per channel) in WDM systems at any transmission rate. In this case, one would need to consider the different possible interactions or scenarios between the randomly generated bits in every channel, whether these are '1' or '0'. Note that the number of possible combinations for the power exchange brought in by the randomness of the bit sequences can be quite extensive as the number of channels N grows (concretely $N!$ in the most favourable case). One would need to test these theories using the exact evolution equations, by considering an enormous number of different initial conditions. Hence, proving the accuracy of such statistical models involves a tremendous complexity and some approximations and further insight are necessary. Several studies have addressed this issue in different ways [7–10]. In some of these works the undepleted pump approximation has been employed. Nevertheless, so far all previous models have failed to explain how every channel may have a different probability distribution (see also [11–14] for some related results).

In the present work, we study the effects of chromatic dispersion on SRS cross-talk in WDM systems. We assume that the specific order of the bits '1' in any pattern does not play a significant role. What is critical however, is their ratio with respect to the total number of bits. Under this assumption, analytical results are obtained for the probability distribution functions of every channel whenever a certain control parameter is small; it corresponds to having either small input powers or small transmission distances; we show that in this case the powers of all channels follow exactly a binomial probability distribution with channel-dependent parameters. In most cases, they can be well approximated by Gaussian distributions, unlike what it is claimed in [7, 8]. Numerical results show an excellent agreement with these analytical expressions. When the control parameter is large, our numerical results show that the power distribution is no longer Gaussian and it is unclear what distribution do powers follow.

2. Problem formulation

2.1. Bit patterns and evolution equation

Let us assume a WDM system of N channels co-propagating in a single mode fibre. Channel 1 represents the lowest propagating wavelength and N the highest. We assume that each wavelength or channel travels at a different group velocity.

Let $Q_k(z)$ be the power in channel k at the propagation distance z , v_k the group velocity and α the fibre loss. The effective core area is given by A . We assume that every channel is separated by a frequency difference δf . Further $G = g'\delta f/2A$, and g' is the slope of the Raman gain of silica glass. The group velocities of channel k with respect to channel 1 can be written in terms of the frequencies as follows: $v_k^{-1} = \beta_0' + (k-1)\beta_0''\delta\Omega + \frac{(k-1)^2}{2}\beta_0'''\delta\Omega^2$, where $\Omega = (k-1)\delta\Omega$, and $\delta\Omega = 2\pi\delta f$. Then, following [4, 15], the power in channel k satisfies the boundary problem:

$$\frac{\partial Q_k}{\partial z} + \frac{1}{v_k} \frac{\partial Q_k}{\partial t} + \alpha(z)Q_k + GQ_k \sum_{m \neq k} (k-m)Q_m = 0 \quad (1)$$

with boundary condition $Q_k(z=0, t)$ random binary data.

The exact solution to this nonlinear system of first order partial differential equations is unknown. Following [4], we suppose that the power in channel k due to SRS crosstalk can

be obtained by using an effective time-independent approximation, the so-called ‘continuous wave’ approximation, and is given by

$$Q_k(z) = \frac{Q_k(0)J e^{-\alpha z + GJkz}}{Q_k(0) e^{GJkz} + d \sum_{m \neq k}^N Q_m(0)b_{k,m} e^{GJmz}}. \tag{2}$$

Here J is the total power in the system given by $J = [d \sum_{m=1, m \neq k}^N Q_m(0)b_{m,k} + Q_k(0)]$ and $b_{k,m} \leq 1$ represents the ratio of ‘1s’ to the total number of bits in the sequence as explained below (see equation (4)). The shape of the ‘1’ bits is also naturally taken into account through the duty cycle d , which represents the fraction of the area under a bit ‘1’ with respect to the total area of the bit slot. Thus, $d = 1$ if the bit occupies the entire bit period, or $1/2$ if only half of the slot is occupied. The RZ (return to zero) bit format is therefore preferred in our model. In what follows, $\exp(-\alpha z)$, which accounts for losses in equation (1), will be omitted with no loss of generality. The loss/gain term can be easily introduced back in the model. Further, we consider an initially equally loaded WDM channel: $Q_k(0) = C$ for all k . In this case, (2) reads

$$Q_k(z) = J \left(1 + d \sum_{m \neq k}^N b_{m,k} e^{GJ(m-k)z} \right)^{-1}. \tag{3}$$

Note that setting $d = b_{k,m} = 1$ we recover the (exact) solution corresponding to the assumption v_k constant for which walk-off effects are not present. To understand the appearance in the general case of the statistical quantities $b_{k,m}$, that measure how the channel k is affected by channel m , we reason as follows. Let us choose channel k as the reference channel. Because of walk-off effects, a ‘1’ carried by $\#k$ will ‘see’ in the other channels a certain pattern of bits passing by. Since the Raman effect is not affected by the direction of the bits, we consider that the speeds or number of bits (whatever their relative directions) are always positive. Channel k will then experience the cross-talk due to $\delta_{k,m}$ bits on channel m where $\delta_{k,m}$ is the number of bits in channel m that overtake $\#k$ over a distance z . A simple dynamical argument shows that $\delta_{k,m} = (v_k - v_m)z / (\tau_0 v_k v_m)$. Further, the above group dispersion relation implies that this number is proportional to the separation between the channels: $\delta_{k,m} = \Delta |k - m|$. The parameter Δ gives the cross-talk between neighbour channels, i.e. $\Delta = \delta_{m-1,m}$ is fixed and equal to $\Delta = z\beta_0'' \delta\Omega / \tau_0$. We assume that Δ is an integer.

Let us take at this point a closer look at the possible bit sequences. Suppose the system counts N channels and we fix the reference channel $k = 1$; thus after the propagation distance, $\#1$ saw Δ bits passing by on channel 2. The corresponding sequence may be: 00, 01, 10, or 11, that is 2^2 different sequences exist. Altogether, the number of pattern combinations that $\#1$ may experience from all the other channels is $2^{N(N-1)\Delta/2}$.

This ominous situation is somewhat simplified if we suppose that the effect of channel m on $\#k$, ($m > k$) depends *only* on the number of ‘1’ bits from m that overtake k , but not their actual position; that is, 10 and 01 pattern results in the same power SRS exchange. The validity of the assumption can be understood by noting that the optical power is only exchanged via SRS among bits of ‘1’. This approximation was checked by numerical simulations using equation (1) and by launching bit patterns with different arrangements. The assumption was found to work perfectly for low loss systems. The previous number of combination in the example given above is then reduced considerably to $(\Delta + 1)(2\Delta + 1) \cdots ((N - 1)\Delta + 1)$. For example, if $N = 8$ the number of possible combinations is $7, 2 \times 10^{16}$ in the general situation but ‘only’ $15!! = 2, 6 \times 10^7$ under the actual assumption. The simplification also affects the number of probability paths corresponding to a certain bit scenario, which is no longer the

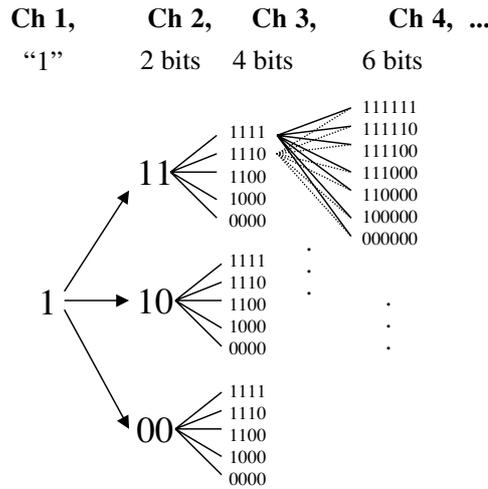


Figure 1. Interaction of channel 1 with neighbouring channels and $\Delta = 2$.

same. For $\Delta = 2$, the probabilities of occurrence for {00}, {01 or 10} and {11} are respectively $C_2^0/4$, $C_2^1/4$ and $C_2^2/4$ where $C_n^m \equiv \binom{n}{m}$. Figure 1 lists the different interactions a bit ‘1’ (on channel $k = 1$) may have with other channels.

Under the former assumption the quantities of interest are $\tilde{b}_{k,m}$ and $b_{k,m}$ which represent, respectively, the total number of ‘1’s’ that occur in a sequence of $\delta_{k,m}$ bits and the ratio of ‘1’ bits to the total number of bits. Because bits are launched independently of each other and have the same probability of occurring it follows that $\tilde{b}_{k,m}$ is a Binomial random variable $\tilde{b}_{k,m} = B(\delta_{k,m}, \frac{1}{2})$, and hence can take $\delta_{k,m} + 1$ different values:

$$\begin{aligned}
 b_{k,m} &= \frac{\tilde{b}_{k,m}}{\delta_{k,m}}, & \tilde{b}_{k,m} &= 0, 1, \dots, \delta_{k,m}, \\
 P(\tilde{b}_{k,m} = j) &= \binom{\delta_{k,m}}{j} \frac{1}{2^{\delta_{k,m}}}, & & \text{for } 0 \leq j \leq \delta_{k,m}.
 \end{aligned}
 \tag{4}$$

3. Small parameter regime. Probability distributions

3.1. Binomial distribution

In [4] it was implicitly assumed that $b_{k,m} = \frac{1}{2} \forall m, k, m \neq k$. This corresponds to having high speeds so as to deal with large sequences of bits. The present model aims to take into account statistical deviations from this assumption. In this case, the random nature of the quantities $b_{k,m}$ is inherited by the powers $Q_k(z)$. Note that the complicated form of expression (2) renders hopeless the task of determining an exact formula for the probability distribution of the random variables $Q_k, k = 1, \dots, N$. On the other hand numerical simulations become computationally unwieldy as the number of channels N increases since, as it has been pointed out, Q_k takes a tremendously large number of possible different values ($N!$ for the simplest case $\Delta = 1$).

The situation can be greatly simplified if one assumes, in addition, that $GzdJQ_k(0)$ is small. In this case we find that the power in k evolves according to the following formula:

$$Q_k(z) = C \left[1 - \frac{CGzd}{\Delta} \sum_{m=1, m \neq k}^N \tilde{b}_{k,m} \text{sign}(m - k) \right]. \tag{5}$$

The validation of this formula for all k is sketched in appendix A assuming that the dimensionless parameter $GC(N - 1)^2dz/2$ is ‘small’, a condition which is valid in most communications networks nowadays. It corresponds to having either low input power levels or a small number N of different channels or small transmission distances. The situation when N is large but $Q_k(0) \equiv C$ is small (low power regime) is of great interest; indeed, the initial power launched in every WDM channel is usually kept relatively low to minimize nonlinear effects. Note that in the appendix it is also shown that the approximation improves as k grows.

From the latter formula and the fact that $\tilde{b}_{k,m} = \delta_{k,m}/2$, it follows that the mean of the power distribution for $\#k$ is given by

$$\mu_k \equiv \bar{Q}_k = C \left[1 - \frac{GzdC}{4} N(N + 1 - 2k) \right]. \tag{6}$$

The variance is found to be given by

$$\sigma_k^2 \equiv \text{Var } Q_k(z) = \frac{(C^2Gzd)^2}{8\Delta} [(N - k)(N - k + 1) + k(k - 1)]. \tag{7}$$

It is convenient to introduce new parameters, $m_k^+ \equiv \frac{\Delta}{2}(N - k)(N - k + 1)$, $m_k^- \equiv \frac{\Delta}{2}k(k - 1)$, $m_k^+ + m_k^- \equiv M_k$ and $k_0 \equiv (N + 1)/2$. In terms of these the latter formula reads

$$\mu_k = C \left[1 + \frac{GzdCN}{2}(k - k_0) \right], \quad \sigma_k^2 = \frac{(C^2Gzd)^2}{4\Delta^2} M_k, \quad M_k = \Delta((k - k_0)^2 + k_0^2 - k_0). \tag{8}$$

The mean grows linearly with k , but is independent of the dispersion parameter Δ . It shows how the power is transferred from the lowest to the highest wavelength number channels. Note that the maximum averaged cross-talk is $\bar{Q}_N/\bar{Q}_1 = (1 + \beta)/(1 - \beta)$ where $\beta = \frac{GzdC}{4}N(N - 1)$.

The variance σ_k^2 is symmetric with respect to k_0 and attains a minimum for $k = k_0$, i.e. in the middle of the cluster. It first decreases with k , and then grows quadratically again. Besides, it is inversely proportional to the dispersion parameter Δ . Hence

$$\mu_1 \equiv C \left[1 - \frac{GzdC}{4}N(N - 1) \right] \leq \mu_k \leq C \left[1 + \frac{GzdC}{4}N(N - 1) \right] \equiv \mu_N \tag{9.1}$$

$$\sigma_{k_0}^2 = \frac{(C^2Gzd)^2}{16\Delta}(N^2 - 1) \leq \sigma_k^2 \leq \frac{(C^2Gzd)^2}{8\Delta}N(N - 1) = \sigma_1^2 = \sigma_N^2. \tag{9.2}$$

To have a complete statistical information the exact probability distribution that Q_k follows is required. As we have pointed out, the number of possible values computed from (2) that Q_k may take, grows very quickly with N . Remarkably, if $GC(N - 1)^2dz/2$ is small most of the points tend to repeat each other and the latter probability distribution can exactly be determined. Concretely, using statistical properties we find that $Q_k(z) \equiv Q$ can take only the values q_j where

$$q_j \equiv C \left[1 - \frac{C}{\Delta}Gzdzj \right] \quad \text{where} \quad -m_k^- \leq j \leq m_k^+ \tag{10}$$

with probabilities (see appendix B)

$$P\{Q = q_j\} = \binom{M_k}{m_k^+ - j} \frac{1}{2^{M_k}}. \tag{11}$$

This means that $Q_k(z) = C[1 - \frac{C}{\Delta}GzdX]$ is a linear function of a random variable X where $X + m_k^- \equiv m_k^- + \sum_{m=1, m \neq k}^N \tilde{b}_{k,m} \text{sign}(m - k)$ follows a binomial distribution $B(M_k, \frac{1}{2})$. Using this and the representation (8) for M_k we can assert the following: for every channel k , Q_k can take only $1 + M_k$ different values with the power distribution probability symmetric respect to its mean μ_k at which the peak power is attained; all channels have the same distribution (i.e., binomial up to a linear transformation) but the defining parameters are different; the distance between peaks of different channels is constant and equals $GzdC^2N/2$. Finally channels symmetrically positioned with respect to k_0 have the same distribution.

We next consider the study of the dependence of channel k on the cross-talk parameter. Assume for convenience that M_k is even; in this case equation (11) gives that the peak probability is attained at $j = (m_k^+ - m_k^-)/2$, i.e., at $q_j = \mu_k$ which is independent of Δ ; but the peak probability $\mathcal{P} \equiv P\{Q_k(z) = \mu_k\}$ depends on M_k and hence on Δ :

$$P\{Q_k(z) = \mu_k\} = \binom{M_k}{M_k/2} \frac{1}{2^{M_k}} \equiv \mathcal{P}(M_k). \tag{12}$$

This expression shows that

$$\frac{\mathcal{P}(M_k + 2)}{\mathcal{P}(M_k)} = \frac{M_k + 1}{M_k + 2} < 1 \tag{13}$$

and hence that \mathcal{P} decreases as M_k increases; hence so it does as Δ increases since M_k is proportional to Δ . For large Δ , Stirling's approximation shows that individual probabilities vanish:

$$P\{Q_k(z) = q_j\} \leq P\{Q_k(z) = \mu_k\} \approx 1/\sqrt{\pi M_k} \xrightarrow{\Delta \rightarrow \infty} 0. \tag{14}$$

However, it is misleading to believe that as $\Delta \rightarrow \infty$ the probability distribution vanishes. Indeed, an application of the well-known central limit theorem of probability theory shows that

$$Q_k(z) = C \left[1 - \frac{C}{\Delta}GzdX \right] \xrightarrow{\Delta \rightarrow \infty} \mu_k \quad \text{or} \quad P\left\{ \lim_{\Delta \rightarrow \infty} Q_k(z) = \mu_k \right\} = 1 \tag{15}$$

i.e., for large speeds $Q_k(z)$ is deterministic and is a linear function of k : $Q_k(z) = C[1 - \frac{GzdC}{4}N(N + 1 - 2k)]$ and we recover the result of [4] (equation (6) of [4] and discussion in following pages. See also [16]). In the next section we comment on numerical simulations that corroborate all these facts.

3.2. Gaussian approximation

Formula (11) gives exactly the distribution of Q_k assuming that $GC(N - 1)^2dz/2$ is small. If, in addition, N is large it is well known that the binomial distribution $B(M_k, \frac{1}{2})$ can be approximated by a normal distribution which implies $Q_k(z) = N(\mu_k, \sigma_k^2)$ (see [17]). Depending on the requirements accuracy, the normal distribution is applicable if $M_k \geq 25$, even though smaller values give also good results. Reminding the reader that $q_j \equiv C[1 - \frac{C}{\Delta}Gz dj]$ and that

$$\frac{\Delta}{4}(N^2 - 1) \leq M_k \leq \frac{\Delta}{2}N(N - 1) \tag{16}$$

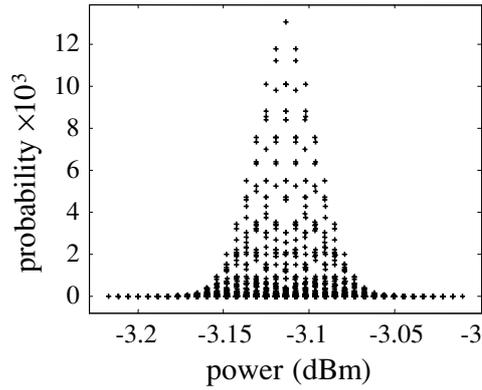


Figure 2. Numerical fit for the probability distribution of the power in channel 1 when $N = 4$, after considering all possible combinations of bit patterns in the other channels.

we see that, the number of channels required for the approximation to hold is, roughly, $N \geq 10/\sqrt{\Delta}$. In this case, for every k , $Q_k(z)$ is approximately Gaussian with pdf

$$P\{Q_k(z) = q_j\} \approx \sqrt{\frac{2}{\pi M_k}} \exp\left[-\frac{2(j - M_k/2)^2}{M_k}\right]. \quad (17)$$

Alternatively, for every real q ,

$$P\{Q_k(z) \leq q\} \approx \frac{1}{\sqrt{2\pi}\sigma_k} \int_{-\infty}^q \exp\left[-\frac{(x - \mu_k)^2}{2\sigma_k^2}\right] dx \quad \text{if } \frac{10}{\sqrt{\Delta}} \leq N. \quad (18)$$

This shows that, at least in this regime, the pdfs are not lognormal unlike what it has been claimed in [8].

4. Examples of probability distributions and numerical computations

We will now illustrate our results with numerical simulations that were carried out using equation (3). Let us consider a non-zero dispersion-shifted fibre operating around $1.55 \mu\text{m}$ with $A_{\text{eff}} = 50 \mu\text{m}^2$. Our examples assume WDM systems with channels equally spaced in the frequency domain with $\delta f = 100 \text{ GHz}$. The fibre dispersion is $D = 4 \text{ ps km}^{-1} \text{ nm}^{-1}$ at $1.55 \mu\text{m}$. At this wavelength the slope of the Raman gain is $g' = 6.7 \times 10^{-18} \text{ mW}^{-1} \text{ GHz}^{-1}$. Every channel is loaded initially with a random sequence of bits in a RZ modulation format with a duty cycle $d = 1/2$ and with a moderate input power of $C = Q_k(0) = 0.50 \text{ mW}$ or -3.0 dBm . The number of channels is taken relatively small to keep the numerical computations under control (recall that the number of combinations grows as $N!$ for $\Delta = 1$); concretely we first consider a four channel system ($N = 4$) operating at 40 Gbit s^{-1} with pulse duration $\tau = 25 \text{ ps}$.

Figure 2 shows the probability distribution of the first channel after a propagation distance of $z = 4.64 \text{ km}$. These results were obtained by directly using equation (3). Note again that the points are partitioned on several distinct vertical lines. Actually, every probability corresponds to a different power level. To numerically compute the pdf, these points must be added in each intervals. The maximum power is obtained at -3.12 dBm which occurs with a probability $\mathcal{P} = 0.132$.

Figure 3(a) shows the Gaussian distribution for these four channels obtained using equation (3). More precisely, the curves represent the envelope of the discrete probabilities.

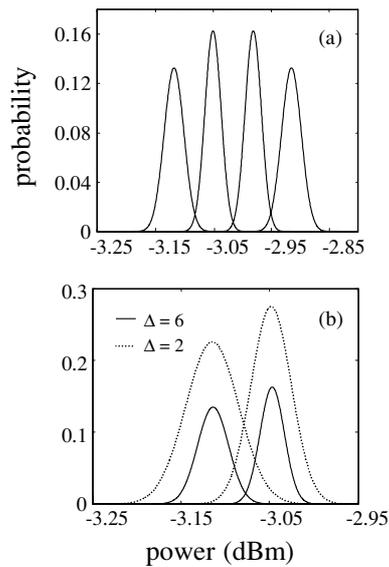


Figure 3. Numerical envelopes of the probability distribution for each of the 4 channels (a), and comparison between a low and a high speed system for the first two channels (b).

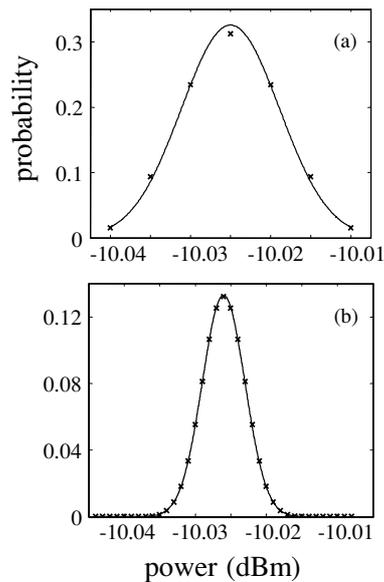


Figure 4. Probability distribution of channel 1 as obtained from numerical results (crosses) and Gaussian fit (solid) with $\Delta = 1$ (a) and $\Delta = 6$ (b).

The distributions are symmetric with respect to the middle channels. In addition to this, the channels that are spectrally farther apart suffer more from SRS and have therefore a bigger deviation from the mean. In figure 3(b), we compare the pdf of the first two channels when $\Delta = 6$ as well as for a slower system with $\Delta = 2$. The deviation is bigger when low speeds are used. Likewise, the peak probability decreases when the speed increases, in agreement with the prediction of equations (13) and (14).

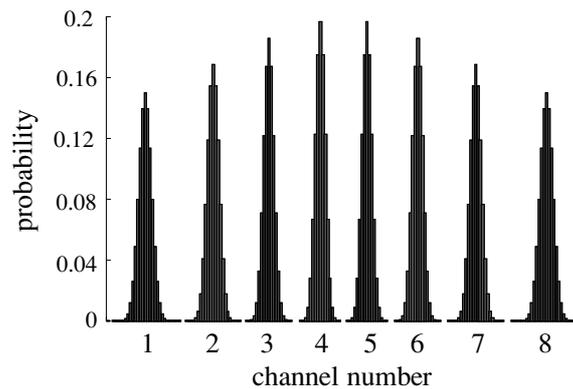


Figure 5. Probability distribution for an eight channel WDM system operating in the low power regime.

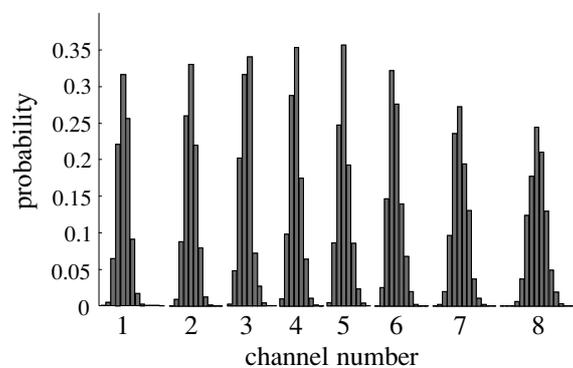


Figure 6. Probability distribution for an eight channel WDM system operating in the high power regime.

Figure 4 compares the results as obtained from equation (3) and those given by the Gaussian fit of equation (17). The exact probabilities are represented by the crosses while the Gaussian approximation is given by the solid curve. Note again that here we use an envelope over the actual points. The system involves four channels but only channel 1 is represented. The first figure exhibits the pdf of a system with $\Delta = 1$, and the second example assumes $\Delta = 6$. To compare with the theory we have considered a low input power of 0.1 mW for each channel. The approximation is, as we expected, better in the second case since the number of points is $M_1 = 6$ in the first situation and $M_1 = 36$ in the second. For $\Delta = 6$ the Gaussian approximation matches the numerical results quite well. Additional simulations also confirmed that even at low dispersion rates the (Gaussian) approximation improved greatly as the total channel number N increases. However, when the number of channels is small, the binomial formula (11) still fits perfectly the numerical results.

The last figures 5 and 6, exhibit an eight channel WDM network operating at 10 Gbits s^{-1} . It corresponds to a slow system with a $\Delta = 1$. The numerical probability distributions obtained from equation (3) are depicted for each channel first for the low power regime (figure 5) and secondly in the high powers range (figure 6). The bar plots resemble a Gaussian distribution in the first case, in accordance with equation (17). As the total input power increases, the highest channel stay relatively close to a Gaussian fit but channels at lower wavelengths get

deformed and are no longer Gaussian; it is unclear whether they approach a lognormal shape as claimed in [9].

5. Discussion and conclusions

In this paper, we investigated how walk-off affects the SRS crosstalk in WDM systems. We take into account group velocity dispersion effects, i.e. we assume that different channels have different group velocities. Effects arising from pulse dispersive effects and four wave mixing are ignored. Our work was carried out under the following assumptions: (i) the initial signal in every channel is independent random binary data, (ii) only the amount of '1' bits plays a significant role but not their specific order, (iii) channels are initially equally loaded, (iv) by using a time average, the SRS crosstalk under walk-off situations can be approximated via an effective time-independent (or 'continuous wave') situation and finally (v) either the initial power or the number of channels or the transmission distance is small. Exact analytical formulae were derived for the probability distribution functions of every channel; they are shown to have binomial distributions—which can be well approximated by a normal distribution when the number of channels is not too small. The corresponding mean and variance are channel dependent, narrower for the middle channels and with a broader width—as a result of having more crosstalk interaction—for the end channels. The pdf gets narrower as the speed is increased. Numerical calculations were performed. In all examples considered we find that the numerical results fit perfectly with the predictions from theory: the biggest difference found between theory predictions and numerical observations is around 2%.

The generalization of these results to the opposite case of having a large parameter, namely having *both* large initial powers and a large number of channels, is also a problem of overriding interest that we shall address in a future publication. As we have pointed out in this case the distributions are not Gaussian anymore and it is unclear what statistical distributions may show. Work in this direction is now in progress.

A second interesting problem is to consider a more general setting corresponding to having a system which is initially loaded with a sequence of '0' and '1' bits that have a different weight, i.e., '0' and '1' are assigned different probabilities; and where different channels are loaded with different initial powers: $Q_k(0) \neq Q_m(0), m \neq k$. It is unclear if, in the low parameter regime, binomial/Gaussian distributions will still obtain for this case.

Even more unclear at this stage is the question of as to whether the introduction of pulse dispersion effects will modify the picture described here. However, unless some effective approximation for this regime is devised, consideration of this problem seems to be out of the scope of the techniques explained here.

Acknowledgments

One of us (JV) wants to warmly acknowledge conversations with D N Christodoulides who brought to our attention this problem. Conversations with Mark Ablowitz at the University of Colorado at Boulder are also appreciated. This project was supported in part by DGICYT PB98/0262 and JCL SA43/00B, NSF (National Science Foundation) and by ARO-MURI.

Appendix A. Power distribution

As it has been pointed out, the power of channel k is given by formulae (2) and (3). We are assuming that every channel has, at the origin, the same power, i.e. $Q_k(0) = C, \forall k$ and hence

$J = C[\sum_{m=1, n \neq k}^N db_{k,m} + 1]$ is the total power at $z = 0$. Note that if $GJkz \ll 1$ one has

$$1 + \sum_{m \neq k}^N db_{k,m} e^{GJ(m-k)z} = \frac{J}{C} + \sum_{m \neq k}^N db_{k,m} (e^{GJ(m-k)z} - 1) \approx \frac{J}{C} + \sum_{m \neq k}^N db_{k,m} GJ(m-k)z.$$

It follows from (3) that

$$Q_k(z) = J \left[1 + \sum_{m \neq k}^N db_{k,m} e^{GJ(m-k)z} \right]^{-1} \approx C \left[1 - \frac{dGCz}{\Delta} \sum_{m \neq k}^N \tilde{b}_{k,m} \text{sign}(m-k) \right].$$

The above approximation requires that $GJ|m-k|z$ be small. Note

$$\bar{J} = C \left[\sum_{m=1, n \neq k}^N d\bar{b}_{k,m} + 1 \right] = (1 + d(N-1)/2)C \approx C(N-1)d/2$$

with the last approximation valid when N is not too small. Likewise

$$Gz(m-k)\bar{J} = \left(1 + \frac{d}{2}(N-1)\right)GCz(m-k) \leq GCz \left(1 + \frac{d}{2}(N-1)\right)(N-1) \approx GCz \frac{d}{2}(N-1)^2.$$

Since the typical deviation of J is small, we can approximate $J \approx \bar{J}$ and the expansion is validated for all k if $GC(N-1)^2 dz/2$ is small. Nevertheless, note that the approximation improves when $k = N$ (since then the exponent is negative).

Appendix B. Derivation of some probabilistic results

It is remarkable that expression (5) giving Q_k as a sum of random variables can be simplified further. Note

$$X \equiv \sum_{m=1, m \neq k}^N \tilde{b}_{k,m} \text{sign}(m-k) = \left(\sum_{m>k}^N - \sum_{m<k}^N \right) \tilde{b}_{k,m} \equiv X_+ - X_-$$

where $X_+ \equiv \sum_{m, m>k}^N \tilde{b}_{k,m}$ is a sum of independent binomial variables $\tilde{b}_{k,m} = B(\Delta|m-k|, \frac{1}{2})$. It is well known that in this situation X_+ is also binomially distributed $X_+ = B(m_k^+, \frac{1}{2})$ where

$$m_k^+ = \Delta \sum_{m, m>k}^N |m-k| = \frac{\Delta}{2}(N-k)(N-k+1).$$

Likewise $X_- \equiv \sum_{m, m<k}^N \tilde{b}_{k,m}$ has a binomial distribution $B(m_k^-, \frac{1}{2})$.

To proceed further, let us recall some standard probabilistic notation (see [17] for a good account of probability theory). Call support of X the set of values X may take with positive probability; let $P(A|B)$ be the conditional probability of the event A given that B happened: $P(A|B) = \frac{P(A \cap B)}{P(B)}$. It follows from the above considerations that X_{\pm} , and $X = X_+ - X_-$ have support (i.e., may take on the values) $\text{Support } X_{\pm} = \{0, 1, 2, \dots, m_k^{\pm}\}$, $\text{Support } X = \{-m_k^-, \dots, 0, 1, 2, \dots, m_k^+\}$. By the theorem on total probability one has that

for $j \in \text{Support } X$ is

$$\begin{aligned}
 P(X = j) &= \sum_{l=0}^{m_k^-} P(X = j|X_- = l)P(X_- = l) \\
 &= \sum_{l=0}^{m_k^-} P(X_+ = j+l|X_- = l)P(X_- = l) \\
 &= \sum_{l=0}^{m_k^-} P(X_+ = j+l)P(X_- = l) \\
 &= \frac{1}{2^{M_k}} \sum_{l=0}^{m_k^-} \binom{m_k^-}{l} \binom{m_k^+}{j+l} = \binom{M_k}{m_k^+ - j} \frac{1}{2^{M_k}}
 \end{aligned}$$

where we have used that X_+ , X_- are independent variables and hence $P(X_+ = j|X_- = l) = P(X_+ = j)$ and the well-known combinatorial identity:

$$\sum_{l=0}^n \binom{n}{l} \binom{m}{j+l} = \binom{m+n}{m-j}.$$

This is equation (11).

References

- [1] Kaminow I P and Koch T L 1997 *Optical Fiber Telecommunications IIIA* (San Diego: Academic)
- [2] Christodoulides D N and Jander R B 1996 Evolution of stimulated Raman crosstalk in wavelength division multiplexed systems *IEEE Photon. Technol. Lett.* **8** 1722–4
- [3] Bigo S, Gauchard S, Bertaina A and Hamaide J-P 1999 Experimental investigation of stimulated Raman scattering limitation on WDM transmission over various types of fiber infrastructures *IEEE Photon. Technol. Lett.* **11** 671–3
- [4] Grandpierre A G, Christodoulides D N, Schiesser W E, McIntosh C M and Toulouse J 2001 Stimulated Raman scattering crosstalk in massive WDM systems under the action of group velocity dispersion *Opt. Commun.* **194** 319–23
- [5] Grandpierre A G, Christodoulides D N and Toulouse J 1999 Theory of stimulated Raman scattering cancellation in wavelength-division-multiplexed systems via spectral inversion *IEEE Photon. Technol. Lett.* **11** 1271–3
- [6] Chraplyvy A R 1984 Optical power limits in multi-channel wavelength-division-multiplexed systems due to stimulated Raman scattering *Electron. Lett.* **20** 58–9
- [7] Forghieri F, Tkach R W and Chraplyvy A R 1995 Effect of modulation statistics on Raman crosstalk in WDM systems *IEEE Photon. Technol. Lett.* **7** 101–3
- [8] Ho K-Po 2000 Statistical properties of stimulated Raman crosstalk in WDM systems *J. Lightwave Technol.* **18** 915–21
- [9] Wang J, Sun X and Zhang M 1998 Effect of group velocity dispersion in stimulated Raman crosstalk in multichannel transmission systems *IEEE Photon. Technol. Lett.* **10** 540–2
- [10] Cotter D and Hill A M 1984 Stimulated Raman crosstalk in optical transmission: effects of group velocity dispersion *Electron. Lett.* **20** 185–7
- [11] Chraplyvy A R and Tkach R W 1993 What is the actual capacity of single-mode fibers in amplified lightwave systems? *IEEE Photon. Technol. Lett.* **5** 666–8
- [12] Christodoulides D N and Joseph R I 1989 Theory of stimulated Raman scattering in optical fibers in the pulse walkoff regime *J. Quantum Electron.* **25** 273–9
- [13] Stolen R H 1980 Nonlinearity in fiber transmission *Proc. IEEE* **68** 1232–6
- [14] Christodoulides D N and Joseph R I 1989 Slow bragg solitons in nonlinear periodic structures *Phys. Rev. Lett.* **62** 1746–9
- [15] Agrawal G P 1995 *Nonlinear Fiber Optics* 2nd edn (San Diego: Academic)
- [16] Zirngibl M 1998 Analytical model of Raman gain effects in massive wavelength division multiplexed transmission systems *Electron. Lett.* **34** 789–90
- [17] Papoulis A 1965 *Probability, Random Variables and Stochastic Processes* (New York: Mac-Graw Hill)