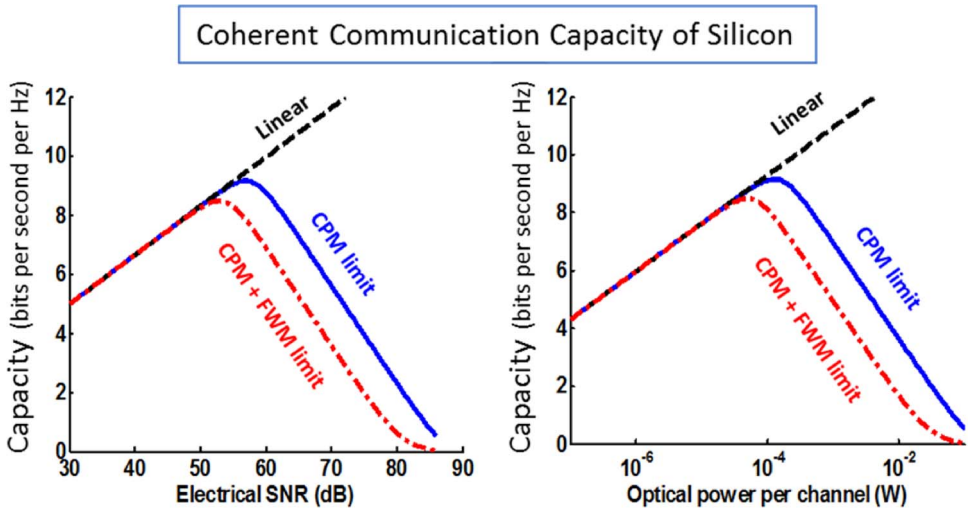
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Noise and Information Capacity in Silicon Nanophotonics

Volume 7, Number 3, June 2015

Dimitris Dimitropoulos
Bahram Jalali



DOI: 10.1109/JPHOT.2015.2427741
1943-0655

Noise and Information Capacity in Silicon Nanophotonics

Dimitris Dimitropoulos[1] **and Bahram Jalali**[1,2,3]

[1]Department of Electrical Engineering, University of California at Los Angeles, Los Angeles, CA 90095 USA
[2]Department of Bioengineering, University of California at Los Angeles, Los Angeles, CA 90095 USA
[3]California NanoSystems Institute, Los Angeles, CA 90095 USA

DOI: 10.1109/JPHOT.2015.2427741

Manuscript received March 5, 2015; revised April 23, 2015; accepted April 26, 2015. Date of publication May 6, 2015; date of current version June 11, 2015. Corresponding author: D. Dimitropoulos (e-mail: ddimitrgr@gmail.com).

Abstract: Modern computing and data storage systems increasingly rely on parallel architectures. The necessity for high-bandwidth data links has made optical communication a critical constituent of modern information systems and silicon the leading platform for creating the necessary optical components. While silicon is arguably the most extensively studied material in history, one of its most important attributes, i.e., an analysis of its capacity to carry optical information, has not been reported. The calculation of the information capacity of silicon is complicated by nonlinear losses, which are phenomena that emerge in optical nanowires as a result of the concentration of optical power in a small geometry. While nonlinear loss in silicon is well known, noise and fluctuations that arise from it have never been considered. Here, we report fluctuations that arise from two-photon absorption, plasma effect, cross-phase modulation, and four-wave mixing and investigate their role in limiting the information capacity of silicon. We show that these fluctuations become significant and limit the capacity well before nonlinear processes affect optical transmission. We present closed-form analytical expressions that quantify the capacity and provide an intuitive understanding of the underlying physics.

Index Terms: Channel capacity, silicon photonics, nanophotonics, nonlinear optics, optical crosstalk.

1. Introduction

Silicon is now recognized as an attractive platform for creating photonic components that provide data communication for massive arrays of servers and routers in data centers [1]–[4]. It can also solve the interconnect bottleneck at the board and eventually at chip scale [5], [6]. Silicon photonics is fuelled by economic, as well as performance, motives. The technology makes use of the same high-volume manufacturing infrastructure that has led to the explosive growth of consumer electronics and the Internet. Moreover, optical interconnects offer lower delays, higher bandwidth, and immunity to electromagnetic interference.

The amount of information that can be transmitted through silicon is a fundamental question that until recently had not been considered [7], [8]. Shannon's classical theory of information [9] shows that the capacity of a channel increases with signal power because, in the presence of ambiguity caused by noise, the number of distinguishable signal levels increases. But the concentration of power in limited cross sections activates new types of noises that place a

fundamental limit on the optical information capacity of silicon. This predicament is emphasized by the economics of silicon manufacturing, requiring efficient use of chip area [10] that drives down the optical wire or waveguide cross section to near the diffraction limit of light (0.25–0.5 μm^2).

The information capacity of a communication channel is the highest rate at which information can be transmitted reliably through the channel. For a simple channel where transmission is linear and the only source of noise is an additive Gaussian source with power N, the capacity, measured in bits is $C = (1/2)\log_2(1 + S/N)$, where S is the signal power [11]. Usually more than one channels are used at multiple wavelength, each operating at the channel capacity, the reason being that multiple wavelengths within a single waveguide improve silicon real estate usage, an important consideration in the economics of silicon manufacturing [10], [12]. As long as the channel is linear, the capacity can be improved by increasing the signal power and the number of wavelength channels.

Optical fibers exhibit nonlinearities at high optical powers and this places an upper limit on the capacity that can be predicted by the linear channel model [13]–[20]. In the case of optical fiber, it has been shown that cross phase modulation, arising from the inherent third-order nonlinearity present in the glass medium (Kerr effect) places a limit on the capacity of optical fibers [13]. In this pioneering work, a Gaussian statistics was assumed for the input distribution. However, in obtaining the capacity, it is necessary to find the optimum distribution that maximizes the bound. This is the case in our treatment of information capacity of silicon.

Silicon photonic devices work with infrared light (wavelength in the 1500 nm range) where the photon energy is insufficient to be absorbed. However at high intensities, two photons will pool their energies together to effect two-photon absorption (TPA) [21]–[23]. The rate, G, of this nonlinear absorption process depends on the light intensity as $G = \beta I^2/2hf$, where I is the intensity and $\beta \sim 0.5$ cm/GW is the two-photon absorption coefficient. The corresponding loss coefficient is $a_{\mathrm{TPA}} = \beta I$, measured in inverse centimeters (cm^{-1}). TPA also initiates a plasma effect: electrons and holes generated by TPA cause free-carrier absorption (FCA) and refraction (FCR), the magnitude of which is proportional to the density of carriers, $n_C = G\tau$, where τ is the carrier lifetime. This plasma effect in the form of the so-called FCA loss is characterized by a loss coefficient $a_{\mathrm{FCA}} = \sigma n_C$, where $\sigma \sim 10^{-17}$ cm^2 is the FCA cross section [21]–[23]. But TPA and FCA cause fluctuations—important phenomena that have not been considered prior to this work. These render the calculation of the information capacity of silicon significantly more complicated than that of optical fibers [7], [8].

The refractive index n depends on the intensity in the form of $n = n_o + n_2 I$ [24]. Here, n_o is the linear index and n_2 is a constant characterizing the strength of the nonlinearity. In coherent multi-channel or wavelength division multiplexing (WDM) communication, the phase in a given channel is influenced by intensities in other channels by a process called cross-phase modulation (CPM), and therefore coherent communication is affected. Four-wave mixing (FWM) can be curbed, in principle, either by dispersion in the waveguide, or by pre-chirping the signal. As we prove in Section 6, in such cases, CPM is the dominant nonlinearity-induced noise source in silicon and limits the information capacity in coherent WDM communication. For the case where FWM cannot be curbed, we show that the noise it generates limits the capacity in coherent WDM communication.

In long-haul communication, optical amplifiers are used to compensate for the loss of long fiber spans. The noise floor is then determined by the amplified spontaneous emission (ASE) of the amplifier [24]. Due to practical and economic reasons, optical amplifiers are not standard elements in silicon photonics. The principal noise sources in silicon photonics, aside from noise due to optical nonlinearity and nonlinear loss, are then thermal noise and shot noise.

As we will show, TPA and FCA nonlinearities have profound implications for information transmission in silicon waveguides, even at low intensity levels where the optical transmission is barely affected by the nonlinear effects. TPA is an instantaneous process because the nonlinear loss does not depend on past transmitted symbols. FCA, on the other hand, is not, since the carriers, generated due to TPA, do not recombine immediately but are present in the waveguide for the duration of the carrier lifetime. An interesting aspect of our findings is the relation

between the capacity and minority carrier lifetime unveiling an intriguing connection between semiconductor physics and information theory.

To begin the topic of noise in silicon photonics, we start by discussing the generation-recombination noise of the photo-generated carrier density in the waveguide and the fluctuations it induces in the optical transmission. Then, we examine the effect of non-linearities in a silicon optical communication channel and show that for incoherent signaling (i) TPA adds noise in WDM intensity modulation schemes, where each transmitted channel is a source of instantaneous (fluctuating) loss for the others and that (ii) FCA adds noise in WDM intensity modulation schemes, as well as in single-channel systems. For WDM, our results show that noise due to TPA is responsible for the value of the peak capacity that can be achieved (as well as the mean input optical power at which the maximum is achieved). In intensity-modulated links, FCA noise determines the rate of capacity decrease with power beyond the maximum. Proof of these are presented in the Appendix sections. We then consider the case of coherent communication where effects of CPM and FWM become important.

1.1. Optical Fluctuations Due to Generation-Recombination Noise

In silicon, TPA creates free carriers in the waveguide whose presence gives rise to FCA. Because the carriers randomly recombine and diffuse in and out of the optical path, their number fluctuates and therefore the FCA loss they cause also fluctuates. We compute the fluctuations of the loss coefficient due to FCA at a point x in the waveguide, in Appendix A.1. As we show there, the fluctuations in absorption coefficient translate into fluctuation in the intensity at the output of the waveguide, and these fluctuations will rise above the shot noise level when the photon number, $\langle n \rangle$, exceeds a threshold $\langle n \rangle > A_{\text{eff}}/\sigma_{\text{FCA}}$, where $\sigma_{\text{FCA}} \sim 10^{-17}$ cm^2 is the FCA cross-section and A_{eff} is the waveguide effective area. For 1 eV photons, 10 GHz signal bandwidth, and a typical waveguide effective area of 0.5 μm^2, this threshold power level is $\sim$800 mW. The spectrum of the loss coefficient fluctuations is determined by the carrier lifetime (see Appendix A.1) and has a cut-off frequency $(2\pi\tau)^{-1}$. As we will show, optical nonlinearities in silicon produce other types of noise that become dominant at much lower power levels.

2. Silicon Waveguide Channel Model

First let us consider the presence of TPA in a WDM scheme where N_c channels are being transmitted, each with power $P_j(z)$ along the dimension z of a waveguide of total length W_{wg}. The total power causes a modulation of the transmission factor of every channel through the waveguide. In a WDM system with a large number of separately coded channels, $N_c \gg 1$, each channel sees random multiplicative noise due to the other channels (see Appendix A.4 for details).

FCA in single-channel intensity-modulated transmission creates a similar problem. Any carrier density generated due to TPA in the waveguide has a characteristic decay lifetime, τ. Therefore, the carrier density present at any given time depends on the signal intensity during a past time interval equal to the carrier lifetime. When the pulse (symbol) duration T_s is much shorter than the lifetime, i.e., $\tau/T_s \gg 1$, as is typically the case, the number of past symbols that contribute to the FCA loss of the current pulse is $\approx \tau/T_s$ (see Appendix A.2 and A.3 for the derivation for statistics of the loss).

We first consider transmission that employs intensity modulation of the optical signal. The input random variable X $(X \geq 0)$ is the optical intensity at the input of the waveguide and the output $Y = LX + N$ is the measured photocurrent. Here the random variable N describes the additive thermal and shot-noise current at the photodetector, and L $(L \geq 0)$ is the channel transfer function that includes TPA and FCA nonlinear losses. Due to the channel nonlinearity, L is in general dependent on X. We start with the case of WDM with incoherent signaling. Let X_i denote the input optical power in the ith WDM channel, the loss L is dependent on the sum of power due to symbols in all the channels, i.e., on $\sum_i X_i$ (see Appendix Equation (A.4.1)). Therefore, when the number of channels is large, the loss, L is nearly independent of the instantaneous power in the current channel because the total optical power is mostly due to the other channels.

Specifically, for the jth channel we can write $\sum_i X_i \cong \sum_{i\neq j} X_i + \langle X_j \rangle$. Further, when each X_i in different channels are independent of each other (i.e., the current practice where channels are coded independently), but each drawn from the same statistical distribution (same coding for all channels), the moments of $\sum_{i\neq j} X_i$ can readily be calculated. We show in Appendix A.4, that the moments of L, have the functional form $\langle L^k \rangle = f_k(\langle X \rangle)$ for the kth moment.

In a similar manner, we derive in Appendix A.2 the functional form of $\log(L)$ for single channel transmission, and show that when the symbol period is much smaller than the carrier lifetime [see (A.2.3)] the loss fluctuation is largely the due to the previous symbols because the carrier population does not decay to its equilibrium value between symbols. Similarly, we show [see (A.3.5)] that for a wide range of power levels where L is not appreciably affected by non-linear losses, for Gamma distributed input symbols, the fluctuations in loss depend only on the mean optical power, i.e., $\langle \Delta L^2 \rangle = f(\langle X \rangle)$. Here, by appreciable we mean the change in $\log(L)$ caused by nonlinear losses is much smaller than unity. As we will see later, the Gamma distribution naturally arises in the solution to the present problem. We show in Appendix C a description of the channel parameters for: (a) a WDM system where TPA noise sets the limit and for (b) single-channel transmission where FCA noise is dominant. A full derivation of the models is presented in a structured manner in sub-sections of Appendix A. *As we will see later (see Fig. 2), through channel fluctuations, lifetime influences the information capacity of silicon. This intimate connection between semiconductor physics and information theory is a fascinating attribute of silicon nanophotonics.*

3. Lower Bound to the Channel Capacity for WDM Systems

To evaluate the maximum information that can be transmitted by a channel, we take the view that the input X and output Y are random variables with probability $p(X)$ and $p(Y)$. The mutual information $I(Y;X)$, is a measure of the dependence between the input and the output, and is given by $I(X;Y) = H(Y) - H(Y|X)$, where $H(Y) = -\langle \log p(Y) \rangle_Y$ and $H(Y|X) = -\langle \log p(Y|X) \rangle_{Y,X}$ with H denoting the information entropy. The maximum of $I(Y;X)$, over the possible input probability distributions, is the bits per channel use that can be reliably transmitted through a channel [11].

The following lower limit on the capacity is obtained from a lower bound on $H(Y)$ through application of the entropy power inequality [11] (details in Appendix B), and from an upper bound on $H(Y|X)$ that is obtained by noting that the random variable $Y|X$ has a lower entropy than a Gaussian variable of the same variance:

$$I_{LB} = \frac{1}{2}\ln\left(e^{2H(N)} + e^{2\langle \ln X\rangle + 2\langle \ln L\rangle + 2H(\ln L)} + e^{2H(X)+2\langle \ln L\rangle}\right) - \frac{1}{2}\ln(2\pi e) - \frac{1}{2}\left\langle \ln\left(\langle N^2\rangle + \langle \Delta L^2\rangle X^2\right)\right\rangle. \quad (1)$$

We now need to maximize this lower bound by varying the symbol distribution $p(X)$. As we discussed in Section 3, the loss L may not appreciably depend on X, but its moments do depend on moments of X. We therefore want to clarify first how the three terms $\langle \ln L\rangle$, $H(\ln L)$ and $\langle \Delta L^2\rangle$ should be treated in this optimization over the probability distribution:

The term $\langle \ln X\rangle - H(N)$ is of the same order of magnitude as $H(X) - H(N)$. The $H(\ln L)$ term can be neglected in the optimization, since it is large and negative, and therefore its exponential will be much smaller than unity (see equation (3)).

For a large range of power levels, $\langle \ln L\rangle$ is dominated by linear losses, with non-linear losses being a small perturbation. The loss fluctuations (that occur only due to nonlinear losses) are even smaller than the loss itself. Linear loss is not affected by the optimization as it is nearly constant, and the nonlinear loss is small. Therefore, $\langle \ln L\rangle$ can be kept constant in the optimization procedure. Within the range of power levels over which this quantity is constant, we find that the capacity exhibits a peak.

The last factor in (1) that we need to consider is $\langle \Delta L^2 \rangle$. This term depends on the moments of X, $M_k(X)$. The dependence can be generalized as $\langle \Delta L^2 \rangle = f(\langle M_1(X) \rangle, \langle M_1(X) \rangle, \ldots, \langle M_K(X) \rangle)$. Let us examine the role of this in the optimization.

First, consider ignoring the variations in $\langle \Delta L^2 \rangle$ in the optimization. Using the method of Lagrange multipliers, the distribution that maximizes the bound in (1), can be shown to have the form

$$p(X) = \frac{CX^{a-1}e^{-\lambda X}}{\left(\langle \Delta L^2 \rangle X^2 + \langle N^2 \rangle\right)^{b/2}}, C, a, b, \lambda > 0 \tag{2}$$

where the parameters a, b, λ are dependent on the mean of the input optical power $\langle X \rangle$ and C is the normalization factor. We note that for $b = 0$, this becomes a Gamma distribution which is the entropy maximizing distribution when the mean and the mean of the logarithm of the random variable are constrained. This can further be maximized by varying a, b until a maximum is found. In this case, optimizing the bound numerically shows that nearly maximal results are obtained with an exponential distribution ($a = 1$, $b = 0$). This is intuitively appealing because the exponential is the distribution that maximizes the entropy when the mean of the variable is constrained [9]. We also note that the sum of exponentially distributed channels (in a WDM system) is a Gamma distribution. In our case the mean is the average optical power. As shown in Appendix E, the effect of $\langle \Delta L^2 \rangle$ is that the resulting bound may be lower than what would be obtained using full numerical optimization in exchange for obtaining an intuitively appealing and closed form result that reveals the peaking of capacity and its dependence of material and device parameters.

We therefore obtain the following results for intensity-modulated WDM transmission:

$$C_{LB}(\text{bits/s} \cdot \text{Hz}) \cong \frac{1}{2}\log_2\left[\frac{1 + \frac{e^{1-2l}}{2\pi}\left(\frac{P}{P_N}\right)^2\left(1 + 2\pi e^{-1-2h}\left(\frac{P}{P_{\text{TPA}}}\right)^2\right)}{1 + 2e^{-2l}\left(\frac{P}{P_N}\right)^2\left(\frac{P}{P_{\text{TPA}}}\right)^2}\right]. \tag{3}$$

Here, $P = \langle X \rangle$ is the average optical power per channel, $P_N^2 = \langle N^2 \rangle$ is the additive noise power as measured in electrical domain, l is the total mean loss coefficient along the waveguide and $P_{\text{TPA}} = A_{\text{eff}}/N_C^{1/2}\beta_{\text{TPA}}L_{\text{eff}}$ is a characteristic power above which the transmission begins to be impaired by TPA and $h = 0.577$ is the Euler-Mascheroni constant. The additive noise power contains both thermal (300 K) and shot noise terms, $P_N = (\hbar\omega/q)^2(kT/R + 2q(q/\hbar\omega)P)B$. The peak of the capacity is reached at a power level $P_{\text{peak}} = (e^l P_{\text{TPA}} P_N)^{1/2}/2^{1/4}$. The optical power at which the capacity is reached is typically high enough for shot noise to dominate thermal noise. In that case, $P_{\text{peak}} \cong (e^{2l}P_{\text{shot}}P_{\text{TPA}}^2/2)^{1/3}$, where $P_{\text{shot}} = 2\hbar\omega B$ and the peak capacity is approximately $C_{\text{peak}} \cong (1/3)\log_2(e^{-2l}P_{\text{TPA}}/P_{\text{shot}}) + (1/2)\log_2(e/(2^{1/3}4\pi))$. We note that when the thermal noise floor is higher than the 300 K equivalent temperature assumed here, the power at which the link is shot noise limited increases.

4. Lower Bound to the Channel Capacity for Single Channel System

As stated before, for single channel communication, intersymbol interference due to FCA is the main bottleneck. By assuming that the power at the input of the waveguide is exponentially distributed, a lower bound to capacity is obtained. By using the prescribed log-normal channel model in (1), we obtain (using notation for average optical power $\langle X \rangle = P$):

$$C_{LB}(\text{bits/s} \cdot \text{Hz}) = \frac{1}{2}\log_2\left[\frac{1 + \frac{e}{2\pi}\frac{e^{-2l}P^2}{P_N^2}\left(1 + 2\pi e^{-1-2h}\frac{P^4}{P_{\text{FCA}}^4}\right)}{1 + \frac{2e^{-2l}P^2}{P_N^2}e^{(P/P_{\text{FCA}})^4}\left(e^{(P/P_{\text{FCA}})^4} - 1\right)}\right] \tag{4}$$

where P_{FCA} is defined in (A.3.6). This expression has units of bits/(s · Hz), whereas expression (1) (after multiplied by $\log_2 e$) has units bits per channel use. To relate these, we note that in non-coherent communication, an optical bandwidth B, results in an electrical bandwidth $B/2$, and therefore the rate of independent samples of the signal (channel uses) per second equals (by Nyquist's theorem) to $1/(2(B/2))$. The peak capacity is obtained approximately when $P^6 = e^{2l} P_{FCA}^4 P_N^2/4$. In the shot noise limited regime, $P_N^2 = PP_{shot}$, and the peak FCA limited capacity occurs at

$$P_{peak} = e^{2l/5}\left(P_{FCA}^4 P_{shot}/4\right)^{1/5} \tag{5}$$

with a peak value of

$$C_{peak} = \frac{1}{2}\log_2\left(\frac{e}{3\pi 4^{1/5}}\right) + \frac{2}{5}\log_2\left(\frac{e^{-2l}P_{FCA}}{P_{shot}}\right). \tag{6}$$

5. Lower Bound to Channel Capacity for Coherent Signaling

So far, we have discussed intensity modulation because this is the domain where silicon photonics operates today. However, due to the proliferation of coherent signaling in long distance fiber optic cables, it is foreseeable that future silicon photonic links may employ coherent communication so we are inclined to consider this case as well. We will examine a WDM system with coherent signaling, with and without four-wave mixing. We will start without four-wave mixing (FWM), and once we get an understanding of the channel, we will find that it is very easy to include it. As before cross-channel TPA causes random inter-channel interference, but now cross-phase modulation (CPM) is also an issue. Below we derive a capacity lower bound when CPM and TPA are simultaneously present. Because the cross-channel TPA coefficient is much smaller than the CPM coefficient, it turns out we can neglect TPA (see below for an exact statement). The limit is then due to CPM induced by the Kerr effect in silicon, as we show below.

We start with the nonlinear Schrodinger equation of propagation for the slow-varying amplitude $\varepsilon_i(z,t)$ at wavelength λ_i $(i = 0, 1, \ldots, N_c/2 - 1)$.

$$\frac{\partial \varepsilon_i}{\partial z} = -\frac{i\beta_2}{2}\frac{\partial^2 \varepsilon_i}{\partial t^2} + (2/A_{eff})(i\gamma_{Kerr} - \delta_{TPA})\left(\sum_j \left|\varepsilon_j(z)\right|^2\right)\varepsilon_i \tag{7}$$

where β_2 is the group velocity dispersion parameter for silicon (dispersion in a silicon waveguide can be tailored [27] and a typical value is ~ 2 ps^2/m), γ_{Kerr} is the Kerr coefficient, and δ_{TPA} the TPA coefficient. The signal has a bandwidth $B(Hz)$ and power $2P$. The coefficient of TPA (for amplitude as opposed to intensity) is $\delta_{TPA} = 0.25$ cm/GW, and the Kerr coefficient is $\gamma_{Kerr} = 1.5$ cm/GW [29]. In optical fiber, which lacks TPA, CPM was found to limit the capacity of coherent WDM communication [13]. While dispersion is significant in long spans of fiber links, it is much smaller in short lengths of integrated waveguides. Therefore the analysis of silicon here takes a different route compared to that in fiber.

Mathematically, we can evaluate the effect of dispersion as follows. Since $\partial^2 \varepsilon_i/\partial t^2 \sim (2\pi B)^2$, when the following condition holds:

$$\beta_2 \pi^2 B^2 \ll (\gamma_{Kerr}/A_{eff}) N_C P \tag{8}$$

dispersion can be ignored. In this case and for power levels at which the nonlinearities have not significantly impacted the channel transmission function (this is where the capacity

peaks), the complex optical field ε_i at the output of the waveguide at wavelength λ_i is approximately

$$\varepsilon_i(W) \cong \varepsilon_i(0)\exp\left((2/A_{\text{eff}})(i\gamma_{\text{Kerr}} - \delta_{\text{TPA}})\int_0^W dz \sum_j \left|\varepsilon_j(z)\right|^2\right) + n_i$$
$$\cong \varepsilon_i(0)\exp\left((2/A_{\text{eff}})(i\gamma_{\text{Kerr}} - \delta_{\text{TPA}})W\sum_j \left|\varepsilon_j(0)\right|^2\right) + n_i \quad (9)$$

where n_i is the additive noise in term field amplitude. Using the notation $= \varepsilon_i(W)$, $x = \varepsilon_i(0)$, $n = n_i$ and $y_{j\neq i} = \varepsilon_j(W)$, $x_{j\neq i} = \varepsilon_j(0)$

$$y \cong x\exp\left((2/A_{\text{eff}})(i\gamma_{\text{Kerr}} - \delta_{\text{TPA}})W\sum_j x_j x_j^*\right) + n. \quad (10)$$

Therefore, multiplicative noise is present due to interference from the other channels (the current channel can be ignored if the number of channels is large). The variable $\sum_{j\neq i}|\varepsilon_j(0)|^2$ can be decomposed to its mean value plus a variable that describes the fluctuations around the mean. Each channel is coded independently and has the same statistics. Because the fluctuation is the sum of independent and identically distributed variables, for a large number of channels, it can be approximated as a Gaussian random variable by the central limit theorem. With decomposition of the input into its in-phase and quadrature components, $x_j = \exp(j\theta_j)(r_j + im_j)$, where r_j and m_j are independent channels and θ_j is a constant phase, we can write the second moment (variance) of channel power as $M = \langle r_j^4\rangle - \langle r_j^2\rangle^2 = \langle m_j^4\rangle - \langle m_j^2\rangle^2$. Using this treatment in (10), we can approximate the input-output relationship as

$$y \cong xe^{(2/A_{\text{eff}})(i\gamma_{\text{Kerr}} - \delta_{\text{TPA}})N_C WP + (2/A_{\text{eff}})(i\gamma_{\text{Kerr}} - \delta_{\text{TPA}})\sqrt{N_C M}Z} + n \quad (11)$$

where Z is a zero mean Gaussian random variable with unity standard deviation, where we have used the relation $\langle x_j x_j^*\rangle = 2P$ for average power, P.

We now use a result from [13] which proves that a lower bound to the channel capacity can be obtained by assigning Gaussian distributions to input and output. In this case, the variance of channel power fluctuations is related to the mean power as $M = 2P^2$.

To compute the capacity we fist simplify the notation. The component of channel power transmission due to mean optical power P is $L_o = e^{-4N_c(P\cdot\delta_{\text{TPA}})W}$. By using the equivalent input amplitude

$$\tilde{x} = x.\exp\left((2/A_{\text{eff}})i\gamma_{\text{Kerr}}N_C WP + i\frac{P^2}{P_1 P_o}\right) \quad (12)$$

we decompose the channel into a real and an imaginary component. Here, P_o, P_1 are power levels that characterize CPM and TPA, respectively

$$P_o = \frac{A_{\text{eff}}}{2\gamma_{\text{Kerr}}W\sqrt{2N_C}} \quad (13)$$
$$P_1 = \frac{A_{\text{eff}}}{2\delta_{\text{TPA}}W\sqrt{2N_C}}. \quad (14)$$

To compute the capacity, we need the following correlators:

$$\langle y\tilde{x}^*\rangle = L_o^{1/2}\exp\left(\frac{1}{2}\left(\frac{P}{P_1}\right)^2 - \frac{1}{2}\left(\frac{P}{P_o}\right)^2\right)\langle \tilde{x}\tilde{x}^*\rangle \tag{15}$$

$$\langle yy^*\rangle = L_o\exp\left(2\left(\frac{P}{P_1}\right)^2\right)\langle \tilde{x}\tilde{x}^*\rangle + \langle nn^*\rangle. \tag{16}$$

while a part of the output is correlated to the input signal, the rest of the output power is noise. Using these to calculate the signal-to-noise ratio, we find for the capacity bound of a coherent WDM link

$$C_{LB} = (N_C B)log_2\left(1 + \frac{L_o P\exp\left(\left(\frac{P}{P_1}\right)^2 - \left(\frac{P}{P_o}\right)^2\right)}{P_n + L_o P\left[\exp\left(2\left(\frac{P}{P_1}\right)^2\right) - \exp\left(\left(\frac{P}{P_1}\right)^2 - \left(\frac{P}{P_o}\right)^2\right)\right]}\right) \tag{17}$$

where $P_n = \langle nn^*\rangle/2$.

Whereas, in the presence of only CPM, the peak capacity occurs at intensity $P_{\max} \cong (P_o^2 P_n/2)^{1/3} \ll P_o$ [13], in the case of both CPM and TPA, it occurs at

$$P_{\max} \cong \left(\frac{P_n/(2L_o)}{P_o^{-2} + P_1^{-2}}\right)^{1/3}. \tag{18}$$

when $\gamma_{\text{Kerr}} \gg \delta_{\text{TPA}}$ (the case in silicon), and since $L_o \cong 1$ in the regime where the peak occurs, the effect of TPA amounts to the following correction to the case where only CPM is present:

$$P_{\max} \cong \left(P_n P_o^2/2\right)^{1/3}\left(1 - \frac{\delta_{\text{TPA}}^2}{3\gamma_{\text{Kerr}}^2}\right). \tag{19}$$

In silicon, we have for the ratio of TPA to CPM coefficient $\delta_{\text{TPA}}/\gamma_{\text{Kerr}} = 1/6$, so the correction caused by TPA is approximately 1 part out of 100 and can be neglected. Using shot noise $P_n = \hbar\omega B$, for $N_c = 10$ channels, each of 2 GHz bandwidth, with waveguide mode area $A_{\text{eff}} \sim 0.5\ \mu\text{m}^2$ and length $W \approx 1$ cm, we find $P_o = 0.37$ W and $P_n = 256$ pW, which gives the power for peak capacity at $P_{\max} = 0.26$ mW/channel, or 2.6 mW total.

Returning to (8), we can verify that the dispersion-induced and nonlinearity-induced phase contributions in the propagation equation (7) have values $\beta_2\pi^2 B^2 \sim 8 \times 10^{-7}\ \text{cm}^{-1}$ and $(\gamma_{\text{Kerr}}/A_{\text{eff}})N_c P \sim 7.6 \times 10^{-4}\ \text{cm}^{-1}$, respectively. Therefore, neglecting dispersion is justified since at the point where nonlinear effects start to limit the capacity, the power is high enough for (8) to hold (in fact even if dispersion is an order magnitude higher, the assumption still holds). Also we note that, while the power is still low enough that the waveguide loss has barely been impacted by channel nonlinearity, noise caused by the nonlinearity is higher than shot noise and limits the capacity.

5.1. Four Wave Mixing (FWM)

As mentioned earlier dispersion in silicon waveguide link is small (compared to a fiber link). Consequently, the wave vector mismatch, $\Delta k_{jkli} = k_j + k_k - k_l - k_i \ll 1$ is small and FWM must be considered if pre-chirping methods are not used or are not practical. We will now show the

role of FWM as the source of fluctuations and its impact on the capacity. We include FWM in the propagation equation for the slowly varying amplitude:

$$\frac{\partial \varepsilon_i}{\partial z} = -\frac{i\beta_2}{2}\frac{\partial^2 \varepsilon_i}{\partial t^2} + \left(\frac{2}{A_{\text{eff}}}\right)(i\gamma_{\text{Kerr}} - \delta_{\text{TPA}})\left(\sum_j |\varepsilon_j(z,t)|^2\right)\varepsilon_i + i\left(\frac{\gamma_{\text{Kerr}}}{A_{\text{eff}}}\right)\cdot \sum_{j,k,l} \varepsilon_j \varepsilon_k \varepsilon_l^* e^{j(k_j + k_k - k_l - k_i)\cdot z}. \quad (20)$$

The dispersion-induced and nonlinearity-induced phase are also negligible (e.g., the overall phase in a single channel due to CPM is $\sim 6 \times 10^{-4}$ radian as shown above). This suggests that the FWM contribution (last term) in (20) can be considered constant along the waveguide. Also, because CPM has already been accounted for in the 2nd term, the sum in the FWM term must be restricted to $j \neq i$ and $k \neq i$. This plus the fact that data in different channels are uncorrelated, implies that the noise due to FWM can be considered as additive. This noise can be handled by modifying the additive noise power to include this in addition to shot noise

$$P_n = \langle nn^* \rangle / 2 = \hbar\omega B + \frac{1}{2}\left(\frac{\gamma_{\text{Kerr}} W}{A_{\text{eff}}}\right)^2 \sum_{j,k,l,n,m,q} \langle \varepsilon_j \varepsilon_k \varepsilon_l^* \varepsilon_n^* \varepsilon_m^* \varepsilon_q \rangle \quad (21)$$

$$i = j + k - l = n + m - q. \quad (22)$$

Equation (21) gives the noise in the i-th frequency band (that is in two channels, in-phase and quadrature). In what follows, we denote the number of frequency bands as N (total number of I & Q channels, $N_C = 2N$). Both the second and the fourth moments of the input amplitude appear in the summation in (21). As before, we will take the channel variables to be Gaussian variables in (21), since this is how the lower capacity bound for coherent case is achieved [13]. The second and 4th moments are therefore $\langle \varepsilon_j \varepsilon_j^* \rangle = 2P$ and $\langle \varepsilon_j \varepsilon_j^* \varepsilon_j \varepsilon_j^* \rangle = 8P^2$, respectively. For the ith channel we perform the summation in (21) in Appendix F. The result of the summation is

$$\begin{aligned} P_n &= \langle nn^* \rangle / 2 \\ &= \hbar\omega B + \left(\frac{\gamma_{\text{Kerr}} W}{A_{\text{eff}}}\right)^2 \times 8P^3 \times \left[(N-1)(N-1) - \frac{i(i+1)}{2} - \frac{(N-1-i)(N-i)}{2}\right]. \end{aligned} \quad (23)$$

The noise power therefore depends on the channel position and has the form

$$P_n = \langle nn^* \rangle / 2 = \hbar\omega B + P\left(\frac{P}{P_{\text{fwm}}^{i,N}}\right)^2 \quad (24)$$

where $P_{\text{fwm}}^{i,N}$ is a characteristic FWM power level for an I or Q channel in the ith band. For example

$$P_{\text{fwm}}^{i=0,N} = \frac{P_o}{\sqrt{(N_C-2)(N_C-4)/(8N_C)}}, \quad P_{\text{fwm}}^{i=K,N=2K+1} = \frac{P_o}{\sqrt{\frac{3}{2}(N_C-2)(N_C-10/3)/(8N_C)}}. \quad (25)$$

Including this new FWM noise, the new capacity lower bound becomes

$$C_{LB} = 2B\sum_{i=0}^{N-1} \log_2\left(1 + \frac{L_o P \exp\left(\left(\frac{P}{P_1}\right)^2 - \left(\frac{P}{P_o}\right)^2\right)}{P\left(\frac{P}{P_{\text{fwm}}^{i,N}}\right)^2 + P_n + L_o P\left[\exp\left(2\left(\frac{P}{P_1}\right)^2\right) - \exp\left(\left(\frac{P}{P_1}\right)^2 - \left(\frac{P}{P_o}\right)^2\right)\right]}\right). \quad (26)$$

The noise values in different WDM channels, $P_{\text{fwm}}^{i,N}$ have the same order of magnitude, and for large N_C approach the value $P_{\text{fwm}}^{i,N} \sim P_o/\sqrt{N_C}$. All terms in the summation in (26) peak around the same power level.

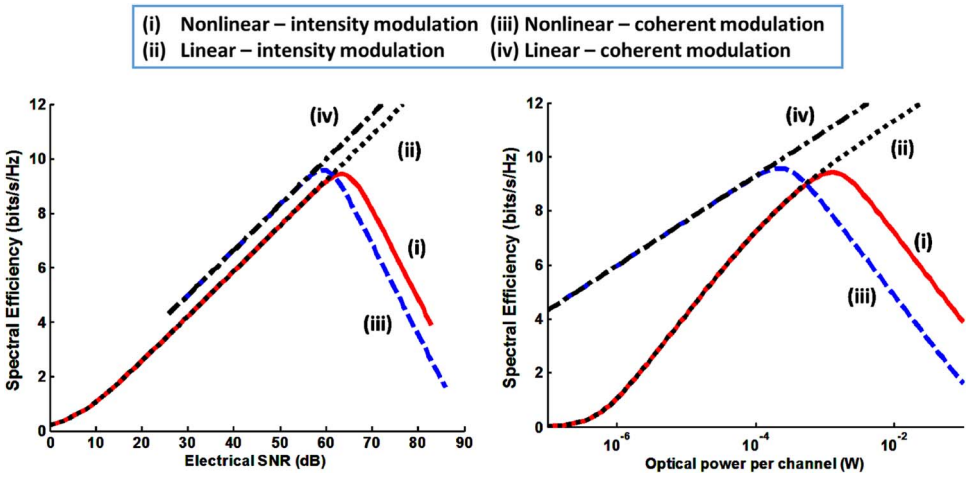


Fig. 1. (a) Channel capacity of silicon versus signal-to-noise ratio for a 10 channel (five wavelengths and two quadrature channels each) wavelength division multiplexed (WDM) communication. Capacities for both intensity modulation (incoherent signaling) and coherent modulation are shown. For the coherent case, the capacity is for one of the two quadrature channels; in other words, the total capacity for both in-phase (I) and quadrature (Q) channels will be twice the value shown. The dotted lines show the capacity for a linear channel with additive white Gaussian noise (AWGN). (b) The same capacity versus optical power. In the case of WDM communication. Ten WDM channels ($N_C = 10$), $W = 4$ Gbps per channel, waveguide length $L_{wg} = 1$ cm, effective area $A_{eff} = 0.5\ \mu m^2$ linear loss 0.5 dB/cm, and electrical receiver impedance $R = 50\ \Omega$. When the detector thermal noise floor is higher than the 300 K assumed here, the power at which peak capacity occurs becomes higher.

Two key observations that distinguish FWM from CPM+TPA:

a) The CPM+TPA lower bound can be exceeded using phase modulation. Here the signal constellation will have a constant power level that leaves the exponent in equation (9) constant. Therefore there is a way to overcome the limit. This is not the case with the FWM limit, since the FWM term in (20) is phase sensitive.

b) The FWM noise variance is approximately N_C times stronger and therefore it dominates over CPM, when the FWM process cannot be curbed. In that case, each of the terms in the sum in (26) peaks at a power level $\sim P_{max}/(N_C)^{1/3}$ (in the limit when N_C becomes large). The power level at which capacity peaks is now lower but it is still high enough that dispersion term in the propagation equation can be ignored as it is smaller than nonlinearity-induced phase. This easily holds for the numerical example we saw earlier in this section.

6. Results

The TPA induced noise penalty on the information capacity of a WDM link is illustrated in Fig. 1 for a set of typical parameters described in the caption. For both intensity modulation, the linear regime of increasing capacity, with optical power, is interrupted when the TPA-induced crosstalk sets in at high optical powers (or equivalently at high SNR). For coherent signaling the peaking of capacity is due to FWM (it is assumed that FWM cannot be curbed) in which case FWM induced fluctuations exceed those due to CPM). After that point, the capacity, as a function of SNR, decreases with the same rate that it was previously increasing. The reason for this drop, is not the nonlinear loss, but rather the induced crosstalk that, due to random nature of data in different channels, appears as random noise. Capacity is a measure of dynamic range, and in the nonlinear regime an increase in the input power translates in an even higher increase of the crosstalk level resulting in the observed peak in capacity for both intensity modulation [curve (i)] and coherent modulation [curve (iii)]. The maximum capacity attained is about 9 bits/s/Hz for intensity modulation. For comparison we also show: the capacity lower bounds for a linear silicon channel with only additive thermal noise (exponential input distribution). Here, the capacity

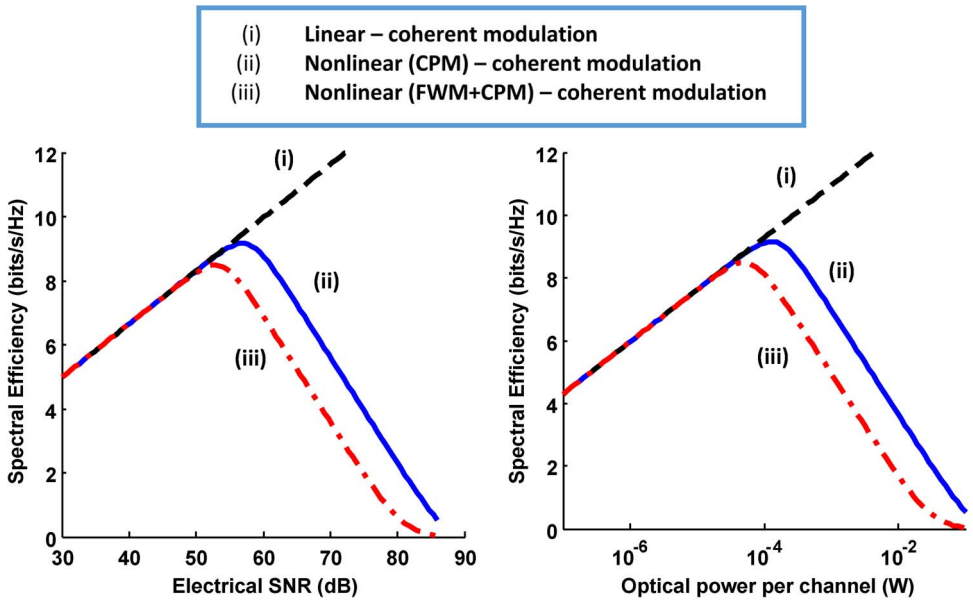


Fig. 2. Channel capacity for coherent WDM communication with 100 channels (50 wavelengths and two quadrature channels each). We compare (i) communication with AWGN (linear channel, (ii) communication limited only by CPM nonlinearity, and (iii) communication limited by both CPM and FWM nonlinearities. The capacity is for one of the two quadrature channels; in other words, the total capacity for both in-phase (I) and quadrature (Q) channels will be twice the value shown. Parameters used: $W_{wg} = 1$ cm, $A_{\text{eff}} = 0.5$ m^2, linear loss 0.5 dB/cm, and electrical receiver impedance $R = 50\ \Omega$. When the detector thermal noise floor is higher than the 300 K assumed here, the power at which peak capacity occurs becomes higher.

monotonically increases with increasing power as $C_{\text{awgn,non}-\text{coh}} = (1/2)\log_2(1 + (e/2\pi)(P^2/P_N^2))$ for intensity modulation. For coherent case, the bound values shown are for only one of the I or Q channels. The total capacity (combined I and Q channels) is twice the values shown, i.e., the capacity for coherent communication is higher than intensity modulation as expected.

We also show (see Fig. 2) results for WDM coherent communication in a $N_c = 100$ channel system (50 wavelengths and two quadrature channels on each). For such a high number of channels, the result of FWM is much more pronounced than the result of CPM. We show the capacity limit for both the cases, of FWM being present and absent. In the latter case, CPM is the main restricting factor, but in the former FWM is the dominant limiting mechanism.

The SNR and power dependence of these results merits some discussion. In the coherent case shot-noise limited detection is assumed (with unity photon/electron conversion efficiency), with an additive noise electrical SNR, $\text{SNR}_e = P/(\hbar\omega B)$, at the photodetector. This SNR can always be achieved in coherent communication by mixing the input signal with a strong local oscillator [24]. On the other hand, non-coherent communication (i.e., intensity modulation) may or may not be shot noise limited. In the photodetector the current fluctuations have both thermal and shot noise terms $\langle \Delta i^2 \rangle = (2q\langle i \rangle + kT/R)B$, where $\langle i \rangle = qP/\hbar\omega$ is the average photocurrent and R is the receiver impedance. For low enough optical signal power, thermal noise dominates the photo-current fluctuations and we get $\text{SNR}_e \sim P^2$, but as the optical power increases, shot noise dominates resulting in a linear dependence of SNR on optical power, i.e., $\text{SNR}_e = P/(2\hbar\omega B) \sim P$. These results are in agreement with both power [25] and SNR dependence [26] of the capacity of classical non-coherent optical communication in linear channels.

Similar peaking of the capacity bound is also observed for single channel communication as shown in Fig. 3 However, here the dominant source of fluctuations is the inter-symbol interference (ISI) caused by FCA. The same waveguide dimension and carrier lifetime of 1 ns as in WDM case are used, only the symbol period is now 100 ps, so that approximately a total of 9 past symbols contribute to the ISI. The linear regime of increasing capacity is interrupted, at a

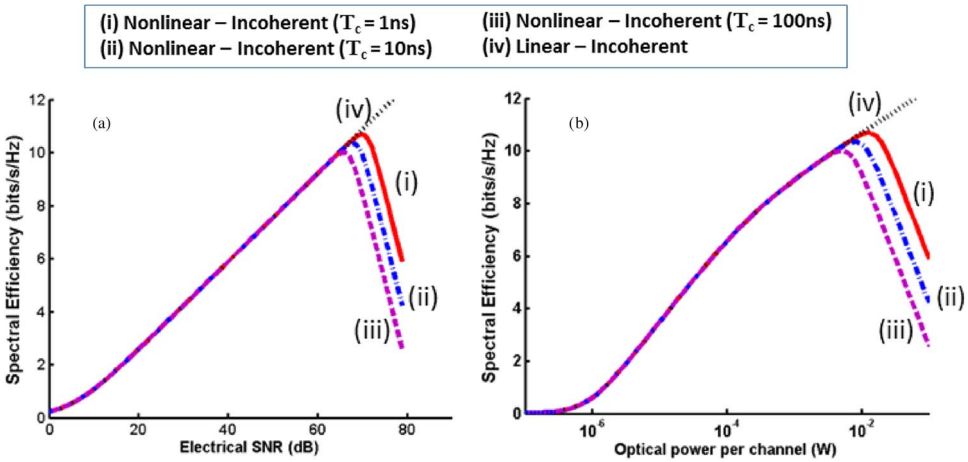


Fig. 3. (a) Channel capacity of silicon versus signal-to-noise ratio for single-channel communication. The dotted line shows the capacity for a linear channel with additive white Gaussian noise (AWGN). (b) The same capacity versus optical power. 10 Gbps data rate, 1–100 ns lifetime, $W_{wg} = 1$ cm, effective area $A_{\text{eff}} = 0.5\ \mu\text{m}^2$, linear loss 0.5 dB/cm, and electrical receiver impedance $R = 50\ \Omega$.

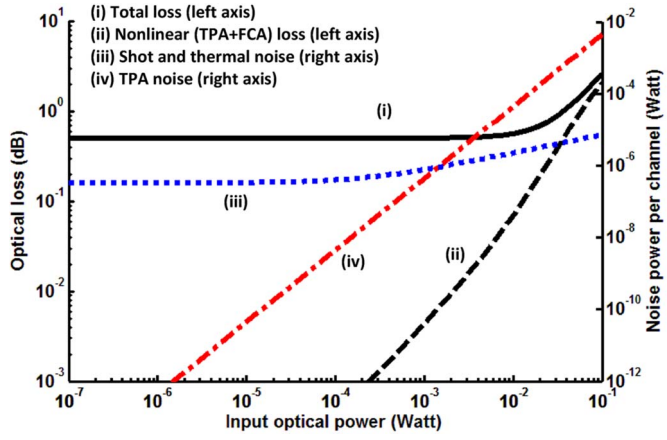


Fig. 4. Loss and fluctuation in a silicon waveguide channel for WDM communication. Two Photon Absorption (TPA) plus Free Carrier Absorption (FCA) loss and total waveguide loss (including passive insertion loss) are shown (left y-axis). Also, shot and thermal noise powers, as well as nonlinear TPA noise are plotted (right y-axis). 1 ns lifetime, $W_{wg} = 1$ cm, effective area $A_{\text{eff}} = 0.5\ \mu\text{m}^2$, linear loss 0.5 dB/cm, and electrical receiver impedance $R = 50\ \Omega$.

peak capacity of ~10.6 bits/s/Hz and now the capacity fall-off rate from the peak is twice the rate of increase in the linear regime. This steeper fall-of, as compared to the WDM case, is due to the steeper increase of ISI with input intensity. *A remarkable observation in both WDM and single-channel cases is that noise due to nonlinear loss (TPA, FCA and FWM) limits the capacity at power levels well below the level at which the nonlinear loss itself becomes significant (e.g., see* Fig. 4 *for the case of incoherent signaling*).

The results also show a dependence of capacity on carrier lifetime. Lifetime in silicon waveguide depends on waveguide geometry and on surface recombination velocity [30]

$$\frac{1}{\tau_{\text{eff}}} = \frac{1}{\tau_b} + \frac{S}{H} + \frac{w + 2(H-h)}{wH} S' + 2\frac{h}{H}\sqrt{\frac{D}{w^2}\left(\frac{1}{\tau_b} + \frac{S+S'}{h}\right)} \tag{27}$$

where S and S' are the surface recombination velocity at the interface between silicon and the buried oxide, and the silicon surface, respectively. H and w are the waveguide height and width,

and h is the rib height ($h = 0$ corresponds to channel waveguide). Also, τ_b is the bulk lifetime and D is the minority carrier diffusion coefficient. The analytical model of this dependence can be used to obtain the capacity bound for different waveguide geometries.

In the case of non-coherent (intensity modulation) WDM, we assumed the current practice, in which channels are coded separately, and to the degree that each channel is subject to a random delay relative to the others, TPA and TPA induced fluctuations are unavoidable. However, if the channels can be synchronized and kept synchronized by having low waveguide dispersion to guarantee insignificant walk-off, joint channel coding is possible and it allows overcoming the peaking of the capacity observed above. This results in monotonically increasing channel capacity but comes at the expense of an overhead in bit rate (by use of a constant weight code).

7. Conclusion and Future Work

We have presented an analysis of noise and information capacity of silicon waveguides and showed that fluctuations arising from nonlinear processes place a limit on the capacity. We first discussed the generation-recombination noise due to carriers generated during the TPA (two photon absorption) process. Through the FCA (free carrier absorption) process, these carriers produce optical transmission fluctuations. This served as an introduction to discuss noise in silicon photonics. We then derived the FCA induced noise and showed that it places a limit on single-channel incoherent communication. We also showed that in incoherent WDM transmission, noise due to TPA can limit the capacity. An examination of coherent WDM schemes, reveals that CPM and FWM pose more severe capacity restrictions than TPA. When FWM cannot be effectively curbed, it is the dominant limiting mechanism; otherwise, CPM places the limit. Future work can also quantify the effect of free carrier refraction (FCR).

This work provides the lower bounds on capacity and it would be interesting in the future to see whether upper limits can be obtained. Naturally, the observed limits are much higher than what is achieved today with binary intensity modulation. As silicon photonics advances into multilevel coherent signaling, it will move closer to the capacity limit.

One assumption in single-channel incoherent communication was the current practice where each symbol was coded separately from the previous symbol. One could examine more complicated schemes, or schemes that attempt to alleviate the non-linear inter-symbol interference (ISI). Equalization schemes for dealing with non-linear ISI can be found in the literature [28].

In WDM signaling, again we assumed the current practice of using separate, independent channels. However, if the channels can be synchronized and symbols can be encoded simultaneously, the total optical power can be made constant. Such joint coding allow one to extend the incoherent WDM communication capacity limit discussed here. The waveguide dispersion can be made low enough to guarantee there won't be significant walk-off and the symbols remain synchronous during propagation. However, if each channel is subject to a random delay relative to other channels then synchronous interchannel coding will not be a viable approach. When it comes to coherent communication, when FWM is present, the form of inter-channel interference is more complex and the capacity limit is more difficult to overcome with joint channel coding.

The tradeoff imposed by coding will be an overhead in terms of bit rate. If for example N WDM channels are being transmitted and PAM constellations with $M = 2^K$ symbols are used for each channel, one could restrict transmission to combinations of symbols in different channels that lead to a constant optical power present in the waveguide. If each of the M symbols at a given instant is always only transmitted by exactly N/M channels, then for a given code, a smaller number of codewords will have to be used. This penalty in the bits-per-use, is the tradeoff for sustaining and increasing the channel capacity with increasing power. Future work could also be in coding schemes that attempt to maintain a constant total transmitted power in the manner described above.

Recently, it has been shown that with in-line regeneration, capacity higher than Shannon limit can be achieved [31]. Also, it's been previously mentioned that for “static” point-to-point

channels, defined as channels that remain the same regardless of which codebook is used (a standard assumption in information theory), the channel capacity is a non-decreasing function of power [32]. This is not the case in our situation as statistical moments of input distribution do affect the channel transmission.

APPENDIX A
Noise Sources

A.1. Carrier Generation-Recombination Noise and Optical Intensity Fluctuations

In a silicon waveguide, carriers are randomly generated in a volume $V = A_{\text{eff}}\Delta x$ at the waveguide core (in an interval Δt) with probability $VG\Delta t$, where $G = \beta(P/A_{\text{eff}})^2/(2\hbar\omega)$ is the average rate (units $\text{cm}^{-3}\text{s}^{-1}$) of generation due to TPA. In this model, the carrier number $n = A_{\text{eff}}\int_0^W N(x)dx$ (where $N(x)$ denotes the carrier density) in a segment of length W of the waveguide, is Poisson distributed and therefore $\langle \Delta n^2\rangle = \langle n\rangle$. This means that the carrier density fluctuations must have the following correlation function:

$$\langle \Delta N(x)\Delta N(x')\rangle = \frac{\langle N(x)\rangle}{A_{\text{eff}}}\delta(x-x'). \tag{A.1.1}$$

The fluctuations of the optical loss coefficient is given by $a_{\text{FCA}}(x) = \sigma_{\text{FCA}}N(x)$. The propagation equation for the optical intensity $dI/dx = -a(x)I(x)$ can be used to obtain the fluctuations in the optical intensity by decomposing $I(x) = \langle I(x)\rangle + \Delta I(x)$, $a(x) = \langle a(x)\rangle + \Delta a_{\text{FCA}}(x)$

$$\frac{d}{dx}\langle I(x)\rangle = -\langle a(x)\rangle\langle I(x)\rangle \tag{A.1.2}$$

$$\frac{d}{dx}\Delta I(x) \cong -\langle a(x)\rangle\Delta\langle I(x)\rangle - \langle I(x)\rangle\Delta a_{\text{FCA}}(x) \tag{A.1.3}$$

where in (A.1.3), second-order terms have been dropped. At any point x in the waveguide, the intensity fluctuation is

$$\Delta I(x) = \int_0^x dx' I(x')\Delta a_{\text{FCA}}(x')\exp\left(-\int_{x'}^x \langle a(x'')\rangle dx''\right). \tag{A.1.4}$$

The correlation function for optical intensity can be obtained using the correlation function for the carrier density fluctuations

$$\left\langle \Delta I^2(x)\right\rangle = \frac{\sigma_{\text{FCA}}}{A_{\text{eff}}}\int_0^x dx'\langle a(x')\rangle\langle I(x')\rangle^2\exp\left(-2\int_{x'}^x \langle a(x'')\rangle dx''\right). \tag{A.1.5}$$

When the waveguide length is much smaller than the penetration length $1/\langle a(0)\rangle$, we can take $\langle I(x)\rangle \cong \langle I(0)\rangle$ and $\exp(-2\int_{x'}^x\langle a(x'')\rangle dx'') \cong 1$

$$\left\langle \Delta I^2(x)\right\rangle \cong \frac{\sigma_{\text{FCA}}}{A_{\text{eff}}}\langle I(x)\rangle^2. \tag{A.1.6}$$

This can be compared with intensity fluctuations due to shot noise (calculated for a data symbol period T)

$$\left\langle \Delta I^2_{\text{shot}}\right\rangle = \langle I\rangle\hbar\omega/(A_{\text{eff}}T). \tag{A.1.7}$$

This noise source overwhelms shot noise when $\langle n_{\text{phot}}\rangle > A_{\text{eff}}/\sigma_{\text{FCA}}$, where $\langle n_{\text{phot}}\rangle = \langle I\rangle(A_{\text{eff}}T/\hbar\omega)$ is the average photon number.

A.2 Free Carrier Absorption Loss Coefficient Fluctuations

The generation of carriers in the waveguide can give rise to inter-symbol interference. The free carriers that are generated by the passage of an optical pulse (due to TPA) have a finite lifetime, when the pulse width T is much smaller than the carrier lifetime $\tau = \gamma^{-1}$, the generated carrier density affects subsequent pulses. Given the fact that sequence of data symbols is random, the loss that a symbol sees is due to the passage of the previous pulses and will be fluctuating. Let's examine this fluctuation in the region where $\upsilon \cdot T > W$, where υ is the group velocity of the pulses, W denotes the waveguide length and T is the symbol length in time. In this case the pulse width in space is larger than the waveguide length, and therefore, the fluctuations will be correlated along the waveguide length.

The TPA process generates carries with a generation rate (units $\mathrm{cm^{-3}s^{-1}}$) $\beta(P/A_{\mathrm{eff}})^2/(2\hbar\omega)$, where β is the TPA coefficient and $\hbar\omega$ is the photon energy. Since the carriers that are generated decay exponentially, the optical loss due to FCA will have a time dependence

$$a_{\mathrm{FCA}}(x,t) = \frac{1}{P'_{\mathrm{TPA}}P'_{\mathrm{FCA}}W}\int_{-\infty}^{t}\gamma e^{-\gamma(t-t')}P^2(x,t')dt' \tag{A.2.1}$$

with the notation $P'_{\mathrm{TPA}} = A_{\mathrm{eff}}/(\beta W)$ and $P'_{\mathrm{FCA}} = (A_{\mathrm{eff}}/\sigma_{\mathrm{FCA}})(2\hbar\omega/\tau)$. These parameters can be interpreted as critical powers for TPA and FCA, respectively for single channel communication. (lowercase symbol "x" is used here in the appendices for the spatial variable whereas lowercase "z" was used in the main section of the paper).

The equation of propagation is then:

$$\frac{1}{2}\left(\frac{\partial P}{\partial(x/W)} + \frac{1}{(\upsilon/W)}\frac{\partial P}{\partial t}\right) = -(a\cdot W)P - P^2/P'_{\mathrm{TPA}} - (a_{\mathrm{FCA}}W)P. \tag{A.2.2}$$

We can simplify this using coordinates $u = (x+\upsilon t)/2$ and $w = t - x/\upsilon$, where w is time in the coordinate system of the pulse, and u labels the position of the pulse in the waveguide as it propagates. If at time w is the peak of the current pulse, then $w - T$ is the peak of the previous. With the new coordinates the transmission loss due to FCA that a pulse see as it propagates along the waveguide is $\exp(-\int_{w/2}^{W+w/2} a_{\mathrm{FCA}}(u,w)P^2(u,w)du)$. Therefore, to compute the fluctuations due to FCA the moments $\langle a_{\mathrm{FCA}}(u,w)a_{\mathrm{FCA}}(u',w)\rangle$ and $\langle a_{\mathrm{FCA}}(u,w)\rangle$ need to be computed. Since we are interested in high bitrates we take $\gamma\cdot T \ll 1$, meaning that the symbol duration is much less than the carrier lifetime. We discretize the integral by assuming that within a given pulse the exponential changes negligibly, and we use the discrete notation $P_n(u) = P(u, w = w_0 + nT)$, where the peak of the current pulse is located at $w_0 = 0$, so that $n > 0$, labels the previous pulses

$$a(u,w_0) \cong \frac{\gamma T}{P'_{\mathrm{TPA}}P'_{\mathrm{FCA}}W}\sum_{n=0}^{\infty} e^{-n(\gamma T)}P_n^2(u). \tag{A.2.3}$$

It is straight-forward to compute moments now as

$$\langle a(u,w_0)\rangle \cong \frac{\langle P^2(u)\rangle}{P'_{\mathrm{TPA}}P'_{\mathrm{FCA}}W} \tag{A.2.4}$$

$$\langle a(u,w_0)a(u',w_0)\rangle \cong \frac{1}{\left(P'_{\mathrm{TPA}}P'_{\mathrm{FCA}}L\right)^2}\left[\sum_{n=0}^{\infty} e^{-n(2\gamma T)}\langle P^2(u)P^2(u')\rangle + \sum_{n\neq m} e^{-(n+m)(\gamma T)}\langle P^2(u)\rangle\langle P^2(u')\rangle\right] \tag{A.2.5}$$

which can be used to obtain for the fluctuations

$$\langle \Delta a(u,w_0)\Delta a(u',w_0)\rangle \cong \left(\frac{\gamma T}{2}\right)\frac{\langle P^2(u)P^2(u')\rangle - \langle P^2(u)\rangle\langle P^2(u')\rangle}{\left(P'_{\mathrm{TPA}}P'_{\mathrm{FCA}}W\right)^2}. \tag{A.2.6}$$

A.3 Fluctuations in Free Carrier Absorption

The objective of this section is to use the fluctuations of the FCA loss coefficient (see Section A.2) to find the FCA loss fluctuation over the entire waveguide s_{FCA} as a function of mean power $\langle P(0)\rangle$ at the input

$$s_{FCA}^2 = \int_0^W \int_0^W \langle \Delta a_{FCA}(x) \Delta a_{FCA}(y)\rangle dxdy. \tag{A.3.1}$$

Here, W is the waveguide length. We are interested in the regime where $\langle P(0)\rangle \ll P'_{FCA} < P'_{TPA}$, because this is where peak capacity is obtained in the single channel communication case. In this regime, the FCA loss can be ignored in the transmission function since FCA is weaker than TPA, and the TPA loss can be treated as a perturbation in the transmission function. Using the analytical intensity input-output relationship (see Section A.2), the mean square of the power at point x is

$$\langle P^2(x)\rangle = \int_0^\infty \frac{e^{-2aL} P^2(0) p(P(0))}{\left(1+\frac{P(0)}{P_{TPA}}\right)^2} d[P(0)]. \tag{A.3.2}$$

When the input power is Gamma distributed as $p(P(0)) \sim (P(0))^{m-1} e^{-mP(0)/\langle P(0)\rangle}$, we can compute, by expanding the slower varying denominator in the integral in a Taylor series, to obtain

$$\langle P^2(x)\rangle \cong e^{-2ax} \frac{m+1}{m} \langle P(0)\rangle \left[1 - 2\frac{\langle P(0)\rangle}{P_{TPA}} \frac{(m+2)u(x)}{m}\right] \tag{A.3.3}$$

where $u(x) = (1-e^{-ax})/a$. In a similar manner, we can obtain the two point average

$$\langle P^2(x)P^2(y)\rangle \cong \langle P^2(x)\rangle \langle P^2(y)\rangle \frac{(m+2)(m+3)}{m(m+1)} \left[1 - 2\frac{\langle P(0)\rangle}{P_{TPA}} \frac{u(x)+u(y)}{m}\right]. \tag{A.3.4}$$

Using (A.3.4) and (A.2.6) in (A.3.1) we obtain:

$$s_{FCA}^2 = \left(\frac{\gamma T}{2}\right) \frac{\langle P(0)\rangle^4}{\left(P'_{TPA}P'_{FCA}\right)^2} \frac{(4m+6)(m+1)}{m^3} \left[1 + c^{(1)}\langle P(0)\rangle + \ldots\right] \tag{A.3.5}$$

where we identify the characteristic power level P_{FCA} as

$$P_{FCA} = \left(P'_{TPA}P'_{FCA}\right)^{1/2} \left(\frac{2m^3}{(\gamma T)(4m+6)(m+1)}\right)^{1/4}. \tag{A.3.6}$$

A.4 Two-Photon Absorption Loss Fluctuations

To obtain the fluctuations, we first solve (A.2.2) ignoring the FCA loss, since at the power levels where peak capacity occurs in WDM, the FCA term can be omitted from the transmission function. It is straight-forward to find for the output intensity

$$P(W) = \frac{P(0)}{1 + (W_{eff}/W)\left(P(0)/P'_{TPA}\right)} \tag{A.4.1}$$

where W denotes the waveguide length, and $W_{eff} = (1-e^{-aW})/a$. In the case of WDM, consider N_C channels present each with power $P_i(0)$ at the input of the waveguide. If each $P_i(0)$ is drawn independently from a Gamma distribution $P_i^{m-1}(0)e^{-mP_i(0)/\langle P(0)\rangle}$, then the total power $P_{TOT}(0) = \sum_i P_i(0)$ is distributed according to $[P_{TOT}(0)]^{N_C m-1} e^{-mP_{TOT}(0)/\langle P(0)\rangle}$. We can compute the moments of the transmission function $L = [1 + (W_{eff}/W)(P_{TOT}(0)/P'_{TPA})]^{-1}$ approximately

when TPA is weak ($\langle P(0)\rangle \ll P_{\text{TPA}}$). In this regime, while total loss is not significantly affected by TPA, fluctuations are increased by its presence. With the notation $P''_{\text{TPA}} \triangleq (W_{\text{eff}}/W)P'_{\text{TPA}}/N_C$ we have for the kth moment of L

$$\langle L^k\rangle \cong 1 - k(\langle P(0)\rangle/P''_{\text{TPA}}) + \frac{k(k+1)(N_C m+1)}{2N_C m}(\langle P(0)\rangle/P''_{\text{TPA}})^2 \tag{A.4.2}$$

which gives for the loss fluctuations

$$\langle \Delta L^2\rangle \cong (1/(N_C m))(\langle P(0)\rangle/P_{\text{TPA}})^2 \tag{A.4.3}$$

where here, $P_{\text{TPA}} = P''_{\text{TPA}}N_C^{1/2}$. Therefore, the moments of L in this particular case, depend only on the first moment of the input. Fluctuations in the transmission function, L, increase with average optical power. To 1st order in $\langle P(0)\rangle/P_{\text{TPA}}, \ln(L)$ is approximated by a constant times a Gamma distributed variable, and it's entropy $H(\ln L)$ is therefore approximately ($N_C \gg 1$)

$$H(\ln L) \cong \frac{1}{2}ln\left(\frac{2\pi e}{m}\frac{\langle P(0)\rangle^2}{P_{\text{TPA}}^2}\right) \tag{A.4.4}$$

and its mean value is $\langle \ln L\rangle \cong 1 - \langle L\rangle$.

A.5 Free-Carrier Absorption Loss Fluctuations in WDM Signaling

As we showed in Section A.1, when many previous pulses ($\gamma T \ll 1$) contribute to the FCA loss of the propagating pulse, the fluctuations are smoothed out. For the case of WDM transmission of N_C number of channels, when $\gamma T \sim 1$, a worst case scenario for the fluctuations of loss are obtained if we consider only the immediately previous symbol contributing to the FCA loss of the current symbol. Since the previous symbol, in WDM is actually a sum of N_C symbols over different channels, if the intensity of each channel is exponentially distributed the sum of the N_C intensities is Gamma distributed. We can compute the loss fluctuations due to FCA by starting from (A.2.1) with similar steps to Sections A.2 and A.3 to get

$$s^2_{\text{FCA}} \sim \frac{P^4 N_C(N_C+1)(4N_C+6)}{(P'_{\text{TPA}}P'_{\text{FCA}})^2}. \tag{A.5.1}$$

The total TPA plus FCA fluctuations are then approximately

$$s^2_{\text{TPA}} + s^2_{\text{FCA}} \sim \frac{P^2 N_C}{(P'_{\text{TPA}})^2}\left[1 + \frac{P^2(N_C+1)(4N_C+6)}{(P'_{\text{FCA}})^2}\right]. \tag{A.5.2}$$

Therefore, as long as the capacity peak occurs at a (per channel) power level where $(N_C+1)^{1/2}(4N_C+6)^{1/2}P < P'_{\text{FCA}}$, the peak will be determined by TPA.

APPENDIX B
Mutual Information Lower Bound

The key to obtaining the lower bound is the entropy power inequality (EPI) for two independent continuous random variables [8], Z_1, Z_2. EPI states that $e^{2H(Z_1+Z_2)} \geq e^{2H(Z_1)} + e^{2H(Z_2)}$ with equality if Z_1, Z_2 are normally distributed.

To simplify the entropy of the channel output $Y = LX + N$, we apply the entropy inequality on the variables LX and N

$$H(Y) \geq \frac{1}{2}\ln\left(e^{2H(LX)} + e^{2H(N)}\right). \tag{B.1}$$

With a view to applying the EPI again on variables $\ln(L)$ and $\ln(X)$, we switch from using the variable LX to using $\ln(LX)$. Using the transformation law for probability distributions,

$p(\ln(LX))d(\ln(LX)) = p(LX)d(LX)$, it is easy to see that $H(LX) = H(\ln(LX)) + \langle \ln(LX)\rangle$. Application of the EPI again gives

$$H(Y) \geq \frac{1}{2}\ln\left(e^{2H(N)} + e^{2\langle \ln X\rangle + 2\langle \ln L\rangle + 2H(\ln L)} + e^{2H(X)+2\langle \ln L\rangle}\right). \tag{B.2}$$

To simplify the conditional entropy distribution, we note that the entropy of the random variable $Y|X$ is always lower than the entropy of a normal variable with equal variance. Therefore

$$-\langle \ln p(Y|X)\rangle_{Y|X} \leq \frac{1}{2}\ln\left(2\pi e\sigma^2_{Y|X}\right) \tag{B.3}$$

where $\sigma^2_{Y|X} = \langle Y^2\rangle_{L,N} - \langle Y\rangle^2_{L,N}$ is the variance of the variable $Y|X$. Using the above inequality, and simplifying the variance of $Y|X$, we obtain

$$H(Y|X) \leq \frac{1}{2}\ln(2\pi e) + \frac{1}{2}\langle \ln(\langle N^2\rangle + \langle \Delta L^2\rangle)\rangle_X. \tag{B.4}$$

Then, the above two bounds are combined to yield a bound on the mutual information $I(Y;X)$ given in (1). The inequality $\langle \ln(1 + cX^2)\rangle \leq \ln(1 + c\langle X^2\rangle)$ can simplify equation (B.4) more if needed.

APPENDIX C
Fluctuations Due to Free Carrier Absorption (FCA) and Two Photon Absorption (TPA) and Characteristic Powers for FCA and TPA

This section provides the channel parameters derived in Appendix A. The mean total loss in the waveguide, due to linear losses, TPA and FCA, is l. The loss fluctuation (standard deviation) parameter is decomposed as $s^2 = s^2_{\text{FCA}} + s^2_{\text{TPA}}$, where s_{FCA} is ignored in WDM and s_{TPA} is ignored in the single-channel calculation as justified in Appendix A. For a single channel or multiple-channel WDM, s_{TPA} and s_{FCA} are, respectively

$$s^2_{\text{FCA}} = \frac{P^4}{P^4_{\text{FCA}}},\ P_{\text{FCA}} = \left(P'_{\text{TPA}}P'_{\text{FCA}}\right)^{1/2}\left(\frac{\tau}{10T}\right)^{1/4} \tag{C.1}$$

$$s^2_{\text{TPA}} \cong \frac{P^2}{P^2_{\text{TPA}}},\ P_{\text{TPA}} = (W_{\text{eff}}/W_{wg})P'_{\text{TPA}}/N_C^{1/2} \tag{C.2}$$

with P the optical power per channel at the input of the waveguide, $P'_{\text{TPA}} = A_{\text{eff}}/(\beta W_{wg})$ and $P'_{\text{FCA}} = (A_{\text{eff}}/\sigma)(2\hbar\omega/\tau)$ are characteristic TPA and FCA powers, respectively. The effective area of the waveguide is A_{eff}, W_{wg} is the waveguide length, and $W_{\text{eff}} = (1 - e^{-a_{\text{LIN}}W_{wg}})/a$ where a_{LIN} is the linear loss coefficient.

APPENDIX D
Four Wave Mixing

The refractive index n depends on the intensity in the form of $n = n_o + n_2 I$. Here, n_o is the linear index and n_2 is a constant characterizing the strength of the nonlinearity. The nonlinear coefficient n_2 in silicon is large rendering nonlinearity observable even in short distances on a chip. In the case of single channel (non WDM) the principle effect is self-phase modulation (SPM) where the refractive index and hence the phase depends on the intensity. In multi-channel (WDM) communication, this nonlinearity creates two other effects. The phase in a given channel is influenced by intensities in other channels in a process called cross-phase modulation (CPM). Also, electric field amplitudes in two channels influence the phase in the third channel in a process called four-wave mixing (FWM). FWM depends on phase matching between the three fields and

is curbed, naturally or intentionally via pre-chirping [24], via dispersion. FWM was also ignored in calculating the capacity of optical fibers [13] because dispersion in fiber is significant.

APPENDIX E
Note on Optimum Input Symbol Distribution

Here, we add a supporting note on the optimization of the lower bound in Section 4. In particular we aim to show what would happen if $\langle \Delta L^2 \rangle$ is included in the optimization. The method of Lagrange multipliers shows the distribution must now have the form

$$p(X) = \frac{CX^{a-1}e^{-\lambda X}}{\left(\langle \Delta L^2 \rangle X^2 + \langle N^2 \rangle\right)^{b/2}} \exp\left(\sum_{k=0}^{K} \lambda_k M_k(X)\right). \tag{E.1.1}$$

This distribution is further maximized by varying the λ_k's along with a, b (C is the normalization constant). Ignoring $\langle \Delta L^2 \rangle$ in the optimization is equivalent to setting λ_k's to zero. While the resulting bound may be suboptimal (lower than maximum value the bound can reach), the benefit is obtaining a closed form result that clearly reveals the peaking of capacity due to fluctuations caused by nonlinear losses: one that is intuitively appealing and one which clearly unveils the relevant physics of the problem.

APPENDIX F
Note on FWM Noise Computation

Here we show how to perform the summation of terms in equation (21). The terms in the sum that are non-zero have $j = n$, $k = m$ or $j = m$, $k = n$. Also, we must have $l = q$ (note that $j \neq l$, $k \neq l$, $n \neq q$, $m \neq q$). Therefore, the sum is simplified

$$\frac{1}{2}\left(\frac{\gamma_{\text{Kerr}} W}{A_{\text{eff}}}\right)^2 \left(2\sum_{j\neq,k,l} \left\langle \varepsilon_j \varepsilon_j^* \varepsilon_k \varepsilon_k^* \varepsilon_l \varepsilon_l^* \right\rangle + \sum_{j=k,l} \left\langle \varepsilon_j \varepsilon_j^* \varepsilon_k \varepsilon_k^* \varepsilon_l \varepsilon_l^* \right\rangle\right). \tag{F.1}$$

The sum must be evaluated for Gaussian distributed in-phase and quadrature components in which case $\langle \varepsilon_j \varepsilon_j^* \rangle = 2P$ and $\langle \varepsilon_j \varepsilon_j^* \varepsilon_j \varepsilon_j^* \rangle = 8P^2$.

For the ith frequency band $(i = 0, 1, \ldots, N-1)$ we want to calculate the number of combinations of the indices j, k, l that have $i = j + k - l$, where $j \neq i$ and $k \neq i$ (since remaining contributions are included in the CPM term in the propagation equation).

For $= 0, 1, \ldots, i-1, i+1, \ldots, N-1$, $d = i - j = k - l$ runs over values $i, i-1, \ldots, 1, -1, \ldots, -(N-1-i)$. To see how many values of k, l correspond to each value of j note that in a matrix of dimension $N \times N$, where k labels columns and l labels row, values of d label matrix diagonals, where $d = 0$ is the main diagonal, and $d > 0$ labels an upper diagonal. Diagonal d is comprised of $N - |d|$ matrix elements.

Summing up all the values of k, l that correspond to the above values of d, means summing up over diagonals labeled by the values of d. From this summation we must exclude the values that have $k = i$, a total of $N - 1$ cases: For every value of d in that range, there is a value of l for which $k = i$.

Therefore, the number of terms is equal to

$$\sum_{d=1}^{i} (N - |d|) + \sum_{d=-(N-1-i)}^{-1} (N - |d|) - (N-1). \tag{F.2}$$

Using the sum $1 + 2 + \cdots + q = q(q+1)/2$, we find

$$(N-1)^2 - \frac{i(i+1)}{2} - \frac{(N-1-i)(N-i)}{2}. \tag{F.3}$$

Using the above results, we obtain the total noise power in a single channel due to FWM:

$$\frac{1}{2}\left(\frac{\gamma_{\text{Kerr}}W}{A_{\text{eff}}}\right)^2 \times 16P^3 \times \left[(N-1)(N-1) - \frac{i(i+1)}{2} - \frac{(N-1-i)(N-i)}{2}\right]. \quad \text{(F.4)}$$

References

[1] A. F. Benner, M. Ignatowski, J. A. Kash, D. M. Kuchta, and M. B. Ritter, "Exploitation of optical interconnects in future server architectures," *IBM J. Res. Develop.*, vol. 49, no. 4/5, pp. 755–775, Jul. 2005.
[2] P. K. Pepeljugoski *et al.,* "Data center and high performance computing intreconnects for 100 Gb/s and beyond," in *Proc. Opt. Fiber Commun. Conf.*, 2007, pp. 1–3.
[3] J. Wolfe, *Why IBM and Intel are Chasing the $100 B Opportunity in Nanophotonics*. New York, NY, USA: Forbes, Feb. 2012.
[4] C. Dillow, "Seeing the data center in a new light," *Fortune*, Sep. 2013. [Online]. Available: http://tech.fortune.cnn.com/2013/09/11/intel-silicon-photonics/
[5] B. Jalali, "Making silicon lase," *Sci. Amer.*, vol. 296, no. 2, pp. 58–65, Feb. 2007.
[6] D. A. B. Miller, "Optical interconnects to electronic chips," *Appl. Opt.*, vol. 49, no. 25, pp. F59–F70, Jul. 2010.
[7] D. Dimitropoulos and B. Jalali, "Optical information capacity of silicon," arXiv preprint arXiv:1409.0095, Aug. 2014.
[8] D. Dimitropoulos and B. Jalali, "Optical information capacity of silicon," in *Proc. OFC*, 2015.
[9] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, Jan. 2001.
[10] B. Jalali, and S. Fathpour, "Silicon photonics," *J. Lightw. Technol.*, vol. 24, no. 12, pp. 4600–4615, Dec. 2006.
[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 2006.
[12] M. J. Kobrinsky, "On-chip optical interconnects," *Intel Technol. J.*, vol. 8, no. 2, pp. 129–142, 2004.
[13] P. P. Mitra and J. B. Stark, "Nonlinear limits to the information capacity of optical fibre communications," *Nature*, vol. 411, no. 6841, pp. 1027–1030, Apr. 2001.
[14] R.-J. Essiambre, G. J. Foschini, G. Kramer, and P. J. Winzer, "Capacity limits of information transport in fiber-optic networks," *Phys. Rev. Lett.*, vol. 101, no. 16, Oct. 2008, Art. ID. 163901.
[15] J. Tang, "The Shannon channel capacity of dispersion-free nonlinear optical fiber transmission," *J. Lightw. Technol.*, vol. 19, no. 8, pp. 1104, Aug. 2001.
[16] K. S. Turitsyn, S. A. Derevyanko, I. V. Yurkevich, and S. K. Turitsyn, "Information capacity of optical fiber channels with zero average dispersion," *Phys. Rev. Lett.*, vol. 91, no.20, Nov. 2003, Art. ID. 203901.
[17] L. G. L. Wegener *et al.,* "The effect of propagation nonlinearities on the information capacity of WDM optical fiber systems: Cross-phase modulation and four-wave mixing," *Phys. D, Nonlinear Phenomena*, vol. 189, no. 1, pp. 81–99, Feb. 2004.
[18] M. H. Taghavi, G. C. Papen, and P. H. Siegel, "On the multiuser capacity of WDM in a nonlinear optical fiber: Coherent communication," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5008–5022, Nov. 2006.
[19] R. J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.
[20] A. D. Ellis, J. Zhao, and D. Cotter, "Approaching the non-linear Shannon limit," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 423–433, Feb. 2010.
[21] B. Jalali, "Silicon photonics: Nonlinear optics in the mid-infrared," *Nature Photon.*, vol. 4, no. 8, pp. 506–508, Aug. 2010.
[22] N. Suzuki, "FDTD analysis of two-photon absorption and free-carrier absorption in Si high-index-contrast waveguides," *J. Lightw. Technol.*, vol. 25, no. 9, pp. 2495–2501, Sep. 2007.
[23] H. Yamada *et al.,* "Nonlinear-optic silicon-nanowire waveguides," *Jpn. J. Appl. Phys.*, vol. 44, no. 9R, p. 6541, Sep. 2005.
[24] G. P. Agrawal, *Fiber-Optic Communication Systems*. Hoboken, NJ, USA: Wiley 1997.
[25] Y. Yamamoto and H. A. Haus, "Preparation, measurement and information capacity of optical quantum states," *Rev. Modern Phys.*, vol. 58, no. 4, p. 1001, Oct.1986.
[26] A. Lapidoth, "On phase noise channels at high SNR," in *Proc. IEEE Inf. Theory Workshop*, 2002, pp. 1–4.
[27] D. Dimitropoulos, V. Raghunathan, R. Claps, and B. Jalali, "Phase-matching and nonlinear optical processes in silicon waveguides," *Opt. Exp.*, vol. 12, no. 1, pp. 149–160, Jan. 2004.
[28] S. Benedetto and E. Biglieri, "Nonlinear equalization of digital satellite channels," *IEEE J. Sel. Areas Commun.*, vol. 1, no. 1, pp. 57–62, Jan.1983.
[29] O. Boyraz, T. Indukuri, and B. Jalali, "Self-phase-modulation induced spectral broadening in silicon waveguides," *Opt. Exp.*, vol. 12, no. 5, pp. 829–834, Mar. 2004.
[30] D. Dimitropoulos, R. Jhaveri, R. Claps, J. C. S. Woo, and B. Jalali, "Lifetime of photogenerated carriers in silicon-on-insulator rib waveguides," *Appl. Phys. Lett.*, vol. 86, Feb. 2005, Art. ID. 071115.
[31] K. S. Turitsyn and S. K. Turitsyn, "Nonlinear communication channels with capacity above the linear Shannon limit," *Opt. Lett.*, vol. 37, no. 17, pp. 3600–3602, 2012.
[32] E. Agrell, "Conditions for a monotonic channel capacity," arXiv:1209.2820, 2012.