Research Article | Vol. 26, No. 21 | 15 Oct 2018 | OPTICS EXPRESS 27118
Optics EXPRESS

Upper and lower bounds to the information rate transferred through the Pol-Mux channel

Arnaldo Spalvieri,[1] Luca Reggiani,[1] and Laura Dossi[2,*]

[1] *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy*
[2] *IEIIT, Consiglio Nazionale delle Ricerche, c/o DEIB, Politecnico di Milano, Italy*
[*] *laura.dossi@ieiit.cnr.it*

Abstract: Pol-Mux transmission is a well established technique that enhances spectral efficiency by simultaneously transmitting over horizontal and vertical polarizations of the electrical field. However, cross-coupling of the two polarizations impairs transmission. Under the assumption that the cross-coupling matrix is a Markov process with free-running state, we propose upper and lower bounds to the information rate that can be transferred through the channel. Simulation results show that the two bounds are tight for values of the cross-coupling power of practical interest and modulation formats up to 16-QAM (quadrature amplitude modulation).

OCIS codes: (060.1660) Coherent communications; (060.4080) Modulation; (060.4510) Optical communications.

References and links

1. M.S.A.S. Al Fiad, M. Kuschnerov, S.L. Jansen, T. Wuth, D. van den Borne, and de H. Waardt, "11 x 224-Gb/s POLMUX-RZ-16QAM transmission over 670 km of SSMF with 50-GHz channel spacing," IEEE Photon. Technol. Lett. **22**(15), 1150–1152 (2010).
2. V.A.J.M. Sleiffer, M.S.A.S. Al Fiad, D. van den Borne, M. Kuschnerov, V. Veljanovski, M. Hirano, Y. Ya mamoto, T. Sasaki, S.L. Jansen, T. Wuth, and H. de Waardt, "10 x 224-Gb/s POLMUX-16QAM transmission over 656 km of Large-A_e PSCF with a spectral efficiency of 5.6 b/s/Hz," IEEE Photon. Technol. Lett. **23**(20), 1427–1429 (2011).
3. P. Boffi, M. Ferrario, L. Marazzi, P. Martelli, P. Parolari, A. Righetti, R. Siano, and M. Martinelli, "Stable 100-Gb/s POLMUX-DQPSK transmission with automatic polarization stabilization," IEEE Photon. Technol. Lett. **21**(11), 745–747 (2009).
4. Gerard J. Foschini, and Michael J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," Wireless Pers. Commun. **6**(3), 311–335 (1998).
5. Seb.J. Savory, "Digital coherent optical receivers: Algorithms and subsystems," IEEE J. Sel. Top. Quantum Electron. **16**(5), 1164–1179 (2010).
6. L. Li, Z. Tao, L. Liu, W. Yan, S. Oda, T. Hoshida, and Jens C. Rasmussen, "Nonlinear polarization crosstalk canceller for dual-polarization digital coherent receivers," presented at Optical Fiber Communication, collocated National Fiber Optic Engineers Conference (OFC/NFOEC), IEEE, Piscataway, NJ, USA, 21 March 2010.
7. P. Layec, A. Ghazisaeidi, G. Charlet, Jean-Christophe Antona, and S. Bigo, "Generalized maximum likelihood for cross-polarization modulation effects compensation," J. Lightwave Technol. **33**(7), 1300–1307 (2015).
8. J. Li, R. Schmogrow, D. Hillerkuss, Philipp C. Schindler, M. Nazarathy, C. Schmidt-Langhorst, Shalva-Ben Ezra, I. Tselniker, C. Koos, W. Freude, and J. Leuthold, "A self-coherent receiver for detection of PolMUX coherent signals," Opt. Express **20**(19), 21413–21433 (2012).
9. R. H. Etkin, and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channels," IEEE T. Inform. Theory **52**(4), 1576–1608 (2006).
10. L. Barletta, M. Magarini, S. Pecorino, and A. Spalvieri, "Upper and lower bounds to the information rate transferred through first-order Markov channels with free-running continuous state," IEEE T. Inform. Theory **60**(7), 3834–3844 (2014).
11. L. Barletta, M. Magarini, and A. Spalvieri, "Estimate of information rates of discrete-time first-order Markov phase noise channels," IEEE Photonic Tech. L. **23**(21), 1582–1584 (2011).
12. L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," J. Lightwave Technol. **30**(10), 1480–1486 (2012).
13. L. Barletta, M. Magarini, and A. Spalvieri, "A new lower bound below the information rate of Wiener phase noise channel based on Kalman carrier recovery," Opt. Express **20**(23), 2547–25477 (2012).

#330877 https://doi.org/10.1364/OE.26.027118
Journal © 2018 Received 11 May 2018; revised 12 Jul 2018; accepted 16 Jul 2018; published 3 Oct 2018

Optics EXPRESS

1. Introduction

Simultaneous transmission of modulated signals over the horizontal and vertical polarizations of the electrical field is a well established technique [1–3] that allows to improve spectral efficiency by using the same frequency twice. In its essence, this technique relies upon the principle of MIMO (Multiple Input Multiple Output) systems, that have become popular after the seminal paper [4]. To cancel interference arising from non-ideal orthogonality between the horizontal and the vertical polarizations, linear processing can be adopted [5], even if it is well known that non-linear techniques achieve better performance in presence of interference and additive noise, see e.g. [6, 7].

Either implicitly or explicitly, most of the receivers studied in the literature assume that the MIMO channel matrix is static or quasi-static. However, the experimental results of [6] show that the coherence time of the channel is quite small, say, in the order of 10 to 30 symbol intervals for 112 Gb/s dual-polarization QPSK (Quadrature Phase Shift Keying). Hence tracking the channel becomes an issue. Tracking techniques can be based on pilot symbols, as proposed, for instance, in [8], but, independently of the channel tracking method, a low coherence time of the channel matrix, hence a fast time-varying channel, will make noisy the channel estimate (in practice only a short time window spanning a few signal samples can be used for channel estimation at a given time instant) thus impacting the information rate that can be transmitted through the channel. This observation motivates the study of the information rate transferred through the Pol-Mux channel. Channel capacity of the fading MIMO channel is a classical topic in the general framework of information theory, see e.g. [9] and, in that context, also the information rate of channels with free-running state has been studied [10]. In the context of optical transmission the information rate is well studied for the phase noise channel, at least for the channel model with free-running state, see e.g. [11–13], but less has been done for the Pol-Mux channel, which can be seen indeed as a variant of the phase noise channel where

- the modulus is not constant
- the channel is MIMO.

Therefore, starting from the lower bound for the phase noise channel of [13], we adapt it here to the Pol-Mux channel and introduce a new upper bound based on the Kalman filter.

2. Channel model

Let the lowercase characters indicate possibly complex scalars and column vectors and let the uppercase characters indicate matrices. The notation a_k^{k+i} is used to indicate a column vector (or matrix, when the elements are vectors) made by the chunk of sequence $(a_k, a_{k+1}, \cdots, a_{k+i})^T$, while $\{a_k\}$ is used to indicate the semi-infinite sequence $(a_0, a_1, \cdots)$. The notation $\mathcal{I}_m$ is used to indicate the $m \times m$ identity matrix and the superscript ^{H} denotes Hermitian transposition. The output of the Pol-Mux channel at time k is

$$y_k = M_k x_k + w_k, \quad k = 1, 2, \cdots, \tag{1}$$

where x_k is the k-th sample of the i.i.d. input modulation complex vector data sequence, with zero mean vector and covariance matrix

$$E\{x_k x_k^H\} = \mathcal{I}_2, \tag{2}$$

M_k is the channel matrix and w_k is the k-th element of the i.i.d. complex Gaussian vector noise sequence with zero mean vector and covariance matrix

$$E\{w_k w_k^H\} = \sigma^2 \mathcal{I}_2. \tag{3}$$

For small to moderate polarization crosstalk, the matrix M_k can be modelled as [6]

$$M_k = \begin{pmatrix} 1 & \lambda_{1,k} \\ \lambda_{2,k} & 1 \end{pmatrix}, \tag{4}$$

where

$$\lambda_k = (\lambda_{1,k}, \lambda_{2,k})^T$$

is the k-th element of a complex Gaussian random vector sequence which is hereafter modelled here as a free-running 1-causal ARMA (Autoregressive Moving Average) process, hence

$$\lambda_k = \sum_{i=1}^{p} b_i v_{k-i} + \sum_{i=1}^{q} a_i \lambda_{k-i}, \tag{5}$$

where v_k is the k-th sample of a white Gaussian random vector sequence with zero mean and covariance matrix

$$E\{v_k v_k^H\} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{6}$$

In other words, $\{\lambda_k\}$ is the filtered version of $\{v_k\}$, where the filter is made of two shift registers, one for $\{v_{1,k}\}$ and the other one for $\{v_{2,k}\}$, each one with m memories, and with 1-causal feedback taps a_1^q and 1-causal forward taps b_1^p, with

$$m = \max\{p, q\}.$$

Using the z-transform you write

$$\lambda(z) = v(z)\frac{b(z)}{1 - a(z)}, \tag{7}$$

where

$$b(z) = \sum_{i=1}^{m} b_i z^{-i}, \quad a(z) = \sum_{i=1}^{m} a_i z^{-i}. \tag{8}$$

To cast the model in the framework of linear dynamic systems we need to define the state of the system. To this aim, let us define the vector sequence

$$\omega_k = (\omega_{1,k}, \omega_{2,k})^T = v_k + \sum_{i=1}^{m} a_i \omega_{k-i}, \quad k = 0, 1, \cdots,$$

hence ω_{k-m}^{k-1} is the content of the two shift registers at the k-th channel use. Note that λ_k depends only on ω_{k-m}^{k-1} as

$$\lambda_k = \sum_{i=1}^{m} b_i \omega_{k-i}$$

and, given ω_{k-m}^{k-1}, the sequence λ_k^∞ is independent of λ_1^{k-1}. Therefore you can take

$$s_k = (1, (\omega_{1,k-m}^{k-1})^T, 1, (\omega_{2,k-m}^{k-1})^T)^T \tag{9}$$

as the state of the linear dynamic system at time k, thus writing the measurement equation and the state transition equation as

$$y_k = H_k s_k + w_k, \tag{10}$$

$$s_{k+1} = F s_k + (0, v_{1,k}, (0_1^{m-1})^T, 0, v_{2,k}, (0_1^{m-1})^T)^T, \tag{11}$$

Optics EXPRESS

with

$$H_k = \begin{bmatrix} x_{1,k} & x_{2,k}(b_1^m)^T & 0 & (0_1^m)^T \\ 0 & (0_1^m)^T & x_{2,k} & x_{1,k}(b_1^m)^T \end{bmatrix}, \tag{12}$$

where 0_1^m is a column vector of m zeros, and the $2(m+1) \times 2(m+1)$ state transition matrix is

$$F = \begin{bmatrix} F_{m+1} & O_{m+1} \\ O_{m+1} & F_{m+1} \end{bmatrix}, \tag{13}$$

where

$$F_{m+1} = \begin{bmatrix} 1 & (0_1^{m-1})^T & 0 \\ 0 & (a_1^{m-1})^T & a_m \\ 0_1^{m-1} & \mathcal{I}_{m-1} & 0_1^{m-1} \end{bmatrix}, \tag{14}$$

and O_m is the all-zero square matrix of size $m \times m$. The state transition probability is

$$p(s_{k+1}|s_k) = g_c(Fs_k, Q; s_{k+1}), \tag{15}$$

where $g_c(\mu, \Sigma_m; x)$ indicates a m-dimensional complex Gaussian probability density function over the complex vector space spanned by x with mean vector μ and covariance matrix Σ_m and Q is the covariance matrix of the process noise $(0, v_{1,k}, (0_1^{m-1})^T, 0, v_{2,k}, (0_1^{m-1})^T)^T$, that is

$$Q = \begin{bmatrix} Q_1 & Q_\rho \\ Q_\rho & Q_1 \end{bmatrix}, \tag{16}$$

with

$$Q_1 = \begin{bmatrix} 0 & 0 & (0_1^{m-1})^T \\ 0 & 1 & (0_1^{m-1})^T \\ 0_1^{m-1} & 0_1^{m-1} & O_{m-1} \end{bmatrix}, \tag{17}$$

$$Q_\rho = \begin{bmatrix} 0 & 0 & (0_1^{m-1})^T \\ 0 & \rho & (0_1^{m-1})^T \\ 0_1^{m-1} & 0_1^{m-1} & O_{m-1} \end{bmatrix}. \tag{18}$$

The joint source and channel output probability, given the hidden state, is

$$\begin{aligned} p(y_k, x_k|s_k) &= p(x_k|s_k)p(y_k|x_k, s_k) \\ &= p(x_k)p(y_k|x_k, s_k) \end{aligned} \tag{19}$$

where

$$p(y_k|x_k, s_k) = g_c(H_k s_k, \sigma^2 \mathcal{I}_2; y_k). \tag{20}$$

The conditional probability of channel output given the hidden state is

$$p(y_k|s_k) = \sum_{x_k \in \mathcal{X}_k} p(x_k)p(y_k|x_k, s_k). \tag{21}$$

3. Upper and lower bounds to the information rate by the Kalman filter

Let

$$I(x;y) = H(x) - H(x|y), \tag{22}$$

where, for conventional M-QAM (Multi-Level Quadrature Amplitude Modulation) and M-PSK (Multi-Level Phase Shift Keying)

$$H(x) = \log_2 \mathrm{M}. \tag{23}$$

For the conditional entropy, by chain rule one writes

$$H(x|y) = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} H(x_k | x_1^{k-1}, y_1^N), \tag{24}$$

which, by the Shannon-McMillan-Breiman theorem, can be evaluated as

$$H(x|y) = -\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \log_2 p(x_k | x_1^{k-1}, y_1^N). \tag{25}$$

Since conditioning does not increase entropy, we have the following upper and lower bounds to the conditional entropy

$$\begin{aligned} H(x|y) &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} H(x_k | x_1^{k-1}, y_1^N) \\ &\le \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} H(x_k | x_1^{k-1}, y_1^k) \\ &= -\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \log_2 p(x_k | x_1^{k-1}, y_1^k), \end{aligned} \tag{26}$$

$$\begin{aligned} H(x|y) &= \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} H(x_k | x_1^{k-1}, y_1^N) \\ &\ge \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} H(x_k | x_1^{k-1}, x_{k+1}^N, y_1^N) \\ &= -\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \log_2 p(x_k | x_1^{k-1}, x_{k+1}^N, y_1^N), \end{aligned} \tag{27}$$

that one can use in a straightforward way in the right side of (22) together with (23) to get lower and upper bounds to the information rate.

Let us consider the upper bound (26). The probabilities inside the logarithm can be evaluated by the Kalman filter as follows. The knowledge of past transmitted symbols that appear in the conditioning is imported in the Kalman filter by including all the conditions in the measurement, hence by updating the Kalman filter in data-aided mode. Let us write the channel output as

$$y_k = H_k s_k + w_k = h_k(s_k) + w_k. \tag{28}$$

The predicted measurement at time k is

$$\hat{y}_k = H_k \hat{s}_k, \tag{29}$$

Optics EXPRESS

where $\hat{s}_k$ denotes the state predicted by the Kalman filter at time k, that is the expectation of the hidden state given past measurements

$$\hat{s}_k = E\{s_k | y_1^{k-1}, x_1^{k-1}\}. \tag{30}$$

As innovations process we take

$$u_k = y_k - \hat{y}_k = H_k(s_k - \hat{s}_k) + w_k. \tag{31}$$

Starting from an initial pair $(\hat{\Sigma}_1, \hat{s}_1)$, where

$$\hat{\Sigma}_k = E\{(s_k - \hat{s}_k)(s_k - \hat{s}_k)^H\}, \tag{32}$$

for $k = 1, 2, \cdots$, the state prediction vector and the prediction error covariance matrix evolve as

$$\hat{s}_{k+1} = F(\hat{s}_k + K_k u_k), \tag{33}$$

$$\hat{\Sigma}_{k+1} = F\Sigma_k F^T + Q, \tag{34}$$

where

$$\Sigma_k = ((\hat{\Sigma}_k)^{-1} + \sigma^{-2} H_k^H H_k)^{-1}, \tag{35}$$

$$K_k = \sigma^{-2} \Sigma_k H_k^H. \tag{36}$$

The desired probability is evaluated as

$$\begin{aligned} p(x_k | x_1^{k-1}, y_1^k) &= p(x_k | y_k, x_1^{k-1}, y_1^{k-1}) \\ &= \frac{p(x_k | x_1^{k-1}, y_1^{k-1}) p(y_k | x_k, x_1^{k-1}, y_1^{k-1})}{\sum_{x_k \in \mathcal{X}_k} p(x_k | x_1^{k-1}, y_1^{k-1}) p(y_k | x_k, x_1^{k-1}, y_1^{k-1})} \\ &= \frac{p(x_k) p(y_k | x_1^k, y_1^{k-1})}{\sum_{x_k \in \mathcal{X}_k} p(x_k) p(y_k | x_1^k, y_1^{k-1})}, \end{aligned} \tag{37}$$

where, using the predicted state and the prediction error covariance matrix computed by the Kalman filter, one has

$$\begin{aligned} p(y_k | x_1^k, y_1^{k-1}) &= \int_{\mathcal{S}} p(s_k, y_k | x_1^k, y_1^{k-1}) ds_k \\ &= \int_{\mathcal{S}} p(s_k | x_1^k, y_1^{k-1}) p(y_k | s_k, x_1^k, y_1^{k-1}) ds_k \\ &= \int_{\mathcal{S}} p(s_k | x_1^{k-1}, y_1^{k-1}) p(y_k | s_k, x_k) ds_k \\ &= \int_{\mathcal{S}} g_c(\hat{s}_k, \hat{\Sigma}_k; s_k) g_c(H_k s_k, \sigma^2 \mathcal{I}_2; y_k) ds_k \\ &= g_c(H_k \hat{s}_k, H_k^H \hat{\Sigma}_k H_k + \sigma^2 \mathcal{I}_2; y_k). \end{aligned} \tag{38}$$

Similarly, for the lower bound to the conditional entropy, one has

$$p(x_k | x_1^{k-1}, x_{k+1}^N, y_1^N) = \frac{p(x_k) p(y_k | x_1^N, y_1^{k-1}, y_{k+1}^N)}{\sum_{x_k \in \mathcal{X}_k} p(x_k) p(y_k | x_1^N, y_1^{k-1}, y_{k+1}^N)}, \tag{39}$$

Optics EXPRESS

with

$$
\begin{aligned}
p(y_k|x_1^N, y_1^{k-1}, y_{k+1}^N) &= \int_{\mathcal{S}} p(s_k, y_k|x_1^N, y_1^{k-1}, y_{k+1}^N) ds_k \\
&= \int_{\mathcal{S}} p(s_k|x_1^N, y_1^{k-1}, y_{k+1}^N) p(y_k|s_k, x_1^N, y_1^{k-1}, y_{k+1}^N) ds_k \\
&= \int_{\mathcal{S}} p(s_k|x_1^{k-1}, y_1^{k-1}, x_{k+1}^N, y_{k+1}^N) p(y_k|s_k, x_k) ds_k \\
&= \int_{\mathcal{S}} g_c(\hat{s}_{fb,k}, \hat{\Sigma}_{fb,k}; s_k) g_c(H_k s_k, \sigma^2 \mathcal{I}_2; y_k) ds_k \\
&= g_c(H_k \hat{s}_{fb,k}, H_k^H \hat{\Sigma}_{fb,k} H_k + \sigma^2 \mathcal{I}_2; y_k),
\end{aligned} \tag{40}
$$

where $\hat{s}_{fb,k}$ and $\hat{\Sigma}_{fb,k}$ are the estimates produced by combining a forward and a backward Kalman filter as

$$
\hat{s}_{fb} = \hat{\Sigma}_b (\hat{\Sigma}_f + \hat{\Sigma}_b)^{-1} \hat{s}_f + \hat{\Sigma}_f (\hat{\Sigma}_f + \hat{\Sigma}_b)^{-1} \hat{s}_b, \tag{41}
$$

$$
\hat{\Sigma}_{fb} = (\hat{\Sigma}_f^{-1} + \hat{\Sigma}_b^{-1})^{-1}. \tag{42}
$$

4. Simulation results

The consideration of realistic spectra of the cross-pol coefficients is out of the scope of the present paper and we left it to future studies. For practical methods, to estimate the strength of cross-pol interference the reader is referred to [6], where the strength of interference is given by the autocorrelation of interference at time zero. In the following we express the strength of interference by using the SIR (Signal-to-Interference Ratio), which is the inverse of the interference autocorrelation at time zero. To derive simulation results, we set $\rho = 0$ and for each one of the two random coefficients appearing in the Pol-Mux matrix we take the first-order ARMA model

$$
\lambda(z) = v(z) \frac{(1 - z_p) z^{-1}}{1 - z_p z^{-1}}, \tag{43}
$$

where $-1 < z_p < 1$ is the pole of the first-order ARMA model. The filtered sequence has zero mean, unit power spectral density at frequency zero and power

$$
E\{\lambda_k^2\} = \frac{1 - z_p}{1 + z_p},
$$

hence the SIR is

$$
\text{SIR} = \frac{1 + z_p}{1 - z_p}. \tag{44}
$$

In the common case where z_p is close to 1, the filtered sequence is a first-order low-pass random sequence with -3 dB normalized bandwidth

$$
B_{-3} \approx \frac{1 - z_p}{2\pi}.
$$

Figure 1 gives the upper and lower bounds to the information rate of 4-QAM, 16-QAM and 64-QAM obtained with $z_p = 0.977$, corresponding to SIR=19.3 dB. With such moderate interference the two bounds are close to each other, also for 64-QAM. Moreover, at high values of SNR (Signal-to-Noise Ratio) information rates reach the maximum value allowed by the

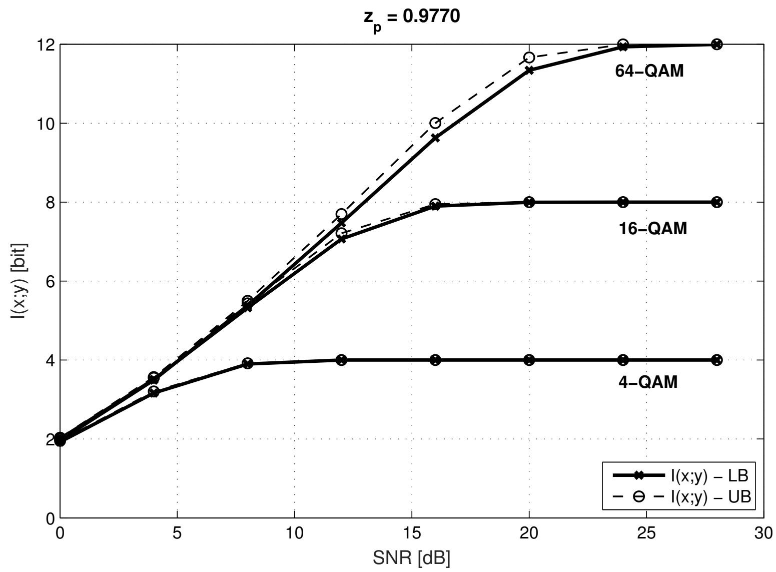

Fig. 1. Upper and lower bounds to the information rate for various modulation formats and $z_p = 0.977$. The Signal-to-Noise Ratio (SNR) is SNR=$\frac{1}{\sigma^2}$.

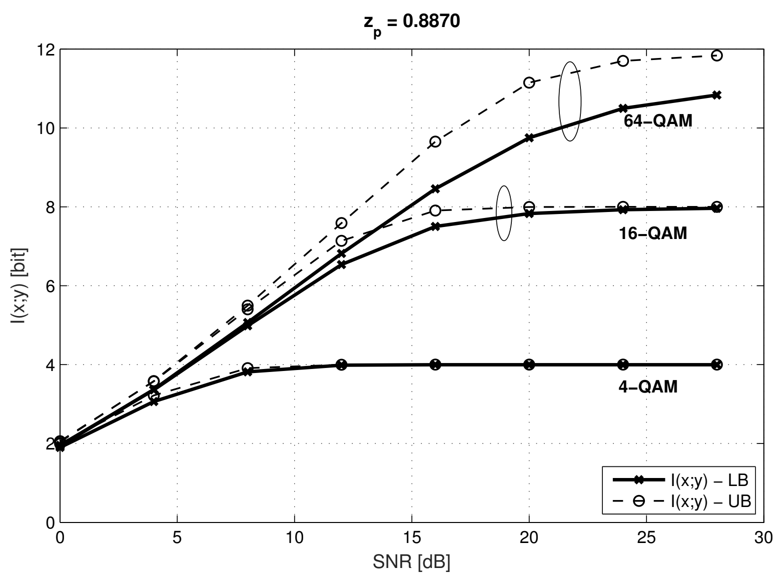

Fig. 2. Upper and lower bounds to the information rate for various modulation formats and $z_p = 0.887$. The Signal-to-Noise Ratio (SNR) is SNR=$\frac{1}{\sigma^2}$.

constellation sizes, achievable with the pure AWGN (Additive White Gaussian Noise) channel: 4 bits for $2 \times$ 4-QAM, 8 bits for $2 \times$ 16-QAM and 12 bits for $2 \times$ 64-QAM.

Figure 2 gives the same upper and lower bounds obtained with $z_p = 0.887$, that is SIR=12.2 dB. In the practice it seems to be a strong interference condition, since the minimum SIR reported in the experimental results of [6] is around 14 dB. In this case, the information rate with 64-QAM and at high SNR remains well below the information rate achieved with the AWGN channel, thus confirming that the Pol-Mux interference becomes the limiting factor of the information rate transferred through the channel. We also note that the spread between upper and lower bounds

Optics EXPRESS

becomes large with 64-QAM and at high SNR, where the capability of tracking the MIMO channel becomes crucial. Actually, the lower bound renounces to the blind part of tracking thus renouncing to some tracking capability, while the upper bound upgrades the blind tracking to a data-aided tracking, thus enhancing tracking capabilities over what can actually be done.

5. Conclusions

We have proposed upper and lower bounds to the information rate of the Pol-Mux channel and shown simulation results for a specific channel model. The results show that with moderate interference our bounds are so close that virtually compute the exact information rate. For strong interference and modulation formats with high spectral efficiency there is still some spread between the two, leaving space to future investigations.