

Computing the information rate of discrete-time Wiener phase noise channels by parametric Bayesian tracking

LUCA BARLETTA

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, I-20133 Milan, Italy
luca.barletta@polimi.it

Abstract: A new upper bound (UB) on the information rate (IR) transferred through the additive white Gaussian noise channel affected by Wiener's laser phase noise is proposed in the paper. The bound is based on Bayesian tracking of the noisy phase. Specifically, the predictive and posterior densities involved in the tracking are expressed in parametric form, therefore tracking is made on parameters. This makes the method less computationally demanding than known non-parametric methods, e.g. methods based on phase quantization and trellis representation of phase memory. Simulation results show that the UB is so close to the lower bound that we can claim of having virtually computed the actual IR.

© 2016 Optical Society of America

OCIS codes: (060.4510) Optical communications; (060.1660) Coherent communications; (060.2330) Fiber optics communications; (200.3050) Information processing.

References and links

1. R.-J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel "Capacity limits of optical fiber networks," *J. Lightw. Technol.* **28**(7), 662–701 (2010).
2. T. Mizuochi, Y. Miyata, K. Kubo, T. Sugihara, K. Onohara, and H. Yoshida, "Progress in soft-decision FEC," *OFC/NFOEC* (2011), paper NWC2.
3. M. Magarini, A. Spalvieri, F. Vacondio, M. Bertolini, M. Pepe, and G. Gavioli, "Empirical modeling and simulation of phase noise in long-haul coherent optical systems," *Opt. Express* **19**(23), 22455–22461 (2011).
4. G. J. Foschini, G. Vannucci, "Characterizing filtered light waves corrupted by phase noise," *IEEE Trans. Inf. Theory* **34**(6), 1437–1448 (1988).
5. M. Peleg, S. Shamai (Shitz), and S. Galan, "Iterative decoding for coded noncoherent MPSK communications over phase-noisy AWGN channel," *Proc. IEE Commun.* **147**, 87–95 (2000).
6. G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Selected Areas Commun.* **23**(9), 1748–1757 (2005).
7. A. Barbieri and G. Colavolpe, "Soft-output decoding of rotationally invariant codes over channels with phase noise," *IEEE Trans. Commun.* **55**(11), 2125–2133 (2007).
8. L. Barletta, M. Magarini, and A. Spalvieri, "Staged demodulation and decoding," *Opt. Express* **20**(21), 23728–23734 (2012).
9. M. Magarini, L. Barletta, and A. Spalvieri, "Efficient computation of the feedback filter for the hybrid decision feedback equalizer in highly dispersive channels," *IEEE Trans. Wirel. Commun.* **11**(6), 2245–2253 (2012).
10. B. Goebel, R.-J. Essiambre, G. Kramer, P. J. Winzer, and N. Hanik, "Calculation of mutual information for partially coherent Gaussian channels with application to fiber optics," *IEEE Trans. Inf. Theory* **57**(9), 5720–5736 (2011).
11. P. Hou, B. J. Belzer, and T. R. Fischer, "Shaping gain of the partially coherent additive white Gaussian noise channel," *IEEE Commun. Letters* **6**(5), 175–177 (2002).
12. L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," *J. Lightw. Technol.* **30**(10), 1480–1486 (2012).
13. A. Barbieri and G. Colavolpe, "On the information rate and repeat-accumulate code design for phase noise channel," *IEEE Trans. Commun.* **59**(12), 3223–3228 (2011).
14. L. Barletta, M. Magarini, and A. Spalvieri, "Tight upper and lower bounds to the information rate of the phase noise channel," *IEEE Intern. Symposium Inf. Theory (2013)*, 2284–2288.
15. 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 Trans. Inf. Theory* **60**(7), 3834–3844 (2014).
16. J. Dauwels and H.-A. Loeliger, "Computation of information rates by particle methods," *IEEE Trans. Inf. Theory* **54**(1), 406–409 (2008).
17. D. Simon, *Optimal State Estimation*. (John Wiley & Sons, 2006).

1. Introduction

Multiplicative phase noise is a major source of impairment in radio and optical channels. Recent studies about the phase noise that arises in optical channels and about its effects in coherent optics can be found in [1–3]. The basic model of laser phase noise is a Wiener process, [3], [4], where the power spectral density (PSD) of the spectral line is a slope of -20 dB/decade. Several methods have been proposed in the literature to combat the detrimental effects of Wiener phase noise (WPN). Among these methods we cite iterative demodulation and decoding techniques of [5–7] and the insertion of pilot symbols (possibly with staged decoding [8].) Note however that single carrier systems affected by non-ideal frequency responses of the channel plus phase noise require specific mitigation techniques [9] that are not addressed in this paper.

The capacity of AWGN channels affected by phase noise with white PSD is studied in [1], [10,11]. Computation of information rates (IRs) for the multiplicative WPN plus AWGN channel, which is a channel with memory and continuous state, is a challenging problem. Monte Carlo approaches based on non-parametric Bayesian tracking (BT) techniques of the hidden phase with memory, *e.g.* phase state-space quantization [12, 13] and particle filtering [14, 15], have been recently proposed for computing IRs. In [16] the method of particle filtering is adopted to work out an approximation to the IR.

The method proposed here is based on parametric representation of densities. This leads to efficient computation, since, as shown in the paper, the parameters can be updated in the iteration steps of the BT method by closed-form equations. Conversely, in non-parametric methods, densities are estimated and updated point-wisely in their support space, leading to a more demanding computation.

2. Channel and Source Model

Let u_i^k indicate the vector $(u_i, u_{i+1}, \dots, u_k)$ with $u_i^k \in \mathcal{U}_i^k$, and let U indicate a stationary and ergodic process, $U = (U_0, U_1, \dots)$, whose generic realization is the sequence (u_0, u_1, \dots) . When \mathcal{U}_i^k is a continuous set, $p(u_i^k)$ is used to indicate the multivariate probability density function (PDF), while when \mathcal{U}_i^k is a discrete set $p(u_i^k)$ indicates the multivariate mass probability and $|\mathcal{U}_i^k|$ denotes the number of elements in \mathcal{U}_i^k .

The k -th output of the channel is

$$y_k = x_k e^{j\phi_k} + w_k, \quad \phi_k = [\phi_{k-1} + \gamma v_i] \bmod 2\pi, \quad k = 1, 2, \dots, \quad (1)$$

where j is the imaginary unit, Y is the complex channel output process, X is the channel complex input modulation process, and Φ is the phase noise process which is assumed to be independent of X and W . We assume that the input process X is made of i.i.d. random variables with zero mean and unit variance. Process W is a complex AWGN process with zero mean and variance SNR^{-1} . Process Φ is modeled as a Wiener process folded in the range $[-\pi, \pi)$ where the frequency noise V is a i.i.d. sequence of Gaussian random variables with zero mean and unit variance, γ is a scalar, and ϕ_0 is uniformly distributed in $[-\pi, \pi)$.

Equations (1) can be cast in the framework of continuous-state first-order Markov channels, where the state at time k is ϕ_k . For this class of channels

$$p(\phi_k | \phi_{k-1}, y_1^{k-1}, x_1^{k-1}) = p(\phi_k | \phi_{k-1}). \quad (2)$$

Moreover, from (1) one realizes that the channel is memoryless given the state, therefore

$$p(y_k, x_k | \phi_k, y_1^{k-1}, x_1^{k-1}) = p(y_k, x_k | \phi_k). \quad (3)$$

3. Upper Bound

Let $h(U)$ denote the relative entropy rate of process U . According to [15], an upper bound (UB) $\bar{I}(X; Y)$ on the IR $I(X; Y)$ is $\bar{I}(X; Y) = \bar{h}(Y) - h(Y|X, \Phi) - h(\Phi) + \bar{h}(\Phi|Y, X)$. The terms $h(\Phi)$ and $h(Y|X, \Phi)$ can be evaluated in a straightforward manner, see [15]. The two UBs $\bar{h}(Y)$ and $\bar{h}(\Phi|Y, X)$ can be based as in [15] on the normalized Kullback-Leibler (KL) divergence. Assuming that U is stationary and ergodic, one can invoke the Shannon-McMillan-Breiman theorem and the chain rule, getting

$$\bar{h}(U) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \left(\frac{1}{q(u_k | u_0^{k-1})} \right), \quad (4)$$

where u_0^n is generated according to the actual multivariate PDF $p(u_1^n)$, and the initial condition u_0 is given. The bound can be extended to the conditional relative entropy rate in a straightforward manner. The *auxiliary* PDF $q(\cdot)$ can be seen as an approximation to $p(\cdot)$. In this perspective, the KL divergence is a measure of the quality of the fit between $p(\cdot)$ and $q(\cdot)$, and the UB is equal to the actual relative entropy rate when the fit is ideal, that is when $q(\cdot) = p(\cdot)$.

4. Bayesian Tracking

In the first-order model of phase noise (1) the hidden state at time k is the phase ϕ_k . BT of the hidden state is based on the two-step iteration for $k = 1, 2, \dots$

$$\bar{p}(\phi_k | z_1^{k-1}) = \int_{-\pi}^{\pi} p(\phi_k | \phi_{k-1}) p(\phi_{k-1} | z_1^{k-1}) d\phi_{k-1}, \quad p(\phi_k | z_1^k) = \frac{\bar{p}(\phi_k | z_1^{k-1}) p(z_k | \phi_k)}{p(z_k | z_1^{k-1})}, \quad (5)$$

where $\bar{p}(\phi_k | z_1^{k-1})$ is the *predictive* density, $p(\phi_k | z_1^k)$ is the *posterior* density, z_1^k is the channel measurement up to time k , and the denominator of the above equation is computed by the Chapman-Kolmogoroff equation

$$p(z_k | z_1^{k-1}) = \int_{-\pi}^{\pi} \bar{p}(\phi_k | z_1^{k-1}) p(z_k | \phi_k) d\phi_k. \quad (6)$$

Note that the state transition probability $p(\phi_k | \phi_{k-1})$ appears in (5) in place of $p(\phi_k | \phi_{k-1}, z_1^{k-1})$ thanks to the Markovian property (2). Moreover, thanks to (3), the channel transition probability $p(z_k | \phi_k)$ can be used in place of $p(z_k | \phi_k, z_1^{k-1})$ in (5).

The predictive density (and, as a consequence, the posterior density as well) cannot be expressed in closed form and therefore it should be approximated in some way. We adopt parametric BT to obtain approximations to the probability density functions appearing in the relative entropy rates $h(Y)$ and $h(\Phi|Y, X)$. When $h(\Phi|Y, X)$ is evaluated, one puts $Z = (Y, X)$ and the auxiliary probability that appears inside the logarithm of (4) is evaluated as

$$q(\phi_k | y_1^k, x_1^k, \phi_{k+1}) = \frac{p(\phi_{k+1} | \phi_k) q(\phi_k | y_1^k, x_1^k)}{\int_{-\infty}^{\infty} p(\phi_{k+1} | \phi_k) q(\phi_k | y_1^k, x_1^k) d\phi_k}, \quad p(\phi_k | \phi_{k-1}) = \sum_{i=-\infty}^{\infty} g(\phi_{k-1} + 2\pi i, \gamma^2; \phi_k), \quad (7)$$

where the approximation $q(\phi_k | y_1^k, x_1^k)$ to the actual posterior probability is obtained from the parameters worked out by the parametric BT, and $g(\mu, \sigma^2; x)$ is a Gaussian density with mean μ and variance σ^2 evaluated in x . When $h(Y)$ is considered, $Z = Y$ and the approximation $q(y_k | y_1^{k-1})$ to $p(y_k | y_1^{k-1})$ is obtained from the Chapman-Kolmogoroff equation (6) and used inside the base-2 logarithm of (4).

Two parametric forms are hereafter considered: Tikhonov and Fourier. As expected, Tikhonov parametrization outperforms Fourier parametrization in approximating $p(\phi_k | y_1^k, x_1^k)$, because

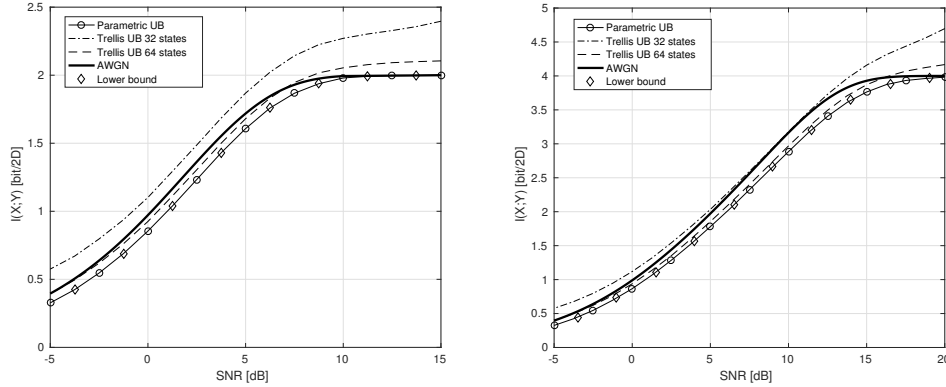


Fig. 1. Lower bound and upper bounds for 4-QAM (left) and 16-QAM (right), with $\gamma = 0.125$. The achievable information rate of the pure AWGN channel ($\gamma = 0$) is also reported.

channel and state transition probabilities are very well approximated by Tikhonov distributions of the phase ϕ_k (see (9) and (11)), hence Tikhonov parameters almost represent a sufficient statistics for computing the IRs. On the other hand, Fourier parametrization is well suited to approximate $p(\phi_k | y_1^{k-1})$, which is a periodic function of period $\pi/2$ and with a small bandwidth (hence a few Fourier coefficients completely describe the function) for most constellations of practical use. In general, using a parametrization that is far from representing a sufficient statistics, will give looser bounds on the IRs. See App. A and B for details about Tikhonov and Fourier parametrizations, respectively. As shown in the appendices, all the integrals involved in the evaluation of auxiliary PDF appearing in (4) can be computed in closed form. This makes the proposed method computationally efficient.

5. Simulation Results

In this section, the new UB is compared to a lower bound worked out by the computationally demanding trellis-based method of [12]. Specifically, the lower bound has been obtained with 128 states. Fig. 1 reports the results obtained with 4-QAM and with $\gamma = 0.125$: This is the largest value of γ obtained in the experimental results of [3] and can be regarded as a case of strong phase noise in cases of practical interest. The lower bound is virtually indistinguishable from the UB obtained with parametric BT in the entire range of signal-to-noise ratios (SNRs). A similar analysis holds for the results obtained with 16-QAM and reported in Fig. 1. Also in this case, the UB obtained with parametric BT virtually gives the actual IRs for $\gamma = 0.125$ and any SNRs. The figures report also the UB computed with the non-parametric approach (*Trellis*) for 32 and 64 states: For both constellations we observe that such UBs are not tight and are often looser than the AWGN information curve, that serves as a trivial IR UB. Increasing the number of states of the non-parametric approach helps in getting tighter bounds, at the expense of a much higher computational complexity. When the actual IR saturates (SNR = {10, 20} dB resp. for 4- and 16-QAM), the gap between the 64-state UB and the actual IR is larger for 16-QAM. However, if the gaps are normalized to the maximum IR, they are the same for the two modulation formats. Moreover, observe that getting a good IR estimate at high SNR is much more difficult with a non-parametric approach rather than a parametric one, because in this regime estimating accurately the density tail is crucial [15]: At high SNR the densities become spiky and tails are better captured by a parametric description, with a negligible computational complexity.

6. Conclusion

A new upper bound (UB) above the information rate transferred through the discrete-time WPN channel has been presented. The results, compared to the lower bound obtained with the computationally demanding non-parametric methods of [12], show that the UB is accurate in all cases of practical interest. Before concluding the paper, a remark is in order about phase noise of order higher than one, which is out of the scope of the present paper and that will be subject of future research. Extension of the parametric-based approach to phase noise with memory of order higher than one, as the one studied in [15], is feasible by proposing an appropriate parametric density. In contrast, quantizing a multidimensional state space according to the approach of [12] would lead to an exponential increase of the number of states of the trellis, making computation unfeasible. Also, non-parametric methods based on particle filtering [14, 15] can suffer from the high space dimensionality, especially at high SNR. To get accurate UBs one is led to increase the number of particles, thus increasing the complexity of the computation.

A. Tikhonov Parametrization

The method is based on the approximation of the posterior density at time $k - 1$ as a Tikhonov density of real variable $\phi_{k-1} \in [-\pi, \pi)$ and complex parameter ω_{k-1}

$$p(\phi_{k-1}|y_1^{k-1}, x_1^{k-1}) \approx t(\omega_{k-1}; \phi_{k-1}) = \frac{e^{\Re\{\omega_{k-1}e^{-j\phi_{k-1}}\}}}{2\pi I_0(|\omega_{k-1}|)}, \quad (8)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind and $\Re\{\cdot\}$ denotes the real part. The state transition probability is hereafter approximated as

$$p(\phi_k|\phi_{k-1}) \approx g(\phi_{k-1}, \gamma^2; \phi_k). \quad (9)$$

Using these approximations in place of the actual probability density functions in (5), and using the properties of Tikhonov densities listed in [18] one has

$$\bar{p}(\phi_k|y_1^{k-1}, x_1^{k-1}) \approx \int_{-\infty}^{\infty} g(\phi_{k-1}, \gamma^2; \phi_k)t(\omega_{k-1}; \phi_{k-1})d\phi_{k-1} \approx t(\bar{\omega}_k; \phi_k) \quad (10)$$

where $\bar{\omega}_k = \omega_{k-1}/(1+\gamma^2|\omega_{k-1}|)$. The channel transition probability at time k can be approximated with a Tikhonov density of variable ϕ_k :

$$p(y_k|x_k, \phi_k) \approx \frac{2\text{SNR} \cdot I_0(2\text{SNR}|y_k x_k^*|)}{\exp(\text{SNR}(|y_k|^2 + |x_k|^2))} t(2\text{SNR}y_k x_k^*; \phi_k). \quad (11)$$

Using (10) and (11) into (5) gives

$$p(\phi_k|y_1^k, x_1^k) \approx t(2\text{SNR}y_k x_k^* + \bar{\omega}_k; \phi_k) = t(\omega_k; \phi_k), \quad \omega_k = 2\text{SNR}y_k x_k^* + \bar{\omega}_k. \quad (12)$$

Having computed ω_k by parametric BT (10) and (12), one can get the contribution at time k to the entropy rate $h(\Phi|Y, X)$ by taking the logarithm of the auxiliary probability (7). Note that in the BT we assume that the posterior and the predictive densities are Tikhonov, thus getting a close approximation to the actual densities without the need of evaluating integrals numerically. However, in the evaluation of (7), the integral appearing in the denominator cannot be evaluated in closed form if we still adopt the Tikhonov parametrization. Therefore, we adopt a Gaussian density based on the complex parameter ω_k obtained by BT as an approximation to the actual posterior density appearing in the right side of (7). Specifically, we put in (7)

$q(\phi_k | y_1^k, x_1^k) = g(\angle\omega_k, |\omega_k|^{-1}; \phi_k)$. This allows to compute in closed form the integral appearing in the denominator of (7):

$$\begin{aligned} \int_{-\infty}^{\infty} p(\phi_{k+1} | \phi_k) q(\phi_k | y_1^k, x_1^k) d\phi_k &= \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} g(\phi_k + 2n\pi, \gamma^2; \phi_{k+1}) g(\angle\omega_k, |\omega_k|^{-1}; \phi_k) d\phi_k \\ &= \sum_{n=-\infty}^{\infty} g(2n\pi + \angle\omega_k, \gamma^2 + |\omega_k|^{-1}; \phi_{k+1}), \end{aligned} \quad (13)$$

where the last equality follows because the integrals appearing inside the sum are convolutions between Gaussian densities.

B. Fourier Parametrization

We assume that constellation \mathcal{X} has a 4-fold symmetry, therefore at time $k-1$ it is possible to approximate the posterior density $p(\phi_{k-1} | y_1^{k-1})$ in the interval $[-\pi, \pi)$ by the Fourier series

$$q(\phi_{k-1} | y_1^{k-1}) = \sum_{n=-m}^m A_n^{(k-1)} e^{j4n\phi_{k-1}}, \quad (14)$$

where $\{A_n^{(k-1)}\}_{n=-m}^m$ are the Fourier coefficients of the auxiliary posterior density at time $k-1$, and m is responsible of the accuracy of the approximation. Using (14) and (9) into (5) one has

$$\bar{q}(\phi_k | y_1^{k-1}) = \int_{-\pi}^{\pi} \sum_{n=-m}^m A_n^{(k-1)} e^{j4n\phi_{k-1}} g(\phi_{k-1}, \gamma^2; \phi_k) d\phi_{k-1} = \sum_{n=-m}^m \bar{A}_n^{(k)} e^{j4n\phi_k}, \quad (15)$$

which provides us with the prediction step $\bar{A}_n^{(k)} = A_n^{(k-1)} e^{-8n^2\gamma^2}$. Using the identity $\exp(x \cos \theta) = \sum_{l=-\infty}^{\infty} I_l(x) \exp(jl\theta)$, one approximates the channel transition probability $p(y_k | \phi_k)$ by the Fourier series

$$q(y_k | \phi_k) = \sum_{n=-\infty}^{\infty} B_n^{(k)} e^{j4n\phi_k}, \quad B_n^{(k)} = \frac{1}{2\pi|\mathcal{X}|} \sum_{x_k \in \mathcal{X}} \frac{2\text{SNR} \cdot I_{|4n|}(2\text{SNR}|y_k x_k^*|)}{\exp(\text{SNR}(|y_k|^2 + |x_k|^2))} e^{-j4n\angle(y_k x_k^*)}. \quad (16)$$

The auxiliary posterior density is found by plugging (15) and (16) into (5) to obtain

$$q(\phi_k | y_1^k) = \sum_{n=-m}^m A_n^{(k)} e^{j4n\phi_k}. \quad (17)$$

Fourier coefficients $A_n^{(k)}$ can be obtained from the convolution $A_n^{(k)} \propto \tilde{A}_n^{(k)} = \sum_{i=-m}^m \bar{A}_i^{(k)} B_{n-i}^{(k)}$ by suitably normalizing the \tilde{A}_n 's such that (17) is a density. Note that the truncation for $n = -m, \dots, m$ of the Fourier expansion in the right hand side of (17) might be such that $q(\phi_k | y_1^k)$ is negative for some values of ϕ_k . It is easy to show that a sufficient condition for (17) to be non-negative is

$$\tilde{A}_0^{(k)} \geq 2 \sum_{i=1}^m |\tilde{A}_i^{(k)}|, \quad (18)$$

that for $m = 1$ is also a necessary condition. At each iteration of the BT we check (18): if it is satisfied then the tracking algorithm can proceed by the normalization $A_n^{(k)} = \tilde{A}_n^{(k)} / (2\pi \tilde{A}_0^{(k)})$, $n = -m, \dots, m$, otherwise we put $A_0^{(k)} = (2\pi)^{-1}$ and $A_n^{(k)} = \tilde{A}_n^{(k)} / (4\pi \sum_{i=1}^m |\tilde{A}_i^{(k)}|)$, $n \neq 0$. In any case, the contribution at time k to the relative entropy rate $h(Y)$ is obtained by putting inside the base-2 logarithm appearing in (4) the auxiliary probability

$$q(y_k | y_1^{k-1}) = \int_{-\pi}^{\pi} \bar{q}(\phi_k | y_1^{k-1}) q(y_k | \phi_k) d\phi_k = 2\pi \tilde{A}_0^{(k)}.$$