Tight Upper and Lower Bounds to the Information Rate of the Phase Noise Channel

Luca Barletta
Institute for Advanced Study
Technische Universität München
D-85748 Garching, Germany
Email: Luca.Barletta@tum.de

Maurizio Magarini, Arnaldo Spalvieri
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano
I-20133 Milan, Italy
Email: {Maurizio.Magarini,Arnaldo.Spalvieri}@polimi.it

Abstract—**Numerical upper and lower bounds to the information rate transferred through the additive white Gaussian noise channel affected by discrete-time multiplicative autoregressive moving-average (ARMA) phase noise are proposed in the paper. The state space of the ARMA model being multidimensional, the problem cannot be approached by the conventional trellis-based methods that assume a first-order model for phase noise and quantization of the phase space, because the number of state of the trellis would be enormous. The proposed lower and upper bounds are based on particle filtering and Kalman filtering. Simulation results show that the upper and lower bounds are so close to each other that we can claim of having numerically computed the actual information rate of the multiplicative ARMA phase noise channel, at least in the cases studied in the paper. Moreover, the lower bound, which is virtually capacity-achieving, is obtained by demodulation of the incoming signal based on a Kalman filter aided by past data. Thus we can claim of having found the virtually optimal demodulator for the multiplicative phase noise channel, at least for the cases considered in the paper.**

I. INTRODUCTION

Multiplicative phase noise is a major source of impairment in radio and optical channels. The presence of phase noise in radio channels is well known and studied from a long time, being phase noise introduced by the local oscillators used in up conversion and down conversion, while multiplicative phase noise is recently becoming a hot topic in the context of coherent optical transmission. Recent studies about the phase noise that arises in optical channels and about its effects in coherent optics can be found in [1], [2]. Several methods have been proposed in the literature to combat the detrimental effects of phase noise. Among these methods we cite iterative demodulation and decoding techniques of [3]–[5] and the insertion of pilot symbols [6], and staged demodulation and decoding [7].

The capacity of the additive white Gaussian noise (AWGN) channel affected by multiplicative phase noise with white power spectral density is studied in [1], [8], [9], while Wiener's phase noise is considered in [10]–[13]. Analytical bounds on capacity of phase noise channels at high signal-to-noise ratio are given in [14]. Despite the quantity and quality of the literature available, we find room for new results by considering the channel impaired by autoregressive moving-average (ARMA) multiplicative phase noise, a phase noise model that is much more realistic than Wiener's phase noise and/or white phase noise in many cases of practical interest.

The ARMA model makes it possible to shape the power spectral density of phase noise by acting on the order and on the parameters of the model. Working out the capacity of a channel affected by a general multiplicative ARMA phase noise process is a challenging problem, because

- the state space is not finite and it is multidimensional, therefore it cannot be approached by techniques like those used for white and Wiener phase noise,
- the observation is a nonlinear function of the state.

The only paper studying the capacity of the channel affected by ARMA phase noise we are aware of is [10], where the method of particle filtering (see [15] for a tutorial on particle filtering) is adopted to work out an approximation to the constrained channel capacity, the constrained capacity being the information rate transferred through the channel with a fixed source. The new results presented in this paper are tight numerical upper and lower bounds to the constrained capacity of the AWGN ARMA phase noise channel.

II. FIRST-ORDER MARKOV CHANNELS WITH CONTINUOUS STATE

Let u_i^k indicate the column vector $(u_k, u_{k-1}, \ldots, u_i)^T$, $i \leq k$, where u_i^k is empty for $i > k$, the superscript ^{T} denotes transposition, and $u_i^k \in \mathcal{U}_i^k$. Also, let U indicate a possibly non-stationary process, $U = (U_0, U_1, \cdots)$, whose generic realization is the sequence $(u_0, u_1, \cdots)$. When $\mathcal{U}_i^k$ is a continuous set, $p(u_i^k)$ is used to indicate the multivariate probability density function, while when $\mathcal{U}_i^k$ is a discrete set $p(u_i^k)$ indicates the multivariate mass probability and $|\mathcal{U}_i|$ denotes the number of elements in $\mathcal{U}_i$.

Consider a first-order Markov channel. The Markovian state process S is characterized by the joint probability

$$p(s_0^n) = p(s_0) \prod_{k=1}^{n} p(s_k | s_{k-1}). \qquad (1)$$

A channel without feedback that is memoryless given the state is characterized by the state transition probability $p(s_k | s_{k-1})$ and by the conditional distribution

$$p(y_1^n | x_1^n, s_1^n) = \prod_{k=1}^{n} p(y_k | x_k, s_k), \qquad (2)$$

where Y is the channel output process and X is the channel input process, that we assume to be discrete. Equation (2) says

that the channel output process is memoryless given the source and the state. Drawing from the parlance of carrier recovery, the channel transition probability $p(y_k|x_k, s_k)$, which is conditioned on channel's input, is hereafter called *data-aided* channel transition probability. We assume that the source is memoryless and independent of the state, that is

$$p(x_1^n|s_1^n) = \prod_{k=1}^{n} p(x_k). \tag{3}$$

Putting together (2) and (3) one finds that the joint source and channel model is memoryless given the state:

$$p(y_1^n, x_1^n|s_1^n) = \prod_{k=1}^{n} p(y_k, x_k|s_k). \tag{4}$$

Using (4) one finds that channel's output is memoryless given the state:

$$p(y_1^n|s_1^n) = \sum_{x_1^n \in \mathcal{X}^n} p(y_1^n, x_1^n|s_1^n) = \sum_{x_1^n \in \mathcal{X}_1^n} \prod_{k=1}^{n} p(y_k, x_k|s_k)$$
$$= \prod_{k=1}^{n} \sum_{x_k \in \mathcal{X}_k} p(y_k, x_k|s_k) = \prod_{k=1}^{n} p(y_k|s_k). \tag{5}$$

Drawing again from the parlance of carrier recovery, the channel transition probability $p(y_k|s_k)$, which is not aware of channel's input, is hereafter called *blind* channel transition probability. From eq. (1) and (5), after straightforward passages one gets

$$p(s_k|s_{k-1}, y_1^{k-1}) = p(s_k|s_{k-1}), \quad k = 1, 2, \cdots, n. \tag{6}$$

Also, by (4) and (5) one finds that the source is memoryless given the state and channel's output:

$$p(x_1^n|y_1^n, s_1^n) = \prod_{k=1}^{n} p(x_k|y_k, s_k). \tag{7}$$

III. BAYESIAN TRACKING

Any measurement process Y that is memoryless given the state can be cast in the general framework of state-space approach for modelling dynamic systems, which is defined by the state transition equation

$$s_k = f_k(s_{k-1}, v_{k-1}), \quad k = 1, 2, \cdots, n, \tag{8}$$

and by the measurement equation

$$y_k = h_k(s_k, n_k), \quad k = 1, 2, \cdots, n, \tag{9}$$

where $f_k(\cdot)$ and $h_k(\cdot)$ are possibly non-linear and time-varying known functions of their arguments, v_k is the process noise vector, and n_k is the measurement noise vector, which is assumed to be independent of v_k. The state-space approach fits the Markov channel, taking the output channel process Y as the measurement process both in the blind and in the data aided case. In the blind case, the measurement equation is a time-invariant function of the state, and the measurement noise is the joint effect of channel noise and input process. The blind case is described by the memoryless probability

$p(y_k|s_k)$ appearing in the product (5). In the data-aided case the measurement noise is only the channel noise and the input process is embedded in the known non-linear and time-varying $h_k(\cdot)$. In this case the measurement probability is $p(y_k|x_k, s_k)$.

A powerful tool in the analysis of dynamical system is the so-called *Bayesian tracking*. Let the Markovian state be continuous. One can track the hidden state by a two-step recursion that, for $k = 1, 2, \cdots, n$, reads

$$p(s_k|y_1^{k-1}) = \int_{\mathcal{S}} p(s_k|s_{k-1})p(s_{k-1}|y_1^{k-1})ds_{k-1}, \tag{10}$$

$$p(s_k|y_1^k) = \frac{p(s_k|y_1^{k-1})p(y_k|s_k)}{p(y_k|y_1^{k-1})}, \tag{11}$$

where $p(s_k|y_1^{k-1})$ is the *predictive* distribution, $p(s_k|y_1^k)$ is the *posterior* distribution, and the denominator of (11) is a normalization factor such that the left-hand side is a probability. The normalization factor can be computed by the Chapman-Kolmogoroff equation

$$p(y_k|y_1^{k-1}) = \int_{\mathcal{S}_k} p(s_k|y_1^{k-1})p(y_k|s_k)ds_k. \tag{12}$$

The state transition probability $p(s_k|s_{k-1})$ appears in (10) in place of $p(s_k|s_{k-1}, y_1^{k-1})$ thanks to (6). Thanks to (5), $p(y_k|s_k)$ can be used in place of $p(y_k|s_k, y_1^{k-1})$ in (11).

When the dynamic system is a linear system with Gaussian noises, Bayesian tracking is performed by the Kalman filter. When the model is not tractable, one can resort to particle filtering techniques to work out an approximation to the wanted distribution.

The probabilities worked out by Bayesian tracking can be used to evaluate entropy rates by Monte Carlo integration as, for instance, in [10], [13]. When the result of Bayesian tracking is an approximation $q(u_1^k)$ to the wanted probability $p(u_1^k)$, then, by the Kullback-Leibler inequality, the approximation can be used to get an upper bound on the wanted entropy rate

$$\overline{h}(U) = -\lim_{k \to \infty} \frac{1}{k} E_p \left\{ \log_2 q(u_1^k) \right\} \geq h(U), \tag{13}$$

where operator E_p denotes expectation with respect to probability $p(\cdot)$.

IV. THE ARMA PHASE NOISE CHANNEL

The k-th output of the channel is

$$y_k = x_k e^{j\phi_k} + w_k, \quad k = 1, 2, \cdots, n \tag{14}$$

where j is the imaginary unit, Y is the complex channel output process, X is the channel complex input modulation process made by i.i.d. random variables with zero mean and unit variance, W is the complex AWGN process with zero mean and variance SNR^{-1}, and Φ is the phase noise process which is assumed to be independent of X and W. Specifically, process Φ is modelled as the 1-causal accumulation modulo 2π of frequency noise, that is

$$\phi_k = [\lambda_{k-1} + \phi_{k-1}]_{\bmod 2\pi}, \tag{15}$$

where the frequency noise process Λ is given by the z-transform

$$\sum_{k=-\infty}^{\infty} \lambda_k z^{-k} = H(z) \cdot \left(\sum_{k=-\infty}^{\infty} v_k z^{-k} \right)$$

where V is a white Gaussian noise process with zero mean and variance γ^2, and

$$H(z) = \frac{\prod_{k=1}^{N}(1 - \beta_k z^{-1})}{\prod_{k=1}^{N}(1 - \alpha_k z^{-1})} = \frac{1 + \sum_{k=1}^{N} b_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}}, \quad (16)$$

where $|\alpha_k| < 1$, $|\beta_k| \leq 1$, $N \geq 0$, and it is understood that $H(z) = 1$ for $N = 0$, leading to the special case of random phase walk, where $\lambda_k = v_k$. $H(z)$ is the transfer function of a filter made by a shift register with feedback taps a_1^N and forward taps b_1^N. Let ω_{k-N}^{k-1} be the content of the shift register at the k-th channel use, that is

$$\sum_{k=-\infty}^{\infty} \omega_k z^{-k} = \frac{\sum_{k=-\infty}^{\infty} v_k z^{-k}}{1 - \sum_{k=1}^{N} a_k z^{-k}}.$$

The state at time k is the $(N+1)$ column vector

$$s_k = (\phi_k, (\omega_{k-N}^{k-1})^T)^T. \quad (17)$$

Let us introduce the state transition matrix

$$F = \begin{bmatrix} 1 & (a_1^N + b_1^N)^T \\ 0 & (a_1^N)^T \\ 0_{N-1} & I_{N-1} \quad 0_{N-1} \end{bmatrix},$$

where I_N is the identity matrix of size $N \times N$ and 0_N is a column vector of N zeros. The state transition equation is

$$s_{k+1} = F s_k + (v_k, v_k, 0_{N-1}^T)^T + (2m\pi, 0_N^T)^T,$$

where m is such that ϕ_{k+1} lies in the interval $[0, 2\pi)$, thus making the state transition equation non-linear.

Given s_k, for $N = 0$ the state transition to s_{k+1} is ambiguous of $2n\pi$, while for $N \geq 1$, due to the presence of ω_k in s_{k+1}, the state transition is not ambiguous. Although not necessary, in the following we will assume $N \geq 1$, referring the reader to [13] for the state transition probability with $N = 0$. For $N \geq 1$ the state transition probability is a $(N+1)$-dimensional Gaussian distribution. Note that, given s_k, N of the $(N+1)$ entries of s_{k+1} are known, the only free random variable being v_k, hence the covariance matrix of the state transition probability has unit rank. Specifically,

$$p(s_{k+1}|s_k) = g_{N+1}(F s_k + (2m\pi, 0_N^T)^T, \Sigma_v; s_{k+1}), \quad (18)$$

where $g_N(\mu, \Sigma; x)$ is a N-dimensional Gaussian distribution over the space spanned by x with mean vector μ and covariance matrix Σ,

$$\Sigma_v = \begin{bmatrix} \gamma^2 & \gamma^2 & 0_{N-1}^T \\ \gamma^2 & \gamma^2 & 0_{N-1}^T \\ 0_{N-1} & 0_{N-1} & 0_{(N-1)\times(N-1)} \end{bmatrix}, \quad (19)$$

where $0_{N \times M}$ is an all-zero $N \times M$ matrix, and

$$2m\pi = \phi_{k+1} - \phi_k - \omega_k - \sum_{i=1}^{N} b_i \omega_{k-i}. \quad (20)$$

The measurement at time k is the y_k given by (14). The data-aided channel transition probability is

$$p(y_k|x_k, s_k) = g_c(x_k e^{j\phi_k}, \text{SNR}^{-1}; y_k), \quad (21)$$

where $g_c(\mu, \sigma^2; t)$ indicates a circular symmetric Gaussian probability density function over the complex plane spanned by t with mean μ and two-dimensional variance σ^2. The joint source and channel probability is

$$p(y_k, x_k|s_k) = p(x_k) g_c(x_k e^{j\phi_k}, \text{SNR}^{-1}; y_k). \quad (22)$$

From the above probability one can compute the blind channel transition probability by (5).

V. UPPER BOUND

Let $h(U)$ denote the entropy rate of process U. Extract $h(Y|X)$ from

$$h(Y|X) - h(Y|X, S) = h(S|X) - h(S|X, Y),$$

to write

$$I(X; Y) = h(Y) - h(Y|X, S) - h(S) + h(S|X, Y), \quad (23)$$

where, by independence between X and the state process S, $h(S)$ has been substituted in place of $h(S|X)$. The upper bound that we propose is

$$\overline{h}(Y) - h(Y|X, S) - h(S) + \overline{h}(S|X, Y) \geq I(X; Y), \quad (24)$$

where $\overline{A}$ indicates an upper bound on A. The two relative entropy rates $h(Y|X, S)$ and $h(S)$ are those of the white Gaussian processes W and V, respectively. The upper bound $\overline{h}(Y) \geq h(Y)$ can be obtained by approximating the conditional probability $p(y_k|y_1^{k-1})$ to the normalization factor of blind Bayesian tracking performed by a particle filter as in [10].

The new contribution of the present paper is the upper bound $\overline{h}(S|X, Y)$, which is worked out as follows. Invoking the chain rule, the Markovian property (6), and the Shannon-McMillan-Breiman theorem, one can evaluate the entropy rate by computer simulation as

$$h(S|X, Y) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} -\log_2 p(s_k|x_1^k, y_1^k, s_{k+1}), \quad (25)$$

where $(x_1^n, y_1^n, s_1^{n+1})$ is a realization of the joint process (X, Y, S). Unfortunately, the actual $p(s_k|x_1^k, y_1^k, s_{k+1})$ of (25) is not tractable. We propose to approximate it as

$$q(s_k|x_1^k, y_1^k, s_{k+1}) = \frac{p(s_{k+1}|s_k) q(s_k|x_1^k, y_1^k)}{\int_{\mathcal{S}} p(s_{k+1}|s_k) q(s_k|x_1^k, y_1^k) ds_k}, \quad (26)$$

with

$$q(s_k|x_1^k, y_1^k) = \sum_{l=-\infty}^{\infty} g_{N+1}(\mu_k, \Sigma_k; \phi_k + 2l\pi, \omega_{k-N}^{k-1}), \quad (27)$$

thus, thanks to (13), getting the upper bound

$$\overline{h}(S|X,Y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} -\log_2 q(s_k|x_1^k, y_1^k, s_{k+1}).$$

The denominator of (26) can be treated by moving the sum (27) outside the integral, and observing that the integral is the convolution between two Gaussian distributions, leading to closed form computation as in the predictive step of the Kalman filter [16, Sec. 3.3]:

$$\int_{\mathcal{S}} p(s_{k+1}|s_k) q(s_k|x_1^k, y_1^k) \, ds_k =$$
$$\sum_{l=-\infty}^{\infty} g_{N+1}(F\mu_k, F\Sigma_k F^T + \Sigma_v; \phi_{k+1} + 2l\pi, \omega_{k-N+1}^k). \quad (28)$$

The parameters μ_k and Σ_k appearing in equations (27) and (28) can be worked out by a linearized Kalman filter [16, Sec. 13.2]. As it will be shown by simulation results, a tighter bound can be obtained by taking for μ_k and Σ_k a sample estimate where the sample is the set of posterior particles of a particle filter. Note that the integral in the denominator of (26) is a normalization factor such that the left side of (26) is a probability. As a consequence, it cannot be evaluated by the predictive particles of the particle filter, because the predictive particles would provide only an approximation to the wanted integral, and using an approximation to the denominator is not sufficient to guarantee that the ratio in (26) is a probability. Also, it is worth pointing out that, while in [10] the phase in the state model is unwrapped, here it is the evaluation of $\overline{h}(S|X,Y)$, that is not made in [10], that forces us to define the state by the wrapped phase (15). As a matter of fact, phase ambiguities of $2n\pi$ are inherently present in the measurement, therefore cycle slips of the Bayesian tracking algorithm would lead to catastrophic errors of $2n\pi$ between the actual unwrapped phase and the distribution of the unwrapped phase recovered by the tracking algorithm.

VI. LOWER BOUND

Assume a discrete input alphabet. The lower bound that we propose is $H(X) - \overline{H}(X|Y) \leq I(X,Y)$, where, by the same arguments leading to (25) and by the Kullback-Leibler inequality (13), one evaluates the upper bound on the conditional entropy rate as

$$\overline{H}(X|Y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} -\log_2 q(x_k|x_1^{k-1}, y_1^n). \quad (29)$$

The upper bound can be based on demodulation, that is on the probability

$$p(x_k|x_1^{k-1}, y_1^n) = \int_{\mathcal{S}} p(s_k, x_k|x_1^{k-1}, y_1^n) ds_k, \quad (30)$$

where the probability inside the integral can be written as

$$p(s_k, x_k|x_1^{k-1}, y_1^n) = p(s_k|x_1^{k-1}, y_1^n) p(x_k|s_k, x_1^{k-1}, y_1^n)$$
$$= p(s_k|x_1^{k-1}, y_1^n) p(x_k|s_k, y_k), \quad (31)$$

where the second equality comes from (7). In what follows the first factor in (31) is approximated to $p(s_k|x_1^{k-1}, y_1^k)$. We point out that the proposed approximation is likely to be tight, because the condition y_{k+1}^n gives only a weak contribution of non-data-aided type to the wanted probability. The proposed approximation leads to

$$q(x_k|x_1^{k-1}, y_1^n) = \int_{\mathcal{S}} q(s_k|x_1^{k-1}, y_1^k) p(x_k|s_k, y_k) ds_k$$
$$= \int_{\mathcal{S}} \frac{q(s_k|x_1^{k-1}, y_1^{k-1}) p(y_k|s_k)}{p(y_k|x_1^{k-1}, y_1^{k-1})} \frac{p(y_k, x_k|s_k)}{p(y_k|s_k)} ds_k$$
$$\propto \int_{\mathcal{S}} q(s_k|x_1^{k-1}, y_1^{k-1}) p(y_k, x_k|s_k) ds_k, \quad (32)$$

which, after normalization, can be used in (29) to get the desired bound. The first factor inside the integral (32) is the predictive probability of Bayesian tracking, while the second factor is a memoryless term that comes from the channel model (22).

VII. SIMULATION RESULTS

The frequency noise used in the simulations is obtained by filtering white Gaussian noise through the transfer function

$$H(z) = \frac{(1 - \beta_1 z^{-1})(1 - \beta_2 z^{-2})}{1 - \alpha_1 z^{-1}}. \quad (33)$$

Special cases of (33) are obtained with $\beta_1 = \alpha_1$ and $\beta_2 = 1$, leading to white phase noise, and $\beta_1 = 0$, $\beta_2 = 1$, $\alpha_1 = 1$, that leads to Wiener's phase noise. Model (33) is proposed in [17] as an approximation to the phase noise spectrum of real-world microwave local oscillators and it has been used with $\alpha_1 = 0.9999$, $\beta_1 = 0.9937$, $\beta_2 = 0.7286$ to get the simulation results that are hereafter presented. The lower bound is computed by adopting as a Bayesian tracking method the linearized predictive Kalman filter, as in [18] and [19], while for the upper bound we use both the Kalman filter and the particle filter. Figure 1 reports the results for 4-ary quadrature-amplitude modulation (QAM) while Fig. 2 reports the results for 16-QAM, in both cases with two values of γ. The two Figures show that the particle filter greatly improves the upper bound over the Kalman filter, especially for large γ. In contrast, the lower bound based on the predictive Kalman filter is so tight that there is no need of using a particle filter for demodulation, also for large values of γ. We have observed that the Kalman filter often produces a covariance Σ_k with a determinant that is much lower than the one that is obtained by the particle filter. What happens is that the folded Gaussian distribution (27) is sampled in the state visited by the simulation, and, when this state is far from the mean vector, the Gaussian is sampled on the tails. In this event, the poor estimation of the covariance leads to dramatically large errors in the evaluation of the differential entropy rate $h(S|X,Y)$. Conversely, the entropy rate $\overline{H}(X|Y)$ that appears in the lower bound is based on the integral of the mentioned Gaussian distribution, hence it is less sensitive to errors in the estimated covariance.
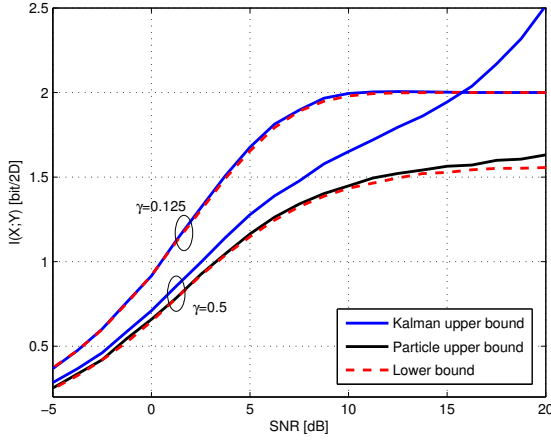
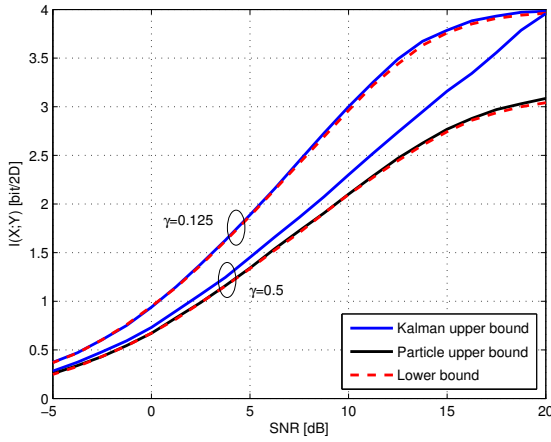Fig. 1: Upper and lower bounds to the information rate of 4-QAM.

Fig. 2: Upper and lower bounds to the information rate of 16-QAM.

VIII. Conclusion

We have presented upper and lower bounds to the constrained information rate transferred through the multiplicative phase noise channel with ARMA phase noise. From the results it appears that the upper and lower bounds are so close to each other that we can claim of having computed the actual information rate, at least for the second-order ARMA phase noise studied in the simulation. An important experimental result presented in the paper is that demodulation based on a predictive linearized Kalman filter aided by past data is virtually capacity achieving, at least in the examples studied in the paper. This is not surprising in view of the result obtained in [20] for the intersymbol interference (ISI) channel, that says that predictive filtering aided by past data (in the case of the ISI channel, the predictive decision-feedback equalizer) virtually leads to channel capacity. A practical mean to replace past data with the decisions coming from a capacity achieving code is the interleaving scheme originally proposed by Eyuboglu in [21] for the ISI channel. Extension of this principle to other channels can be found, for instance, in [22]. Computational complexity of demodulation via Kalman filter can be lowered by using a time invariant filter as described in [23].

References

[1] 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. 15, 2010.

[2] 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," *Optics Express*, vol. 19, no. 23, pp. 22455-22461, Nov. 7, 2011.

[3] M. Peleg, S. Shamai (Shitz), and S. Galan, "Iterative decoding for coded noncoherent MPSK communications over phase-noisy AWGN channel," *Proc. IEE Commun.*, vol. 147, pp. 87-95, Apr. 2000.

[4] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE Journal on Selected Areas in Commun.*, vol. 23, no. 9, pp. 1748-1757, Sept. 2005.

[5] A. Barbieri and G. Colavolpe, "Soft-output decoding of rotationally invariant invariant codes over channels with phase noise," *IEEE Trans. Commun.*, vol. 55, no. 11, pp. 2125-2133, Nov. 2007.

[6] A. Spalvieri and L. Barletta, "Pilot-aided carrier recovery in the presence of phase noise," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 1966-1974, July 2011.

[7] L. Barletta, M. Magarini, and A. Spalvieri, "Staged demodulation and decoding," *Optics Express*, vol. 20, no. 21, pp. 23728–23734, Oct. 8, 2012.

[8] 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*, vol. 57, no. 9, pp. 5720-5736, Sept. 2011.

[9] P. Hou, B. J. Belzer, and T. R. Fischer, "Shaping gain of the partially coherent additive white Gaussian noise channel," *IEEE Commun. Letters*, vol. 6, no. 5, pp. 175-177, May 2002.

[10] J. Dauwels and H.-A. Loeliger, "Computation of information rates by particle methods," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 406-409, Jan. 2008.

[11] L. Barletta, M. Magarini, and A. Spalvieri, "Estimate of information rates of discrete-time first-order Markov phase noise channels," *IEEE Photon. Technol. Lett.*, vol. 23, no. 21, pp 1582–1584, Nov. 1, 2011.

[12] A. Barbieri and G. Colavolpe, "On the information rate and repeat-accumulate code design for phase noise channels," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3223-3228, Dec. 2011.

[13] L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," *IEEE J. Lightw. Technol.*, vol. 30, no. 10, pp. 1480-1486, May 15, 2012.

[14] A. Lapidoth, "On phase noise channels at high SNR," *Inf. Theory Workshop,* Oct. 2002.

[15] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. on Signal Proc.,* vol. 50, no. 2, pp. 174-188, Feb. 2002.

[16] D. Simon, *Optimal State Estimation.* New York: Wiley, 2006.

[17] A. Spalvieri and M. Magarini, "Wiener's analysis of the discrete-time phase-locked loop with loop delay," *IEEE Trans. Circuits and Systems II: Express Briefs,* vol. 55, no. 6, pp.596-600, June 2008.

[18] 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," *Optics Express*, vol. 20, no. 23, pp. 25471-25477, Nov. 5, 2012.

[19] L. Barletta, M. Magarini, and A. Spalvieri, "New lower bound below the information rate of phase noise channel based on Kalman carrier recovery," *International Journal on Electrical Engineering and Informatics*, vol. 4, no. 4, pp. 597-607, Dec. 2012.

[20] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and G. D. Forney, "MMSE decision-feedback equalizers and coding—part I: equalization results," *IEEE Trans. Commun.,* vol. 43, no. 10, pp 2582-2594, Oct. 1995.

[21] M. V. Eyuboglu, "Detection of coded modulation signals on linear severely distorted channels using decision-feedback noise prediction and interleaving," *IEEE Trans. Commun.,* vol. 36, no. 4, pp. 401-409, Apr. 1988.

[22] T. Li and O. M. Collins, "A successive decoding strategy for channels with memory," *IEEE Trans. Inf. Theory,* vol. 53, no. 2, pp. 628–646, Feb. 2007.

[23] L. Barletta, M. Magarini, and A. Spalvieri, "Bridging the gap between Kalman filter and Wiener filter in carrier phase tracking," *IEEE Photon. Technol. Lett.*, vol. 25, no. 11, pp 1035–1038, Jun. 1, 2013.