

# Upper and Lower Bounds to the Information Rate Transferred Through First-Order Markov Channels With Free-Running Continuous State

Luca Barletta, *Member, IEEE*, Maurizio Magarini, *Member, IEEE*, Simone Pecorino, and Arnaldo Spalvieri

**Abstract**—Starting from the definition of mutual information, one promptly realizes that the probabilities inferred by Bayesian tracking can be used to compute the Shannon information between the state and the measurement of a dynamic system. In the Gaussian and linear case, the information rate can be evaluated from the probabilities computed by the Kalman filter. When the probability distributions inferred by Bayesian tracking are nontractable, one is forced to resort to approximated inference, which gives only an approximation to the wanted probabilities. We propose upper and lower bounds to the information rate between the hidden state and the measurement based on approximated inference. Application of these bounds to multiplicative communication channels is discussed, and experimental results for the discrete-time phase noise channel and for the Gauss–Markov fading channel are presented.

**Index Terms**—Mutual information, Bayesian tracking, Kalman filtering, particle filtering, multiplicative channels, coherent communication, phase noise, Gauss–Markov fading channel, channel capacity.

## I. INTRODUCTION

**T**RACKING the state of a dynamic system from noisy measurements is a classical problem in several fields of science. In the Bayesian approach, probabilities are used to model the state evolution and the measurement given the state, and, from the model and the measurements, inference is made on the hidden evolving state. By making inference one builds the probability of the state given all the available measurements, thus embodying all the available statistical information in the inferred distribution. Therefore it can be said that, in some sense, *Bayesian tracking* extracts the information about the state that is brought by the measurements.

The most popular tool for Bayesian tracking of a system with continuous state is the Kalman filter (see, e.g., [1] for a comprehensive book on the Kalman filter). The Kalman filter performs optimal tracking, thus leading to exact inference,

when the equations that describe the state evolution and the measurement are linear and the noise processes that affect the state evolution and the measurement are additive and independent Gaussian processes. When the state transition and/or the measurement equations are non-linear and/or the noise processes are non-Gaussian, the Kalman filter is no more optimal. Among the inferential techniques proposed to face these difficult cases, particle filters have received in the past two decades widespread interest. The basic feature of the particle filter is to provide a non-parametric approximation to the exact distribution, thus making possible to accurately infer multi-modal distributions. Particle filtering techniques have found application in several research areas, including, to cite just a few, communication systems, data fusion, non-linear control, analysis of financial time series. Being a comprehensive survey of the bibliography out of the scope of the present paper, the interested reader is referred to the tutorial papers [2]–[6] to take a look at the world of particle filters and their applications.

In the following, we will focus on discrete-time systems with continuous state. The state process is assumed to be a first order Markov process, the measurement process is assumed to be memoryless given the state, and the distributions of the Markov state process and of the measurement noise are assumed to be known. Specifically, among the broad class of discrete-time systems with continuous Markov state, communication channels with free-running hidden state will be considered in the following.

Two prominent examples of communication channels with free-running continuous hidden state are the multiplicative phase noise channel and the multiplicative fading channel. The presence of multiplicative phase noise in radio channels, introduced by the local oscillators used in up conversion and down conversion, is well known and studied from a long time. Also, multiplicative phase noise is 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 [7]–[9]. It is intuitive that a time-varying channel, as the multiplicative phase noise channel is, can impair the information rate between the source and channel's output, this concept having been investigated several times in the past. Results on the capacity of the additive white Gaussian noise (AWGN) channel affected by memoryless multiplicative phase noise can be found in [10], [11]. The information rate transferred through the channel with memo-

Manuscript received October 24, 2013; revised February 10, 2014; accepted April 8, 2014. Date of publication April 16, 2014; date of current version June 12, 2014. L. Barletta was supported by Technische Universität München–Institute for Advanced Study through the German Excellence Initiative. This paper was presented at the 2013 IEEE International Symposium on Information Theory.

L. Barletta is with the Institute for Advanced Study, Technische Universität München, Garching D-85748, Germany (e-mail: luca.barletta@tum.de).

M. Magarini, S. Pecorino, and A. Spalvieri are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milan I-20133, Italy (e-mail: magarini@elet.polimi.it; pecorino@elet.polimi.it; spalvier@elet.polimi.it).

Communicated by R. Fischer, Associate Editor for Communications.  
Digital Object Identifier 10.1109/TIT.2014.2317694

ryless phase noise is studied in [12], while considerations on the model for continuous-time memoryless phase noise are proposed in [13]. The case of Wiener phase noise, where the phase noise process has memory and should therefore be tracked, is considered in [14]–[21]. To fit the phase noise affecting local oscillator from the real world, the richer ARMA (AutoRegressive Moving Average) model is often considered. The ARMA model better fits many cases of practical interest, because it allows for shaping the power spectral density of phase noise by acting on the order and on the parameters of the model, see e.g. [22] for a second-order ARMA model of phase noise. Working out the information rate transferred through 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 trellis-based techniques based on quantization of the state space as those used with Wiener phase noise [15], [17], because the number of states of the trellis would be enormous, and
- the observation is a nonlinear function of the state, therefore the linearized Kalman filter can be far from being optimal.

The only papers studying the information rate transferred through a channel affected by ARMA phase noise we are aware of are [14], [23], where the method of particle filtering is adopted. Recent investigations on the capacity of the fading channel with hidden Markov state can be found in [24]–[27], the most popular model for the fading spectrum being the first-order ARMA model of [24].

In this paper, upper and lower bounds to the information rate between the measurement and the hidden state are presented. The upper bound, which is based on approximate Bayesian tracking, is quite straightforward and can be found in many already published papers, while the lower bound, which is new, is obtained by Bayesian smoothing. From these bounds we derive upper and lower bounds to the information rate transferred between the input and the output of communication channels with free-running ARMA continuous hidden state. Specifically, the upper bound is already published in [17] and [23], while the lower bound is new. Evaluation of these bounds, which is presented here for the fading channel and for the phase noise channel, is based on the Kalman filter and on the particle filter. The novelty compared to [14], where particle filtering techniques are used to compute the information rate, is that we present here upper and lower bounds, while by [14] one can compute only an approximation. Compared to [23], here the evaluation method of the upper bound is new, because one of the terms appearing in the bound is based on a distribution that is allowed here to be multi-modal, while in [23] that distribution is approximated to a Gaussian one. Also, the evaluation method of the upper bound is different from [17], where trellis-based techniques are adopted. Both the upper bound and the lower bound are substantially tighter than those of [23] especially when, as it happens with the phase noise spectrum used for deriving the numerical results, inference becomes challenging due to strong phase noise and to the high-dimensional state space.

The outline of the paper is as follows. Sections II, III, and IV, focus on the evaluation of the information rate between the measurement and the hidden state. Specifically, Section II is an introductory Section which shows that the actual information rate between the measurement and the hidden state can be evaluated from the probabilities inferred by exact Bayesian tracking. Evaluation of the information rate by the Kalman filter, that will find application in Section VII, is presented as an example. In Section III the case where exact inference is not feasible is considered. To deal with this case, upper and lower bounds to the information rate are proposed. Section IV shows how the bounds of Section III can be computed by particle methods. Communication channels with free-running hidden Markov state are considered in Section V. In that Section upper and lower bounds to the information rate between the source and the output of the communication channel are derived as a by-product of the upper and lower bounds to the information rate inferred about the hidden state of the channel. These bounds are based on data-aided inference for some terms, and on data-aided inference for some others. In Section VI the multiplicative ARMA phase noise channel is analyzed in depth, deriving for it numerical results showing that the upper and lower bounds to the information rate proposed here outperform those available in the literature. To give a more complete view of applicability of the proposed method to multiplicative channels, in Section VII the multiplicative fading channel is considered. Also for this channel numerical results are presented, taking for fading spectrum the first-order ARMA model of [24]. While with the phase noise channel all the terms appearing in the bounds are computed by the particle filter, here, thanks to linearity of the data-aided measurement, the terms based on data-aided inference are computed by the conventional Kalman filter. Finally, in Section VIII the conclusion is drawn.

## II. EVALUATION OF THE INFORMATION RATE BY EXACT BAYESIAN TRACKING

Let the lowercase character  $u$  denote a column vector and let the uppercase calligraphic character  $\mathcal{U}$  denote the space spanned by  $u$ . Let the uppercase character  $U$  indicate a possibly non-stationary process,  $U = U_0, U_1, \dots$ , where the uppercase indexed letter  $U_k$  denotes a random vector whose generic realization  $u_k$  takes its values in  $\mathcal{U}$ . Also, let  $u_i^k$  denote a windowed sequence of vectors between the discrete time instant  $i$  and the discrete time instant  $k$ , that is

$$\begin{aligned} u_i^k &= (u_i, u_{i+1}, \dots, u_k), \quad 0 \leq i \leq k, \\ u_i^k &= \text{empty}, \quad \text{elsewhere.} \end{aligned}$$

For continuous random variables,  $p(u_i^k)$  is a shorthand used to indicate the multivariate probability density function  $p(U_i^k = u_i^k)$ , while, when using discrete random variables, the shorthand  $p(u_i^k)$  indicates the multivariate mass probability of  $U_i^k$  evaluated in  $u_i^k$ . The notation  $|\mathcal{U}|$  denotes the number of elements in the discrete set  $\mathcal{U}$ .

Consider a dynamical system based on the state transition equation

$$S_k = f_{k-1}(S_{k-1}, V_{k-1}), \quad (1)$$

and on the measurement equation

$$Y_k = h_k(S_k, N_k), \quad (2)$$

where, here and in what follows, we let  $k = 1, 2, \dots$ . In the above equations,  $V$  is a process of independent vectors called *process noise*,  $N$  is a process independent of  $V$  made of independent vectors and called *measurement noise*,  $S$  is the state process,  $Y$  is the measurement process, and  $\{f_{k-1}(\cdot)\}$  and  $\{h_k(\cdot)\}$  are sequences of known functions.

The dynamical system can be mapped onto the framework of first-order Markov processes. The Markovian state process  $S$  is characterized by the joint probability

$$p(s_0^n) = p(s_0) \prod_{k=1}^n p(s_k | s_{k-1}). \quad (3)$$

A measurement that is memoryless given the state is characterized by the conditional distribution

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

From the above two equations, after straightforward passages one gets

$$p(s_k | s_0^{k-1}, y_1^{k-1}) = p(s_k | s_{k-1}). \quad (5)$$

The Shannon mutual information rate between the state and the measurement, expressed in bits per measurement, is

$$\begin{aligned} I(S; Y) &= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left\{ \log_2 \left( \frac{p(Y_1^n | S_1^n)}{p(Y_1^n)} \right) \right\} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left\{ \log_2 \left( \frac{\prod_{k=1}^n p(Y_k | S_k)}{\prod_{k=1}^n p(Y_k | Y_1^{k-1})} \right) \right\} \end{aligned} \quad (6)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left\{ \log_2 \left( \frac{p(Y_k | S_k)}{p(Y_k | Y_1^{k-1})} \right) \right\} \quad (7)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n I(S_k; Y_k | Y_1^{k-1}), \quad (8)$$

where  $I(X; Y | Z)$  is the conditional mutual information rate between  $X$  and  $Y$  given  $Z$ , the numerator inside the logarithm in (6) is obtained by (4), and the denominator inside the logarithm in (6) is obtained by chain rule.

By the Shannon-McMillan-Breiman theorem, one can generate a joint sequence  $(s_0^n, y_1^n)$  according to the actual joint state transition probability and measurement probability

$$p(s_0^n, y_1^n) = p(s_0) \prod_{k=1}^n p(s_k | s_{k-1}) p(y_k | s_k) \quad (9)$$

and then evaluate the information rate as a sample estimate of (7):

$$I(S; Y) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \left( \frac{p(y_k | s_k)}{p(y_k | y_1^{k-1})} \right). \quad (10)$$

When the state transition probability and the measurement probability are known and treatable, the conditional probability  $p(y_k | y_1^{k-1})$  can be worked out by *Bayesian tracking*. Let the

Markovian state be continuous. One can track the hidden state by a two-step recursion that, for  $k = 1, 2, \dots$ , reads

$$p(s_k | y_1^{k-1}) = \int_{\mathcal{S}} p(s_k | s_{k-1}) p(s_{k-1} | y_1^{k-1}) ds_{k-1}, \quad (11)$$

$$p(s_k | y_1^k) = \frac{p(s_k | y_1^{k-1}) p(y_k | s_k)}{p(y_k | y_1^{k-1})}, \quad (12)$$

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 (12), that is the probability that we want to use in (10), is a normalization factor such that the left-hand side is a probability, therefore it can be computed by the Chapman-Kolmogorov equation

$$p(y_k | y_1^{k-1}) = \int_{\mathcal{S}} p(s_k | y_1^{k-1}) p(y_k | s_k) ds_k. \quad (13)$$

The state transition probability  $p(s_k | s_{k-1})$  appears in (11) in place of  $p(s_k | s_{k-1}, y_1^{k-1})$  thanks to (5). Thanks to (4),  $p(y_k | s_k)$  can be used in place of  $p(y_k | s_k, y_1^{k-1})$  in (12).

Note that the distribution  $p(s_0)$  of the initial state that, for  $k = 1$ , is the second factor inside the integral in the right side of (11), after a transient whose duration depends on the coherence time of the state process is forgotten. Therefore, since we let  $n \rightarrow \infty$  in (10), we can choose  $p(s_0)$  as we want because this choice does not impact the infinite sum. We have experimentally observed that the distribution  $p(s_0)$  can influence the speed of convergence of the sum to the limit it achieves as  $n \rightarrow \infty$ . In the end, the best initial distribution  $p(s_0)$  that we have found is the Dirac delta function, hence, in the simulation results to be hereafter presented, the first prediction of Bayesian tracking, that is (11) with  $k = 1$ , is

$$p(s_1) = p(s_1 | s_0),$$

meaning that the tracking algorithm starts from the actual initial state  $s_0$ .

When the measurement and the state evolution are expressed by a linear and additive noise model with Gaussian measurement noise and process noise, evaluation of the actual information rate is feasible by the Kalman filter. Specifically, the model is

$$S_k = \mathbf{F}_{k-1} S_{k-1} + V_{k-1}, \quad (14)$$

$$Y_k = \mathbf{H}_k S_k + N_k, \quad (15)$$

where the uppercase boldface character denotes matrices, and  $V_k$  and  $N_k$  are jointly independent and white Gaussian random vectors with zero mean and covariance matrices  $\mathbf{Q}_k$  and  $\mathbf{R}_k$ , respectively. The innovation process  $U$  of process  $Y$  is a white multivariate Gaussian process whose  $k$ -th element is

$$U_k = Y_k - \mathbf{H}_k \bar{\mu}_k = \mathbf{H}_k (S_k - \bar{\mu}_k) + N_k, \quad (16)$$

where

$$\bar{\mu}_k = \mathbb{E} \left\{ S_k | Y_1^{k-1} \right\}$$

is the prediction of state  $S_k$  computed by the Kalman filter. Since

$$h(U_k) = h(Y_k | Y_1^{k-1}), \quad h(N_k) = h(Y_k | S_k, Y_1^{k-1}), \quad (17)$$

where  $h(X_k)$  denotes the differential entropy of  $X_k$ , an unbiased random estimate  $\hat{I}(S_k; Y_k | Y_1^{k-1})$  of the  $k$ -th term of the sum (10) is

$$\begin{aligned} \hat{I}(S_k; Y_k | Y_1^{k-1}) &= \hat{h}(U_k) - \hat{h}(N_k) \\ &= \frac{1}{2} \log_2 \frac{\det(\mathbf{H}_k \bar{\Sigma}_k \mathbf{H}_k^T + \mathbf{R}_k)}{\det(\mathbf{R}_k)} \\ &= \frac{1}{2} \log_2 \det \left( \mathbf{I} + \mathbf{R}_k^{-1} \mathbf{H}_k \bar{\Sigma}_k \mathbf{H}_k^T \right), \end{aligned} \quad (18)$$

where  $\mathbf{I}$  is the identity matrix,

$$\bar{\Sigma}_k = \mathbb{E} \left\{ (S_k - \bar{\mu}_k)^T (S_k - \bar{\mu}_k) | Y_1^{k-1} \right\}$$

is the covariance matrix of the error between the state and its prediction computed by the Kalman filter at time  $k$ , and

$$h(X_k) = \frac{1}{2} \log_2 \left( (2\pi e)^m \det(\Psi_k) \right)$$

is the differential entropy of the  $m$ -variate Gaussian random vector  $X_k$  with covariance matrix  $\Psi_k$ .

### III. UPPER AND LOWER BOUNDS TO THE INFORMATION RATE BY APPROXIMATED BAYESIAN INFERENCE

In many cases of practical interest, although the state transition probability and the measurement probability are known and treatable, it happens that the posterior and predictive probabilities are not treatable due to lack of linearity and/or Gaussianity. In these cases, one can generate a long sequence  $(s_0^n, y_1^n)$  according to the treatable joint probability (9) and work out an approximation to the non-treatable probabilities by approximated Bayesian tracking. To assess the quality of the approximation, we propose to evaluate an upper bound on the information rate based on the distributions inferred by Bayesian filtering, and a lower bound below the information rate based on the distributions inferred by Bayesian smoothing. When the upper bound is close to the lower bound, one can claim of having virtually computed the actual information rate and that the inferred distributions closely fit the actual ones.

#### A. An Upper Bound Based on Bayesian Filtering

The upper bound is

$$\bar{I}(S; Y) = \bar{h}(Y) - h(Y|S) \geq I(S; Y), \quad (19)$$

$$\bar{h}(Y) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \frac{1}{q(y_k | y_1^{k-1})} \geq h(Y), \quad (20)$$

where the probability  $q(y_k | y_1^{k-1})$  is the approximation to  $p(y_k | y_1^{k-1})$  worked out as the normalization factor of the update step of the approximate Bayesian tracking, and  $y_1^n$  is a realization of the actual joint state transition and measurement probability. The inequality in (20) follows by Gibbs' inequality, and it holds for any probability  $q(y_k | y_1^{k-1})$ .

#### B. A Lower Bound Based on Bayesian Smoothing

The lower bound is

$$\underline{I}(S; Y) = h(S) - \bar{h}(S|Y) \leq I(S; Y). \quad (21)$$

Invoking the Shannon-McMillan-Breiman theorem (22), the chain rule (23), the known initial state discussed before the end of Section II (24), the Markov property (25), and Gibbs' inequality (26), we have

$$h(S|Y) = \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \frac{1}{p(s_1^n | y_1^n)} \quad (22)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \frac{1}{p(s_1 | y_1^n) \prod_{k=2}^n p(s_k | y_1^n, s_1^{k-1})} \quad (23)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \frac{1}{p(s_k | y_1^n, s_0^{k-1})} \quad (24)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \frac{1}{p(s_k | y_k^n, s_{k-1})} \quad (25)$$

$$\leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \frac{1}{q(s_k | y_k^{k+l}, s_{k-1})} \quad (26)$$

$$= \bar{h}(S|Y), \quad (27)$$

where the probability  $q(s_k | y_k^{k+l}, s_{k-1})$  is the approximation to  $p(s_k | y_k^n, s_{k-1})$  worked out by a lag- $l$  Bayesian smoother initialized from the state  $s_{k-1}$  visited by the realization at time  $k-1$ , the time lag  $l$  being up to the user. If the state sequence is a reversible function of the process noise  $V$  given the initial state  $s_0$ , then

$$I(S; Y) = I(V; Y) \geq h(V) - \bar{h}(V|Y), \quad (28)$$

where the upper bound on the conditional differential entropy rate can be evaluated as

$$\bar{h}(V|Y) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \frac{1}{q(v_{k-1} | y_k^{k+l}, v_0^{k-2}, s_0)} \quad (29)$$

$$= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \frac{1}{q(v_{k-1} | y_k^{k+l}, s_{k-1})}. \quad (30)$$

### IV. COMPUTING THE BOUNDS BY PARTICLE METHODS

As the measurement equation is nonlinear in the state variable, we need to provide non-parametric approximations to the true distributions, that in general can be multimodal. Particle methods are practical tools for estimating distributions in a non-parametric way, and in this section we use these techniques for computing the upper bounds  $\bar{h}(Y)$  and  $\bar{h}(S|Y)$  introduced in the previous section.

Let  $P$  be the number of particles,  $s_k^{(i)}$  the state visited by the  $i$ -th particle at time  $k$ ,  $w_k^{(i)}$  the weight of the  $i$ -th particle at time  $k$ , and  $\pi(s_k | s_{k-1}, y_k)$  the *importance density* at time  $k$ , which is up to the user. Starting from uniform initial weights  $\{w_0^{(i)} = P^{-1}, i = 1, 2, \dots, P\}$  and from an initial set of

particles  $\{s_0^{(i)} = s_0, i = 1, 2, \dots, P\}$ , the predict step of particle tracking is

$$s_k^{(i)} \sim \pi(s_k | s_{k-1}^{(i)}, y_k), \quad i = 1, 2, \dots, P, \quad (31)$$

where  $\sim$  means *drawn with probability*. The update step of particle tracking is

$$w_k^{(i)} = w_{k-1}^{(i)} \frac{p(y_k | s_k^{(i)}) p(s_k^{(i)} | s_{k-1}^{(i)})}{\alpha_k \pi(s_k^{(i)} | s_{k-1}^{(i)}, y_k)}, \quad i = 1, 2, \dots, P, \quad (32)$$

where  $\alpha_k$  is a normalization factor such that  $\sum_{i=1}^P w_k^{(i)} = 1$ . Given the set of weights and particles one has the approximation

$$p(s_0^k | y_1^k) \approx \sum_{i=1}^P w_k^{(i)} \delta(s_0^k - s_0^{k,(i)}), \quad (33)$$

where  $\delta(\cdot)$  is the Dirac delta function. From (33) one has

$$\begin{aligned} p(s_k | y_1^k) &= \int_{S_0^{k-1}} p(s_0^k | y_1^k) ds_0^{k-1} \\ &\approx \int_{S_0^{k-1}} \sum_{i=1}^P w_k^{(i)} \delta(s_0^k - s_0^{k,(i)}) ds_0^{k-1} \\ &= \sum_{i=1}^P w_k^{(i)} \delta(s_k - s_k^{(i)}). \end{aligned} \quad (34)$$

After updating the particles with (32), a *resampling* procedure may be necessary to prevent particles from collapsing onto one particle of weight 1. Commonly used resampling procedures are described in [1].

In the experimental results presented in the following we adopt  $\pi(s_k | s_{k-1}, y_k) = p(s_k | s_{k-1})$ . With this choice of the importance function, the normalization factor of (32) is

$$\alpha_k = \sum_{i=1}^P w_{k-1}^{(i)} p(y_k | s_k^{(i)}), \quad (35)$$

and the predict step is

$$s_k^{(i)} = f_{k-1}(s_{k-1}^{(i)}, v_{k-1}^{(i)}), \quad i = 1, 2, \dots, P, \quad (36)$$

where  $\{v_{k-1}^{(i)}, i = 1, 2, \dots, P\}$  is a set of independent samples of process noise.

#### A. Evaluation of $\bar{h}(Y)$

As in [14], the probability  $q(y_k | y_1^{k-1})$  used in the upper bound is obtained as the factor that normalizes the weights of the particles in the update step:

$$q(y_k | y_1^{k-1}) = \sum_{i=1}^P w_{k-1}^{(i)} p(y_k | s_k^{(i)}). \quad (37)$$

The entire procedure for Monte-Carlo evaluation of  $\bar{h}(Y)$  is reported in Algorithm 1. The initial state is selected as  $s_0 = 0_m$ , where  $0_m$  is a vector of  $m$  zeros.

---

#### Algorithm 1 Calculate $\bar{h}(Y)$

---

Generate samples:

$$(s_0^n, y_1^n) \sim p(s_0^n, y_1^n) = \delta(s_0) \prod_{k=1}^n p(s_k | s_{k-1}) p(y_k | s_k)$$

$$s_0^{(i)} \leftarrow 0_m \text{ for } i = 1, \dots, P$$

$$w_0^{(i)} \leftarrow P^{-1} \text{ for } i = 1, \dots, P$$

**for**  $k = 1, \dots, n$  **do**

    Generate  $v_{k-1}^{(i)} \sim p(v_{k-1})$  for  $i = 1, \dots, P$

$$s_k^{(i)} \leftarrow f_{k-1}(s_{k-1}^{(i)}, v_{k-1}^{(i)}) \text{ for } i = 1, \dots, P$$

$$w_k^{(i)} \leftarrow w_{k-1}^{(i)} p(y_k | s_k^{(i)}) \text{ for } i = 1, \dots, P$$

$$\alpha_k \leftarrow \sum_{i=1}^P w_k^{(i)}$$

$$w_k^{(i)} \leftarrow w_k^{(i)} / \alpha_k \text{ for } i = 1, \dots, P$$

**if**  $\sum_{i=1}^P (w_k^{(i)})^2 > (0.3P)^{-1}$  **then**

$$\quad (\{s_k^{(i)}\}, \{w_k^{(i)}\}) \leftarrow \text{resample}(\{s_k^{(i)}\}, \{w_k^{(i)}\})$$

**end if**

**end for**

$$\bar{h}(Y) \leftarrow -n^{-1} \sum_{k=1}^n \log_2 \alpha_k$$


---

#### B. Evaluation of $\bar{h}(S|Y)$

At time instant  $k$  and lag  $\ell = 0$  the particles for  $i = 1, \dots, P$  are initialized as

$$s_{k,0}^{(i)} = f_{k-1}(s_{k-1}, v_{k-1,0}^{(i)})$$

with weight

$$w_{k,0}^{(i)} = \frac{p(y_k | s_{k,0}^{(i)})}{\sum_{j=1}^P p(y_k | s_{k,0}^{(j)})},$$

where the set  $\{v_{k-1,0}^{(i)}, i = 1, 2, \dots, P\}$  is a set of independent samples of process noise, and  $s_{k-1}$  is the state visited at time  $k-1$  by the realization  $(s_0^n, y_1^n)$ . For each time lag  $\ell = 1, \dots, l$  the particles and their weights are updated as

$$\begin{aligned} s_{k,\ell}^{(i)} &= f_{k-1+\ell}(s_{k,\ell-1}^{(i)}, v_{k-1,\ell}^{(i)}) \\ w_{k,\ell}^{(i)} &= \frac{w_{k,\ell-1}^{(i)} p(y_{k+\ell} | s_{k,\ell}^{(i)})}{\sum_{j=1}^P w_{k,\ell-1}^{(j)} p(y_{k+\ell} | s_{k,\ell}^{(j)})}, \end{aligned}$$

where  $\{v_{k-1,\ell}^{(i)}, i = 1, 2, \dots, P\}$  for  $\ell = 1, \dots, l$  are sets of independent samples of the process noise. After  $l$  steps, using (33) one gets

$$\begin{aligned} p(s_k | y_k^{k+l}, s_{k-1}) &= \int_{S_{k+1}^{k+l}} p(s_k^{k+l} | y_k^{k+l}, s_{k-1}) ds_{k+1}^{k+l} \\ &\approx \int_{S_{k+1}^{k+l}} \sum_{i=1}^P w_{k,l}^{(i)} \delta(s_k^{k+l} - s_{k,0;l}^{(i)}) ds_{k+1}^{k+l} \\ &= \sum_{i=1}^P w_{k,l}^{(i)} \delta(s_k - s_{k,0;l}^{(i)}), \end{aligned} \quad (38)$$

where  $s_{k,0;l}^{(i)} = (s_{k,0}^{(i)}, s_{k,1}^{(i)}, \dots, s_{k,l}^{(i)})$ . Since the evaluation of (38) in the point  $s_k$  visited by the realization requires that the inferred distribution is actually a probability density function, a smooth kernel should be used in place of the Dirac delta,

leading to

$$q(s_k|y_k^{k+l}, s_{k-1}) = \sum_{i=1}^P w_{k,l}^{(i)} \kappa(s_{k,0}^{(i)}; s_k), \quad (39)$$

where the kernel  $\kappa(\mu; x)$  is a probability density function over the space spanned by  $x$  with mean vector  $\mu$ . In the numerical examples to be presented in the following, the state sequence is a reversible transformation of the process noise given the initial state, therefore the wanted bound can be evaluated by (30) with

$$q(v_{k-1}|y_k^{k+l}, s_{k-1}) = \sum_{i=1}^P w_{k,l}^{(i)} \kappa(v_{k-1,0}^{(\rho(i))}; v_{k-1}), \quad (40)$$

where  $\rho(\cdot)$  is a function used in the resampling procedure and discussed later. The kernel that we adopt is

$$\kappa(\mu; x) = (1 - \alpha)g(\mu, \sigma^2 \mathbf{I}; x) + \alpha u(\mu, \Delta; x), \quad (41)$$

where  $g(\mu, \sigma^2 \mathbf{I}; x)$  is a multivariate Gaussian probability density function with mean vector  $\mu$  and covariance matrix  $\sigma^2 \mathbf{I}$  over the space spanned by  $x$ ,  $u(\mu, \Delta; x)$  is a uniform distribution over a hypercube of center  $\mu$  and side  $\Delta$  over the space spanned by  $x$ , and  $0 < \alpha < 1$ . We take  $\alpha$  small and  $\Delta$  large enough to prevent problems of numerical stability that occur with the pure Gaussian kernel when  $q(v_{k-1}|y_k^{k+l}, s_{k-1})$  is evaluated in a point  $v_{k-1}$  that falls far from all the samples of the set  $\{v_{k-1,0}^{(i)}, i = 1, 2, \dots, P\}$ . To optimize the bound, (30) is computed for several values of  $\sigma^2$  and then the minimum is taken.

It is worth mentioning that, when using the particle resampling procedure, it is important to consider the right particles, and in the right order, of the set  $\{v_{k-1,0}^{(i)}, i = 1, 2, \dots, P\}$  in such a way that, after  $l$  steps, the  $i$ -th particle  $s_{k,l}^{(i)}$  was generated by  $v_{k-1,0}^{(\rho(i))}$ . For example, if  $P = 4$  and the particles  $\{s_{k,l}^{(1)}, s_{k,l}^{(2)}, s_{k,l}^{(3)}, s_{k,l}^{(4)}\}$  are generated by  $\{v_{k-1,0}^{(2)}, v_{k-1,0}^{(2)}, v_{k-1,0}^{(1)}, v_{k-1,0}^{(4)}\}$ , respectively, then  $\rho(1) = \rho(2) = 2$ ,  $\rho(3) = 1$ , and  $\rho(4) = 4$  in (40).

The entire procedure for Monte-Carlo evaluation of  $\bar{h}(S|Y)$  is reported in Algorithm 2, again for  $s_0 = 0_m$ .

## V. INFORMATION RATE TRANSFERRED THROUGH CHANNELS WITH FREE-RUNNING STATE

In this section we show how to use the bounds proposed in Section III in order to derive bounds on the information rate transferred through channels with free-running state and without channel state information at the transmitter. In order to relate the notation used in Section III to the one used in this section, we introduce the concepts of data-aided and blind inference.

Consider a communication channel described by the state transition probability (3) and by the channel probability

$$p(r_1^n | x_1^n, s_1^n) = \prod_{k=1}^n p(r_k | x_k, s_k), \quad (42)$$

where  $R$  is the channel output process and  $X$  is the source process made of discrete random variables. Equation (42) says

---

### Algorithm 2 Calculate $\bar{h}(S|Y)$

---

Generate samples:

$$(s_0^{n+l}, y_1^{n+l}) \sim \delta(s_0) \prod_{k=1}^{n+l} p(s_k | s_{k-1}) p(y_k | s_k)$$

Compute  $v_0^{n+l-1}$  from  $s_0^{n+l}$

**for**  $k = 1, \dots, n$  **do**

Generate  $v_{k-1,0}^{(i)} \sim p(v_{k-1})$  for  $i = 1, \dots, P$

$\rho(i) = i$  for  $i = 1, \dots, P$

$s_{k,0}^{(i)} \leftarrow f_{k-1}(s_{k-1}, v_{k-1,0}^{(i)})$  for  $i = 1, \dots, P$

$w_{k,0}^{(i)} \leftarrow p(y_k | s_{k,0}^{(i)}) / \sum_{j=1}^P p(y_k | s_{k,0}^{(j)})$  for  $i = 1, \dots, P$

**for**  $\ell = 1, \dots, l$  **do**

**if**  $\sum_{i=1}^P (w_{k,\ell-1}^{(i)})^2 > (0.3P)^{-1}$  **then**

$\{s_{k,\ell-1}^{(i)}, \{w_{k,\ell-1}^{(i)}, \{\rho(i)\}\} \leftarrow$

resample  $(\{s_{k,\ell-1}^{(i)}, \{w_{k,\ell-1}^{(i)}, \{\rho(i)\}\})$

**end if**

Generate  $v_{k-1,\ell}^{(i)} \sim p(v_{k-1+\ell})$  for  $i = 1, \dots, P$

$s_{k,\ell}^{(i)} \leftarrow f_{k-1+\ell}(s_{k,\ell-1}, v_{k-1,\ell}^{(i)})$  for  $i = 1, \dots, P$

$w_{k,\ell}^{(i)} \leftarrow w_{k,\ell-1}^{(i)} p(y_{k+\ell} | s_{k,\ell}^{(i)}) / \sum_{j=1}^P w_{k,\ell-1}^{(j)} p(y_{k+\ell} | s_{k,\ell}^{(j)})$   
for  $i = 1, \dots, P$

**end for**

$q(v_{k-1}|y_k^{k+l}, s_{k-1}) \leftarrow \sum_{i=1}^P w_{k,l}^{(i)} \kappa(v_{k-1,0}^{(\rho(i))}; v_{k-1})$

**end for**

$\bar{h}(S|Y) \leftarrow -n^{-1} \sum_{k=1}^n \log_2 q(v_{k-1}|y_k^{k+l}, s_{k-1})$

---

that the channel output process is memoryless given the source and the state. Also, consider the case of free-running state, where the source is memoryless and independent of the state, that is

$$p(x_1^n | s_1^n) = \prod_{k=1}^n p(x_k). \quad (43)$$

Examples of channels with free-running state are multiplicative channels as the phase noise channel and the fading channel.

### A. Data-Aided Inference

Putting together (42) and (43) one finds that the joint source and channel model is memoryless given the state:

$$p(r_1^n, x_1^n | s_1^n) = \prod_{k=1}^n p(r_k, x_k | s_k), \quad (44)$$

hence putting the pair  $(R, X)$  in place of  $Y$  in Sections III and IV we have  $I(R, X; S)$  from exact Bayesian inference, while we have the upper and lower bounds to  $I(R, X; S)$  from approximated Bayesian inference. By independence between  $X$  and  $S$  we have

$$I(R, X; S) = I(R; S|X). \quad (45)$$

The above equation, which read as ‘‘given  $X$ ,’’ shows that inference can be based on the knowledge of  $X$ , as if  $X$  were part of the observation. Therefore, drawing from the parlance of channel estimation, we hereafter call *data-aided* inference the one that is performed when the measurement  $Y$  is the pair  $(R, X)$ , and call *data-aided* channel probability the probability  $p(r_k | x_k, s_k)$ .

### B. Blind Inference

Using (44) one finds that channel's output is memoryless given the state:

$$\begin{aligned} p(r_1^n | s_1^n) &= \sum_{x_1^n \in \mathcal{X}_1^n} p(r_1^n, x_1^n | s_1^n) = \sum_{x_1^n \in \mathcal{X}_1^n} \prod_{k=1}^n p(r_k, x_k | s_k) \\ &= \prod_{k=1}^n \sum_{x_k \in \mathcal{X}_k} p(r_k, x_k | s_k) = \prod_{k=1}^n p(r_k | s_k). \end{aligned} \quad (46)$$

Since Bayesian inference, that is performed using  $R$  as a measurement process, is not aware of channel's input, drawing again from the parlance of channel estimation, we call it *blind* inference. We call the channel transition probability  $p(r_k | s_k)$  *blind* channel probability. The blind information rate is not greater than the data-aided information rate:

$$I(R; S) \leq I(R; S) + I(X; S|R) = I(R, X; S) = I(R; S|X) \quad (47)$$

where (47) follows by nonnegativity of mutual information, chain rule, and independence between  $X$  and  $S$ .

### C. Information Rate

Since

$$I(X; R) = I(X; R|S) + I(S; R) - I(S; R|X) \quad (48)$$

one can sandwich the information rate transferred through the channel as

$$\begin{aligned} \bar{I}(X; R) &= I(X; R|S) + \bar{I}(S; R) - \underline{I}(S; R|X) \\ &\geq I(X; R) \\ &\geq I(X; R|S) + \underline{I}(S; R) - \bar{I}(S; R|X) = \underline{I}(R; X), \end{aligned} \quad (49)$$

$$(50)$$

where, using differential entropy rates, one has

$$\begin{aligned} \bar{I}(X; R) &= \bar{h}(R) + \bar{h}(S|X, R) - h(S|X) - h(R|X, S) \\ &\geq I(X; R) \\ &\geq h(S) + h(R|S) - \bar{h}(S|R) - \bar{h}(R|X) = \underline{I}(R; X). \end{aligned} \quad (51)$$

$$(52)$$

The expression of the upper bound is the same as [17], [23], while the lower bound is new. To compute the differential entropy rates appearing in (51) and (52), we need to work out  $\bar{h}(R)$  and  $\bar{h}(R|X)$  by Bayesian filtering, and  $\bar{h}(S|R)$  and  $\bar{h}(S|X, R)$  by Bayesian smoothing. Recall that  $h(S) = h(S|X)$  is known and that  $h(R|S)$  and  $h(R, X|S)$ , which are those of the memoryless channel, are also assumed to be known. The gap between the upper bound (49) and the lower bound (50) is equal to the gap between upper and lower bounds of blind inference ( $\bar{I}(S; R) - \underline{I}(S; R)$ ) plus the gap between upper and lower bounds of data-aided inference ( $\bar{I}(S; R|X) - \underline{I}(S; R|X)$ ). Also, it holds that

$$\bar{I}(X; R) \geq \bar{h}(R) - \bar{h}(R|X) \geq \underline{I}(X; R), \quad (53)$$

where the sandwiched term is the approximation to the information rate proposed in [14]. We also mention the demodulation lower bound of [23], that we will use as a competitor of (52) in the sections devoted to experimental results. It reads as

$$\underline{I}(X; R) = H(X) - \bar{H}(X|R) \leq I(X; R), \quad (54)$$

where  $H(X)$  is the entropy rate of process  $X$  and

$$\bar{H}(X|R) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \frac{1}{q(x_k | r_1^k, x_1^{k-1})} \quad (55)$$

$$\begin{aligned} &\geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \frac{1}{p(x_k | r_1^k, x_1^{k-1})} \\ &\geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \log_2 \frac{1}{p(x_k | r_1^n, x_1^{k-1})} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \frac{1}{p(x_1^n | r_1^n)} \\ &= H(X|R), \end{aligned} \quad (56)$$

where  $q(x_k | r_1^k, x_1^{k-1})$  is the approximation to  $p(x_k | r_1^k, x_1^{k-1})$  obtained by a demodulator aware of past data.

### VI. DISCRETE-TIME ARMA PHASE NOISE CHANNEL

The concepts developed so far are applied in this section to the ARMA multiplicative phase noise channel. The  $k$ -th output of the channel is

$$R_k = X_k e^{j\Phi_k} + N_k, \quad (57)$$

where  $j$  is the imaginary unit,  $R$  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,  $N$  is the complex AWGN process with zero mean and variance  $\text{SNR}^{-1}$ , and  $\Phi$  is the phase noise process which is assumed to be independent of  $X$  and  $N$ .

The measurement probability in data-aided inference is

$$\begin{aligned} p(r_k, x_k | s_k) &= p(r_k, x_k | \phi_k) = p(x_k | \phi_k) p(r_k | x_k, \phi_k) \\ &= p(x_k) p(r_k | x_k, \phi_k) = p(x_k) g_c(x_k e^{j\phi_k}, \text{SNR}^{-1}; r_k), \end{aligned} \quad (58)$$

where  $g_c(\mu, \sigma^2; x)$  indicates a circular symmetric Gaussian probability density function over the complex plane spanned by  $x$  with mean  $\mu$  and two-dimensional variance  $\sigma^2$ . The measurement probability in blind inference is

$$p(r_k | \phi_k) = \sum_{x_k \in \mathcal{X}} p(x_k) g_c(x_k e^{j\phi_k}, \text{SNR}^{-1}; r_k). \quad (59)$$

Process  $\Phi$  is hereafter modelled as accumulation of frequency noise, that is

$$\Phi(z) = \frac{z^{-1}}{1 - z^{-1}} \Lambda(z), \quad (60)$$

where the frequency noise process  $\Lambda$  is the sequence of coefficients of the polynomial of complex variable  $z$

$$\Lambda(z) = c(z)V(z) \quad (61)$$

where  $V$  is white Gaussian noise with zero mean and variance  $\gamma^2$ , and

$$c(z) = \frac{\prod_{k=1}^m (1 - \beta_k z^{-1})}{\prod_{k=1}^m (1 - \alpha_k z^{-1})} = \frac{1 + \sum_{k=1}^m b_k z^{-k}}{1 - \sum_{k=1}^m a_k z^{-k}}, \quad (62)$$

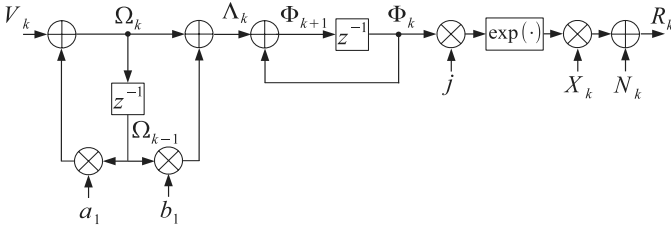


Fig. 1. Block diagram of the system given in equations (57)-(63) with  $m = 1$ .

where  $|\alpha_k| < 1$ ,  $|\beta_k| \leq 1$ , therefore the transfer function  $c(z)$  is causal, monic, and minimum phase. Since the phase is observed through the complex exponential, to prevent the overflow in the accumulation one can periodically reduce it modulo  $2\pi$ .

The ARMA phase noise can be cast in the framework of dynamical systems [1, Sec. 7.2] by defining the state at time  $k$  as the  $(m + 1)$  column vector

$$S_k = (\Phi_k, \Omega_{k-m}^{k-1})^T, \quad (63)$$

where, modelling the filter with transfer function (62) as a shift register with feedback taps  $a_1^m$  and forward taps  $b_1^m$ ,  $\Omega_{k-m}^{k-1}$  is the content of the shift register at the  $k$ -th channel use, that is

$$\Omega(z) = \frac{V(z)}{1 - a(z)}.$$

Figure 1 shows the block diagram of the channel model given by equations (57) to (63) with  $m = 1$ .

The state transition equation is

$$S_k = \mathbf{F}S_{k-1} + (V_{k-1}, V_{k-1}, 0_{m-1}^T)^T,$$

where the state transition matrix is

$$\mathbf{F} = \begin{bmatrix} 1 & (a_1^m + b_1^m)^T \\ 0 & (a_1^m)^T \\ 0_{m-1} & \mathbf{I}_{m-1} & 0_{m-1} \end{bmatrix},$$

with  $\mathbf{I}_m$  denoting the identity matrix of size  $m \times m$ . Given  $S_{k-1}$ ,  $S_k$  is determined if also  $V_{k-1}$  is known, hence the covariance matrix of the state transition probability has unit rank. Specifically,

$$p(s_k | s_{k-1}) = g(\mathbf{F}s_{k-1}, \boldsymbol{\Sigma}_v; s_k), \quad (64)$$

where

$$\boldsymbol{\Sigma}_v = \begin{bmatrix} \gamma^2 & \gamma^2 & 0_{m-1}^T \\ \gamma^2 & \gamma^2 & 0_{m-1}^T \\ 0_{m-1} & 0_{m-1} & 0_{(m-1) \times (m-1)} \end{bmatrix}, \quad (65)$$

where  $0_{m \times m}$  is an all-zero  $m \times m$  matrix. Note that, while the state transition equation is linear, the measurement equation is nonlinear both in data-aided tracking and in blind tracking, hence we have to renounce to exact Bayesian tracking with the Kalman filter. For sufficiently small phase noise and data-aided tracking, one can linearize the complex exponential and use the linearized Kalman filter to perform approximated Bayesian tracking as in [23], [28]–[30].

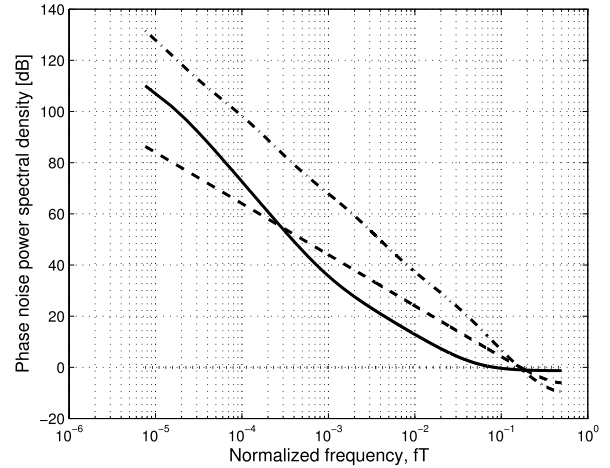


Fig. 2. Power spectral density of phase noise generated by accumulating white Gaussian noise with zero mean and unit variance filtered through a causal, monic, and minimum phase transfer function. Solid line: phase noise model of [22]. Dash-dotted line: phase noise generated by (66) with  $m = 4$  followed by accumulation. Dashed line: Wiener phase noise. Dotted line: white phase noise.

#### A. Numerical Results

As a representative case of a class of frequency noise spectra that are difficult to deal with we take

$$c(z) = \prod_{i=1}^m \frac{1 - (1 - 3 \cdot 4^{-2i+1})z^{-1}}{1 - (1 - 3 \cdot 4^{-2i})z^{-1}}. \quad (66)$$

The  $m$  poles and  $m$  zeros in the right side of (66) are interleaved and spectrally spaced of two octaves from each other. Starting from low frequency, one finds for  $i = m$  the pole at  $z = 1 - 3 \cdot 4^{-2m}$ . This pole is followed by pairs of the type zero-pole, and the sequence of zeros and poles terminates when  $i = 1$  with the zero at  $z = 0.25$ . Denoting by  $T$  the time delay represented by  $z^{-1}$ , the transfer function (66) is that of a low-pass filter with  $-3$  dB normalized frequency

$$f_{-3}T \approx \frac{3 \cdot 4^{-2m}}{2\pi}$$

determined by the pole at  $z = 1 - 3 \cdot 4^{-2m}$ . Figure 2 reports the power spectral density of four different spectra of phase noise.

From Fig. 2 one appreciates that the spectrum of phase noise obtained by frequency noise generated by (66) closely fits the slope of  $-30$  dB/decade at normalized frequency higher than  $f_{-3}T$ , a slope that is often encountered in real world oscillators. The frequency noise that generates a phase noise whose spectrum is a slope of  $-30$  dB/decade is called Flicker frequency noise, or pink frequency noise, and its spectrum shows a slope of  $-10$  dB/decade.

Upper and lower bounds to the information rate between the state and the measurement for blind and data-aided tracking are worked out by the particle filter. The results for 4-QAM and 16-QAM with  $\gamma = 0.5$ ,  $m = 4$ , and  $10^4$  particles are reported in Fig. 3.

The upper and lower bounds of Fig. 3 are used to draw the upper and lower bounds to the information rate between



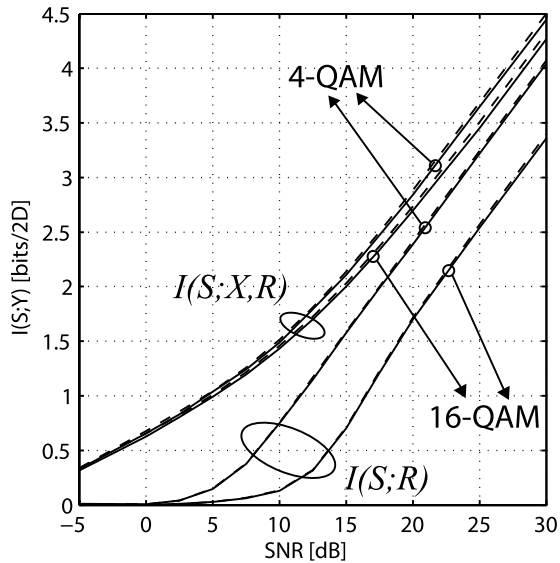


Fig. 3. Phase noise channel. Mutual information  $I(S; Y)$  computed by the particle filter with  $10^4$  particles versus SNR with data-aided tracking ( $I(S; X, R)$ ) and blind tracking ( $I(S; R)$ ). Dashed line: upper bound. Solid line: lower bound.

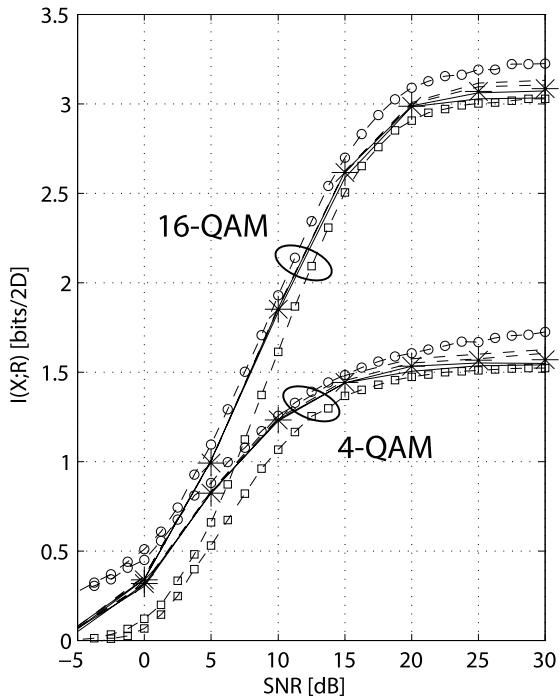


Fig. 4. Phase noise channel. Mutual information  $I(X; R)$  versus SNR. Dashed line: upper bound with, from the upper to the lower curve,  $10^4$  particles,  $5 \cdot 10^4$  particles. Solid line: lower bound with, from the lower to the upper curve,  $10^4$  particles,  $5 \cdot 10^4$  particles. Asterisks: sandwiched term in (53) with  $5 \cdot 10^4$  particles. Circles: upper bound of [23] with  $10^4$  particles. Squares: demodulation lower bound (55) based on data-aided linearized Kalman filter.

the input modulation and the output of the channel reported in Fig. 4. Figure 4 shows that the upper bound (51), when evaluated as proposed here, is substantially tighter than when it is evaluated as proposed in [23]. The reason is that, although also the bound of [23] is based on particle techniques, the inferred probability in [23] is assumed to be Gaussian, the mean and variance of the Gaussian distribution being

computed from the particles, while here the inferred distributions are allowed to be multi-modal. Concerning the lower bounds of Fig. 4, we see that the lower bound (52) outperforms the lower bound proposed in [23], [28], and [29] which relies upon demodulation performed by a linearized Kalman filter.

## VII. DISCRETE-TIME GAUSS-MARKOV FADING CHANNEL

Another example of communication channel with free-running hidden state is the multiplicative fading channel. The  $k$ -th output of the channel is

$$R_k = X_k \Lambda_k + N_k, \quad (67)$$

where  $X$  is the same as in Section VI,  $\Lambda$  is the complex fading process which is assumed to be independent of  $X$  and  $N$ , and  $N$  is complex white Gaussian noise with zero mean and two-dimensional variance  $\mathbb{E}\{|\Lambda_k|^2\} \text{SNR}^{-1}$ . A convenient model for process  $\Lambda$  is again the ARMA model, where the state of the ARMA model and the state transition equation are defined in a straightforward way following the line of the previous section. Blind inference is performed with the particle filter/smoothing taking process  $R$  as the measurement process and

$$p(r_k | s_k) = \sum_{x_k \in \mathcal{X}} p(x_k) g_c(x_k \lambda_k, \text{SNR}^{-1}; r_k) \quad (68)$$

as the measurement probability. Exact data-aided Bayesian filtering is feasible with the Kalman filter, therefore the data-aided information rate  $I(S; R, X)$  can be exactly evaluated using (18) in (8) and substituted in (49) and (50) in place of the bounds, leading to

$$\begin{aligned} \bar{I}(X; R) &= \bar{h}(R) + h(S|X, R) - h(S|X) - h(R|X, S) \\ &= \bar{h}(R) - h(R|X) \\ &\geq I(X; R) \\ &\geq h(S) + h(R|S) - \bar{h}(S|R) - h(R|X) = \underline{I}(R; X). \end{aligned} \quad (69)$$

Since  $\bar{h}(R)$  is worked out by the particle filter, the upper bound (69) coincides with the approximation of [14] and with the upper bound of [23]. Conversely, the lower bound (70) is still different from (55).

### A. Numerical Results

A first-order model is assumed in [24] for the power spectral density of  $\Lambda$ , while in [25] a brickwall spectrum is considered. In what follows, we will take for  $\Lambda$  the first-order model of [24], that is

$$\Lambda(z) = \frac{\sqrt{\gamma} z^{-1} V(z)}{1 - \sqrt{1 - \gamma} z^{-1}}, \quad (71)$$

where the complex process noise  $V$  has zero mean and unit two-dimensional variance, and  $0 < \gamma < 1$  is a parameter that determines the bandwidth of the fading process. The frequency response of the filter has unit energy, therefore the additive white Gaussian channel noise has two-dimensional variance  $\text{SNR}^{-1}$ . Upper and lower bounds to the information rate between channel's input and output for 4-QAM and 16-QAM with  $\gamma = 0.1$  are reported in Fig. 5. Note that, in contrast to

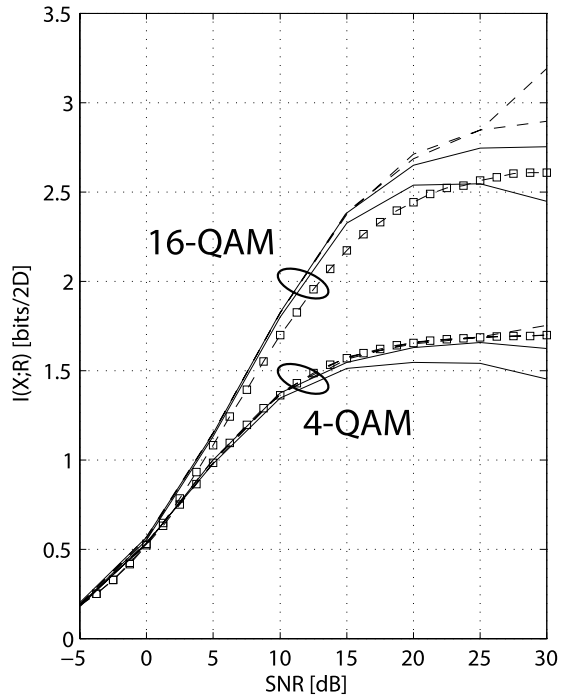


Fig. 5. Gauss-Markov fading channel. Mutual information  $I(X; R)$  versus SNR. Dashed line: upper bound with, from the upper to the lower curve,  $10^4$  particles,  $10^5$  particles. Solid line: lower bound with, from the lower to the upper curve,  $10^4$  particles,  $10^5$  particles. Squares: demodulation lower bound (54) based on data-aided linearized Kalman filter.

the case of phase noise, here, since exact data-aided Kalman filtering is performed, the probability  $q(x_k | r_1^k, x_1^{k-1})$  appearing in (55) is equal to the actual  $p(x_k | r_1^k, x_1^{k-1})$ . Therefore the inaccuracy of the bound (55) is due only to the conditions  $r_{k+1}^n$  that are removed in inequality (56). These conditions bring a contribution of non data-aided type to demodulation which, at low SNR, seems to have minor impact on the information rate extracted by demodulation. In contrast, in the phase noise channel, the inaccuracy introduced in (55) by linearizing the measurement equation can be large, especially at low SNR. Also note that the lower bound (54) is remarkably tight with 4-QAM, while it is less tight with 16-QAM, especially at intermediate-to-high SNR. Again, this can be explained by observing that, with 16-QAM, discarding the conditions  $r_{k+1}^n$  can impact the quality of demodulation much more than with 4-QAM. This can be seen by noting that, at high SNR, the decision error probability is small, therefore the quality of blind, e.g. decision-directed, smoothing is virtually equal to the quality of data-aided filtering. When the fading coefficient is small and the pattern of input data shows symbols with low amplitude up to time  $k$  and symbols of high amplitude in the future time instants, then future measurements, although non data-aided, can potentially contribute more than the past data-aided measurements to the inference made on the fading coefficient. Therefore, in these conditions, renouncing to blind smoothing means renouncing to substantial information about the fading coefficient, hence to substantial information rate.

## VIII. CONCLUSION

In the paper, Shannon information between the hidden Markov state process of a dynamical system and the

measurement process has been evaluated by the probabilities inferred by Bayesian tracking. When the state transition and measurement models are known and treatable but the system is non-linear and/or non-Gaussian, exact inference is not feasible. The main achievements of the paper are upper and lower bounds to the information rate between the hidden state and the measurement that can be computed from approximate Bayesian tracking. The upper bound is based on filtering while the lower bound is based on smoothing. Also, the quality of the approximation to the wanted distributions obtained by approximated inference can be assessed from the bounds. Specifically, if the upper and lower bounds based on the inferred distributions are close to each other, then the inferred distributions are close to the true ones, while if this does not happen then the fit between the inferred distributions and the actual distributions is questionable. Application of the mentioned upper and lower bounds to the information rate transferred through channels with free-running hidden Markov state has been proposed, and specific results have been derived for the phase noise channel. These results show that, compared to the existing literature, our proposed approach allows to better deal with strong phase noise generated by a state space with high dimensionality. The picture is completed by numerical results that show application of our method to the Gauss-Markov fading channel.

## REFERENCES

- [1] D. Simon, *Optimal State Estimation*. New York, NY, USA: Wiley, 2006.
- [2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [3] P. M. Djurić *et al.*, "Particle filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, Sep. 2003.
- [4] J. V. Candy, "Bootstrap particle filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, Jul. 2007.
- [5] D. Creal, "A survey of sequential Monte-Carlo methods for economics and finance," *Econ. Rev.*, vol. 31, no. 3, pp. 245–296, Nov. 2011.
- [6] O. Hinka, F. Hlawatsch, and P. M. Djurić, "Distributed particle filtering in agent networks: A survey, classification, and comparison," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 61–81, Jan. 2013.
- [7] 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.
- [8] G. Colavolpe, T. Foggi, E. Forestieri, and M. Secondini, "Impact of phase noise and compensation techniques in coherent optical systems," *J. Lightw. Technol.*, vol. 29, no. 18, pp. 2790–2800, Sep. 15, 2011.
- [9] 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. Exp.*, vol. 19, no. 23, pp. 22455–22461, Nov. 2011.
- [10] P. Hou, B. J. Belzer, and T. R. Fischer, "Shaping gain of the partially coherent additive white Gaussian noise channel," *IEEE Commun. Lett.*, vol. 6, no. 5, pp. 175–177, May 2002.
- [11] A. Lapidoth, "On phase noise channels at high SNR," in *Proc. Inf. Theory Workshop*, Bangalore, India, Oct. 2002, pp. 1–4.
- [12] 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, Sep. 2011.
- [13] L. Barletta and G. Kramer, "On continuous-time white phase noise channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jul. 2014.
- [14] 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.
- [15] 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.

- [16] 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.
- [17] L. Barletta, M. Magarini, and A. Spalvieri, "The information rate transferred through the discrete-time Wiener's phase noise channel," *J. Lightw. Technol.*, vol. 30, no. 10, pp. 1480–1486, May 15, 2012.
- [18] H. Ghozlan and G. Kramer, "On Wiener phase noise channels at high signal-to-noise ratio," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2279–2283.
- [19] G. Durisi, A. Tarable, C. Camarda, and G. Montorsi, "On the capacity of MIMO Wiener phase-noise channels," in *Proc. Inf. Theory Appl. Workshop*, San Diego, CA, USA, Feb. 2013, pp. 1–7.
- [20] M. Martaló, C. Tripodi, and R. Raheli, "On the information rate of phase-noise limited communications," in *Proc. Inf. Theory Appl. Workshop*, San Diego, CA, USA, Feb. 2013, pp. 1–7.
- [21] H. Ghozlan and G. Kramer, "Multi-sample receivers increase information rates for Wiener phase noise channels," in *Proc. IEEE Globecom*, Atlanta, GA, USA, Dec. 2013.
- [22] A. Spalvieri and M. Magarini, "Wiener's analysis of the discrete-time phase-locked loop with loop delay," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 55, no. 6, pp. 596–600, Jun. 2008.
- [23] L. Barletta, M. Magarini, and A. Spalvieri, "Tight upper and lower bounds to the information rate of the phase noise channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 2284–2288.
- [24] R. H. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread MIMO fading channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1576–1608, Apr. 2006.
- [25] S. Venkatraman, K. Padmanabhan, and O. M. Collins, "The constrained capacity of single and multiuser channels with unknown fading," in *Proc. Inf. Theory Workshop*, Lake Tahoe, CA, USA, Sep. 2007, pp. 349–354.
- [26] G. Durisi, V. I. Morgenshtern, and H. Bölcskei, "On the sensitivity of continuous-time noncoherent fading channel capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6372–6391, Oct. 2012.
- [27] T. Koch and G. Kramer, "On noncoherent fading relay channels at high signal-to-noise ratio," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2221–2241, Apr. 2013.
- [28] 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. Exp.*, vol. 20, no. 23, pp. 25471–25477, Nov. 2012.
- [29] L. Barletta, M. Magarini, and A. Spalvieri, "New lower bound below the information rate of phase noise channel based on Kalman carrier recovery," *Int. J. Electr. Eng. Inf.*, vol. 4, no. 4, pp. 597–607, Dec. 2012.
- [30] 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.

**Luca Barletta** (M'13) received the M.S. degree (cum laude) in telecommunications engineering and the Ph.D. degree in information engineering from Politecnico di Milano, Milano, Italy, in 2007 and 2010, respectively.

He currently is a postdoc member at Technische Universität München - Institute for Advanced Study (TUM-IAS). He formerly was a postdoctoral researcher at Politecnico di Milano and visiting researcher at Bell Labs, Alcatel-Lucent, Holmdel, NJ, USA. His research interests include digital communications and information theory, with emphasis on fiber-optic communication, synchronization, and analysis of random access protocols.

**Maurizio Magarini** (M'04) was born in Milano, Italy, in 1969. He received the Master and Ph.D. degrees in electronic engineering from the Politecnico di Milano, Milano, Italy, in 1994 and 1999, respectively.

In 1994, he was granted the TELECOM Italia scholarship award for his Master thesis. From 1999 to 2001 he was a Research Associate in the Dipartimento di Elettronica, Informazione e Bioingegneria at the Politecnico di Milano where, since 2001, he has been an Assistant Professor. From August 2008 to January 2009 he spent a sabbatical leave at Bell Labs, Alcatel-Lucent, Holmdel, NJ. He has authored and coauthored more than 60 journal and conference papers. His research interests are in the broad area of communication theory. Topics include synchronization, channel estimation, equalization and coding applied to wireless and optical transmission systems.

**Simone Pecorino** received the M.S. degree (cum laude) in telecommunications engineering from Politecnico di Milano, Milano, Italy, in 2012.

He currently is a Ph.D. student in information technology at the Dipartimento di Elettronica, Informazione e Bioingegneria of Politecnico di Milano. His research interests include digital transmissions and information theory, with focus on Bayesian tracking techniques.

**Arnaldo Spalvieri** received his the degree in electronic engineering from the University of Ancona, Italy, in 1985.

From 1986 to 1992 he was in the modem laboratory of Telettra (now Alcatel-Lucent Italia), working at the research and development of modulation systems for high capacity digital radio systems. From 1992 to 1998 he was with the Dipartimento di Elettronica e Informazione at Politecnico di Milano as an Assistant Professor. In 1998 he was appointed Associate Professor at Politecnico di Milano, where he now holds the course Digital Communication. His research is focused on channel coding, channel equalization, synchronization. In 2006 he co-founded the company Binary Core, a spinoff of Politecnico di Milano active in the area of design on FPGA of mo-demodulation systems for point-to-point terrestrial radio and for digital video broadcasting.