Iterative Decoding in the Presence of Strong Phase Noise

Giulio Colavolpe, Alan Barbieri, and **Giuseppe Caire**

Università di Parma
Dipartimento di Ingegneria dell'Informazione
Parco Area delle Scienze 181/A
43100 Parma - ITALY
Tel: +39 0521 905744
Fax: +39 0521 905758
Email: giulio@unipr.it
alan.barbieri@studenti.unipr.it

Mobile Communications Group
Institut EURECOM
2229 Route des Cretes
06560 Valbonne, FRANCE
Tel: +33 (0)4 93002904
Fax: +33 (0)4 93002627
E-mail: giuseppe.caire@eurecom.fr

March 25, 2004

Abstract

We present some iterative decoding algorithms for channels affected by strong phase noise. The proposed algorithms are obtained as an application of the sum-product algorithm to the factor graph representing the joint a posteriori probability mass function of the information bits given the channel output. To overcome the problems due to the presence in the factor graph of continuous random variables, we advocate the method of *canonical distributions*. For several choices of canonical distributions, we derive the corresponding iterative decoding algorithms and compare their performance by computer simulation. We present numerical results for binary LDPC codes and LDPC-coded modulation, with particular reference to some phase-noise models and coded-modulation formats standardized in the next-generation satellite Digital Video Broadcasting (DVB-S2). Our results show that phase noise, with a rate of change typical of the instabilities of satellite transmitter and receiver oscillators, does not entail significant degradation with respect to the case of a perfectly coherent channel.

Index Terms

Factor graphs, Sum-product algorithm, Channels with memory, Phase-noise, Low-density parity-check codes, Iterative detection/decoding,

This work is funded by the European Space Agency, ESA-ESTEC, Noordwijk, The Netherlands, under contract no. 17337/NL/LvH.

The paper will be presented in part at the IEEE Symposium on Information Theory (ISIT'04), Chicago, IL, U.S.A., June-July 2004.

I. INTRODUCTION

The factor graph (FG) representation and the sum-product algorithm (SPA) [1] provide a general and powerful framework to reinterpret a number of well-known algorithms in digital communications, such as the Viterbi algorithm [2], the BCJR algorithm [3], the iterative "turbo" decoding algorithm [4], and the belief propagation algorithm for low-density parity-check (LDPC) codes [5].

In this paper, we use this framework to derive new efficient algorithms for iterative detection and decoding of channel codes transmitted over channels affected by phase noise. The approach is Bayesian, i.e., the unknown channel parameter is modeled as a stochastic process with known statistics. We construct the FG corresponding to the a posteriori joint probability distribution of the information message bits given the received signal and let the SPA compute the posterior marginal distributions. Bit-by-bit decisions are then made, based on the resulting posterior marginals. The FG includes the knowledge of the unknown parameter statistics. The average over the unknown parameter is implicitly performed by the SPA as part of the marginalization. The posterior marginal probabilities computed by the SPA are *exact* if the underlying FG is cycle-free. In this case, the bit-by-bit decision is optimal, i.e., it minimizes the average bit-error probability. More often, the underlying FG has cycles and the resulting SPA is inherently iterative. In this case, the SPA does not yield in general the optimal MAP decision rule. Nevertheless, the iterative SPA has proven to provide very good performance in several problems and therefore it can be regarded as a viable low-complexity solution when the optimal decision rule is just too complex to be implemented in practice. Since the resulting algorithms are naturally iterative, they are particularly suited to the decoding of codes such as LDPC and turbo codes, whose decoding algorithms are typically iterative (and suboptimal) even in the fully coherent setting (all channel parameters known).

Iterative decoding algorithms for channels with unknown phase has attracted an increasing interest in the recent literature. The algorithms developed in [6]–[10] are designed for noncoherent decoding of turbo codes and can be applied to LDPC codes only if trellis-based separate detection is performed. In particular, in [6], receivers for both the block-constant phase model and a discretized random-walk phase model are developed by using a phase discretization approach. In [11], the use of FGs that include both the code constraints and the channel parameter statistics is advocated in a very general setting. By specializing the approach of [11] to particular channel phase statistics, several algorithms for noncoherent detection/decoding have been proposed. In [12], [13] LDPC ensemble optimization via density evolution is considered for a very simplified block-constant phase model quantized over the two levels 0 and π. In [14] a constant and a random-walk phase noise model with Gaussian increments are considered and approximations of the SPA are derived and evaluated whereas in [15], messages in the SPA related to continuous random variables are approximated by means of an impulse at an estimated value and different estimation methods are considered. Finally, in [16] a phase model where the unknown carrier phase is constant over a block of N symbols and independent from block to block is considered, the channel parameters are not explicitly introduced in the FG, and the power allocation to the pilot symbols is optimized by using density evolution.

A non-Bayesian approach is adopted in [17]–[23]. In [17]–[22] the concept of *soft-decision-directed* estimation is introduced. The channel parameters are estimated by using the expectation-maximization (EM) algorithm [17]–[21] or an ad-hoc procedure [22] and the estimation algorithm is embedded into the iterative decoding process. Generally speaking, the non-Bayesian approaches consider the channel phase as a deterministic unknown constant. Tracking time-variations, such as in the case of the random-walk model, is allowed by using some heuristic sliding window adaptation. As a matter of fact, while the non-Bayesian schemes may be suited for the block-constant phase model, their performance degrades significantly in the presence of phase noise since the algorithms are not designed by exploiting the statistical knowledge of the phase time-variations.

In this paper we adopt the FG/SPA framework of [11] and we focus on the random-walk phase noise model with Gaussian increments. While the SPA is well-suited to handle discrete random variables, characterized by a probability mass function (pmf), the channel parameters are typically continuous random variables, characterized by a probability density function (pdf). The SPA for continuous random variables involves integration and computation of continuous pdfs, and it is not suited for direct implementation. A solution for this problem is suggested in [11] and consists of the use of *canonical distributions*, i.e., the pdfs computed by the SPA are constrained to be in a certain "canonical" family, characterized by some parameterization. Hence, the SPA has just to forward the parameters of the pdf rather than the pdf itself. We shall consider several canonical distributions, yielding different algorithms.

The most straightforward parameterization is based on the discretization of the parameter space [6]. Obviously, this approach becomes "optimal" (in the sense that it approaches the performance of the exact SPA) for a sufficiently large number of quantization levels, at the expenses of an increased computational complexity. For this reason, it will be considered as a performance benchmark. The other approaches proposed in this paper are based on Fourier, Tikhonov and Gaussian parameterizations, respectively. In the latter case, the derived algorithm is a modified version of the well-known Kalman smoother. Different pdf families lead to different performance and complexity, showing that the choice of the canonical distribution family is a non-trivial key step for obtaining efficient algorithms. We found that the proposed Tikhonov parameterization yields to an algorithm with unprecedented performance/complexity tradeoff.

The reminder of this paper is organized as follows. Section II introduces the channel model and derives the basic FG and SPA. In Section III, we briefly recall the discretization method and present the details of the new algorithms. In Section IV we present numerical results to compare the proposed algorithms in the context of LDPC decoding, under the random-walk phase noise model and under a European Space Agency (ESA) phase noise model used to evaluate the next-generation Digital Video Broadcasting satellite standard (DVB-S2) [24], [25]. Finally, Section V points out some conclusive remarks.

II. SYSTEM MODEL AND EXACT SUM-PRODUCT ALGORITHM

We consider the transmission of a sequence of complex modulation symbols $\mathbf{c} = (c_0, c_1, \ldots, c_{K-1})$ over an additive white Gaussian noise (AWGN) channel affected by carrier phase noise. Symbols c_k are linearly modulated. Assuming Nyquist transmitted pulses, matched filtering, and phase variations slow enough so as no intersymbol interference arises, the discrete-time baseband complex equivalent channel model at the receiver is given by

$$r_k = c_k e^{j\theta_k} + n_k, \quad k = 0, \ldots, K-1\,. \tag{1}$$

We assume that the sequence $\mathbf{c}$ is a codeword of the channel code $\mathcal{C}$ constructed over an M-ary modulation constellation $\mathcal{X} \subset \mathbb{C}$. We include possible pilot symbols (known to the receiver) and/or possible differential encoding as a part of the code $\mathcal{C}$. The vector of noise samples $\mathbf{n} = (n_0, n_1, \ldots, n_{K-1})$ has i.i.d., complex circularly symmetric components, with $n_k \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma^2)$.[1] The vector of channel phases $\boldsymbol{\theta} = (\theta_0, \theta_1, \ldots, \theta_{K-1})$ is random, unknown to both transmitter and receiver, and statistically independent of $\mathbf{c}$ and $\mathbf{n}$.

A common model for the phase noise process $\{\theta_k\}$ is the random-walk (Wiener) model described by

$$\theta_k = \theta_{k-1} + \Delta_k \tag{2}$$

where $\{\Delta_k\}$ is a white real Gaussian process with $\Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2)$. Under this assumption and assuming $\theta_0 \sim \text{Uniform}[0, 2\pi)$, it follows that

$$p(\theta_k|\theta_{k-1}, \theta_{k-2}, \ldots, \theta_0) = p(\theta_k|\theta_{k-1}) = p_\Delta(\theta_k - \theta_{k-1}) \tag{3}$$

[1] A complex circularly symmetric (resp., real) Gaussian random vector $\mathbf{v}$ with mean $\mathbf{m}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathbf{v} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{m}, \boldsymbol{\Sigma})$ (resp., by $\mathbf{v} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$). We denote the multivariate complex circularly symmetric (resp. real) Gaussian pdf with mean $\mathbf{m}$, covariance matrix $\boldsymbol{\Sigma}$ and argument $\mathbf{x}$ by $g_{\mathbb{C}}(\mathbf{m}, \boldsymbol{\Sigma}, \mathbf{x})$ (resp., by $g(\mathbf{m}, \boldsymbol{\Sigma}, \mathbf{x})$).

where we define $p_\Delta(\phi)$ as the pdf of the increment $\Delta_k \bmod [0, 2\pi)$, i.e.,

$$p_\Delta(\phi) \triangleq \begin{cases} \sum_{\ell=-\infty}^{\infty} g\left(0, \sigma_\Delta^2, \phi - \ell 2\pi\right) & \phi \in [0, 2\pi) \\ 0 & \text{elsewhere.} \end{cases} \tag{4}$$

The Wiener phase noise model will be considered in the following as a working assumption in order to derive efficient iterative detection and decoding algorithms. This assumption will be relaxed in Section IV, where we apply our algorithms to the DVB-S2-compliant ESA model described in [24], [25].

Without loss of generality, we assume that the code $\mathcal{C}$ admits an encoding function $\mu_\mathcal{C} : \mathbb{F}_2^B \to \mathcal{X}^K$, mapping binary information messages $\mathbf{b} \in \mathbb{F}_2^B$ into the codewords. The optimal decision rule that minimizes the average bit-error probability is given by

$$\widehat{b}_i = \arg \max_{b \in \mathbb{F}_2} P_{\text{bit},i}(b|\mathbf{r}) \tag{5}$$

where $P_{\text{bit,i}}(b|\mathbf{r})$ denotes the a posteriori pmf for the i-th information bit given the received signal vector $\mathbf{r} = (r_0, \ldots, r_{K-1})$. Let $P(\mathbf{b}, \boldsymbol{\theta}|\mathbf{r})$ denote the joint posterior probability distribution function[2] of the information bits and of the phase noise vector $\boldsymbol{\theta}$ given $\mathbf{r}$. Clearly, the desired $P_{\text{bit},i}(b|\mathbf{r})$ can be obtained by marginalizing $P(\mathbf{b}, \boldsymbol{\theta}|\mathbf{r})$ with respect to $\boldsymbol{\theta}$ and to all b_j for $j \neq i$. This can be accomplished in an approximated but low-complexity way by the SPA applied on the FG of $P(\mathbf{b}, \boldsymbol{\theta}|\mathbf{r})$, as illustrated in the following.

We assume that the reader is familiar with the FG/SPA framework (that can be found, for example, in the excellent tutorial paper [1]). Therefore, for the sake of space limitation we will not recall here this background. From the definition of the encoding function $\mu_\mathcal{C}$ and the

[2]We use the term probability distribution function to denote a continuous pdf with some discrete probability masses. For a probability distribution function we still use the symbol $P(.)$.

channel model (1) we obtain the factorization[3]

$$
\begin{aligned}
P(\mathbf{b},\boldsymbol{\theta}|\mathbf{r}) &\propto P(\mathbf{b})p(\boldsymbol{\theta})p(\mathbf{r}|\boldsymbol{\theta},\mathbf{b}) \\
&\propto \chi[\mathbf{c}=\mu_{\mathcal{C}}(\mathbf{b})]p(\boldsymbol{\theta})p(\mathbf{r}|\boldsymbol{\theta},\mathbf{c}=\mu_{\mathcal{C}}(\mathbf{b})) \\
&\propto \chi[\mathbf{c}=\mu_{\mathcal{C}}(\mathbf{b})]p(\boldsymbol{\theta})\prod_{k=0}^{K-1} p(r_k|c_k,\theta_k) \\
&\propto \chi[\mathbf{c}=\mu_{\mathcal{C}}(\mathbf{b})]p(\boldsymbol{\theta})\prod_{k=0}^{K-1} f_k(c_k,\theta_k)
\end{aligned} \tag{6}
$$

where we have used the fact that the output signal pdf $p(\mathbf{r})$ does not depend on $\mathbf{b}$, that the information bits are uniform and i.i.d., therefore $P(\mathbf{b}) = 2^{-B}$, that the AWGN channel for given $\boldsymbol{\theta}$ is memoryless, and we have defined the functions

$$
f_k(c_k,\theta_k) \triangleq \exp\left\{\frac{1}{\sigma^2}\mathrm{Re}[r_k c_k^* e^{-j\theta_k}] - \frac{|c_k|^2}{2\sigma^2}\right\} \propto \exp\left\{-\frac{1}{2\sigma^2}|r_k - c_k e^{j\theta_k}|^2\right\}. \tag{7}
$$

and the code indicator function $\chi[\mathbf{c}=\mu_{\mathcal{C}}(\mathbf{b})]$, equal to 1 if $\mathbf{c}$ is the codeword corresponding to $\mathbf{b}$ and to zero otherwise. The FG corresponding to (6) is shown in Fig. 1.

Under the assumption of 1st order Markov model for the phase noise, we can further factor the term $p(\boldsymbol{\theta})$ as $p(\boldsymbol{\theta}) = p(\theta_0)\prod_{k=1}^{K} p_\Delta(\theta_k - \theta_{k-1})$ obtaining

$$
P(\mathbf{b},\boldsymbol{\theta}|\mathbf{r}) \propto \chi[\mathbf{c}=\mu_{\mathcal{C}}(\mathbf{b})]p(\theta_0)\prod_{k=1}^{K-1} p_\Delta(\theta_k - \theta_{k-1})\prod_{k=0}^{K-1} f_k(c_k,\theta_k). \tag{8}
$$

The corresponding FG is sketched in Fig. 2 and represents the starting point for the development of the proposed algorithms.

The SPA applied to the FG in the upper box, corresponding to the code constraints, consists of the well-known standard belief propagation whose efficient implementation depends on the structure of the code $\mathcal{C}$ and needs no details here. Hence, we shall concentrate on the SPA message propagation in the lower part of the graph. Omitting for simplicity of notation the explicit reference to the current iteration, let us denote by $P_d(c_k)$ the message from variable

[3]In this paper, we use extensively the proportionality relationship $f \propto g$, indicating that $f = ag$ for some real constant a, since the SPA is defined up to scaling its messages by positive factors, independent of the variables represented in the graph.

node c_k to factor node f_k, and by $P_u(c_k)$ the message in the opposite direction (see Fig. 2). The message $p_d(\theta_k)$ from factor node f_k to variable node θ_k can be expressed as

$$p_d(\theta_k) \propto \sum_{x \in \mathcal{X}} P_d(c_k = x) f_k(c_k = x, \theta_k)\,. \tag{9}$$

We also assume that in the lower part of the FG, describing the phase-noise evolution, a forward-backward node activation schedule is adopted. Therefore, messages $p_f(\theta_k)$ from factor node $p_\Delta(\theta_k - \theta_{k-1})$ to variable node θ_k, and $p_b(\theta_k)$ from factor node $p_\Delta(\theta_{k+1} - \theta_k)$ to variable node θ_k, can be recursively computed as follows:

$$p_f(\theta_k) \propto \int_0^{2\pi} p_d(\theta_{k-1}) p_f(\theta_{k-1}) p_\Delta(\theta_k - \theta_{k-1})\, d\theta_{k-1} \tag{10}$$

$$p_b(\theta_k) \propto \int_0^{2\pi} p_d(\theta_{k+1}) p_b(\theta_{k+1}) p_\Delta(\theta_{k+1} - \theta_k)\, d\theta_{k+1} \tag{11}$$

with uniform pdfs as initial conditions:

$$p_f(\theta_0) = p_b(\theta_K) = \frac{1}{2\pi}\,. \tag{12}$$

The message $P_u(c_k)$ from f_k to c_k is given by

$$P_u(c_k) \propto \int_0^{2\pi} p_f(\theta_k) p_b(\theta_k) f_k(c_k, \theta_k)\, d\theta_k\,. \tag{13}$$

The vector of messages $\{P_u(c_k) : k = 0, \ldots, K-1\}$ represents the observation (in the form of sequence of a posteriori pmfs) of the coded symbols “seen” through a virtual memoryless channel, and are processed by the upper part of the graph according to the standard belief propagation algorithm. At each iteration, this produces updated messages $\{P_d(c_k) : k = 0, \ldots, K-1\}$ and updated estimates of the a posteriori probabilities $P_{\text{bit},i}$.

Equations (10), (11), and (13), form the main part of the SPA for iterative detection and decoding in the presence of phase noise.

III. Proposed Algorithms

It is clear that the implementation complexity of the exact SPA is impractical, since the messages from and to the variable nodes $\{\theta_k\}$ are continuous pdfs. In order to obtain practical algorithms, we follow the canonical distribution approach proposed in [11].

A. Discretization of the channel parameters

This case corresponds to letting the canonical distribution be a weighted sum of impulses. This approach has been adopted for Viterbi- and BCJR-like receivers in [26] and [6], [15], respectively. We assume that the channel phase θ_k may take on the following L values: $\Theta = \{0, 2\pi/L, \ldots, 2\pi(L-1)/L\}$.[4] Obviously, this approach becomes "optimal" (in the sense that it approaches the performance of the exact SPA) for a sufficiently large number of discretization levels, at the expenses of an increasing computational complexity.

B. Fourier Parameterization

The function $f_k(c_k, \theta_k)$ defined in (7) is periodic in θ_k. Hence, it can be expanded in Fourier series. We use the well-known identity [27, eqn. (9.6.34)]

$$e^{x\cos\theta} = \mathrm{I}_0(x) + 2\sum_{\ell=1}^{\infty} \mathrm{I}_\ell(x)\cos(\ell\theta) \tag{14}$$

where $\mathrm{I}_\ell(x)$ is the modified Bessel function of the first kind of order ℓ. Letting, for a complex number z, $\phi(z) = \arg(z)$, after some straightforward manipulations we obtain

$$f_k(c_k, \theta_k) \propto e^{-\frac{|c_k|^2}{2\sigma^2}} \sum_{\ell=-\infty}^{\infty} \mathrm{I}_\ell\left(\frac{|r_k||c_k|}{\sigma^2}\right) e^{-j\ell\phi(r_k c_k^*)} e^{j\ell\theta_k}\,. \tag{15}$$

Substituting (15) into eqn. (9), we may express

$$p_d(\theta_k) \propto \sum_{\ell=-\infty}^{\infty} A_k^{(\ell)} e^{j\ell\theta_k} \tag{16}$$

having defined

$$\begin{aligned} A_k^{(\ell)} &\triangleq \sum_{x\in\mathcal{X}} P_d(c_k = x) e^{-\frac{|x|^2}{2\sigma^2}} \mathrm{I}_\ell\left(\frac{|r_k||x|}{\sigma^2}\right) e^{-j\ell\phi(r_k x^*)} \\ &= e^{-j\ell\phi(r_k)} \sum_{x\in\mathcal{X}} P_d(c_k = x) e^{-\frac{|x|^2}{2\sigma^2}} \mathrm{I}_\ell\left(\frac{|r_k||x|}{\sigma^2}\right) e^{j\ell\phi(x)} \\ &= e^{-j\ell\phi(r_k)} \sum_{x\in\mathcal{X}} P_d(c_k = x) e^{-\frac{|x|^2}{2\sigma^2}} \mathrm{I}_\ell\left(\frac{|r_k||x|}{\sigma^2}\right) \frac{x^\ell}{|x|^\ell}\,. \end{aligned} \tag{17}$$

[4] In [6], the authors state that for M-PSK signals, $L = 8M$ values are sufficient to have no performance loss.

Note that for M-PSK signals, the expression of coefficients $A_k^{(\ell)}$, neglecting irrelevant terms, simplifies to

$$A_k^{(\ell)} = e^{-j\ell\phi(r_k)} \mathrm{I}_\ell\left(\frac{|r_k|}{\sigma^2}\right) \sum_{x\in\mathcal{X}} P_d(c_k = x) x^\ell \,. \tag{18}$$

In this case, at the first iteration, when the probabilities of symbols $P_d(c_k)$ are all equal to $1/M$ (except for the pilot symbols), these coefficients are zero for $\ell \neq 0, \pm M, \pm 2M, \pm 3M, \ldots$.

Pdfs $p_f(\theta_k)$ and $p_b(\theta_k)$ take on the same form, i.e., they are periodic as well and can be expanded in Fourier series as

$$\begin{aligned} p_f(\theta_k) &= \sum_{\ell=-\infty}^{\infty} B_{f,k}^{(\ell)} e^{j\ell\theta_k} && (19) \\ p_b(\theta_k) &= \sum_{\ell=-\infty}^{\infty} B_{b,k}^{(\ell)} e^{j\ell\theta_k} \,. && (20) \end{aligned}$$

Substituting (16) and (19) into eqn. (10), we obtain

$$\begin{aligned} \sum_{\ell=-\infty}^{\infty} B_{f,k}^{(\ell)} e^{j\ell\theta_k} &= \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} A_{k-1}^{(m)} B_{f,k-1}^{(n)} \int_0^{2\pi} e^{j(m+n)\theta_{k-1}} p_\Delta(\theta_k - \theta_{k-1})\, d\theta_{k-1} \\ &= \sum_{\ell=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} A_{k-1}^{(m)} B_{f,k-1}^{(\ell-m)} \int_0^{2\pi} e^{j\ell\theta_{k-1}} p_\Delta(\theta_k - \theta_{k-1})\, d\theta_{k-1} \,. \end{aligned} \tag{21}$$

We notice that, for practical values of σ_Δ, the pdf $p_\Delta(\phi)$ is essentially zero for argument ϕ outside an interval centered in 0 of size much smaller than 2π. Hence, we can write

$$\begin{aligned} \int_0^{2\pi} e^{j\ell\theta_{k-1}} p_\Delta(\theta_k - \theta_{k-1})\, d\theta_{k-1} &\simeq \int_{-\infty}^{\infty} e^{j\ell\theta_{k-1}} g(0, \sigma_\Delta^2, \theta_k - \theta_{k-1})\, d\theta_{k-1} \\ &= D_\ell(\sigma_\Delta) e^{j\ell\theta_k} \end{aligned} \tag{22}$$

where we define

$$D_\ell(\sigma_\Delta) \triangleq e^{-\frac{\sigma_\Delta^2 \ell^2}{2}} \,. \tag{23}$$

By using (22) in (21) we obtain

$$\sum_{\ell=-\infty}^{\infty} B_{f,k}^{(\ell)} e^{j\ell\theta_k} = \sum_{\ell=-\infty}^{\infty} \left[D_\ell(\sigma_\Delta) \sum_{m=-\infty}^{\infty} A_{k-1}^{(m)} B_{f,k-1}^{(\ell-m)} \right] e^{j\ell\theta_k} \tag{24}$$

yielding the forward recursion for the Fourier coefficients $B_{f,k}^{(\ell)}$:

$$B_{f,k}^{(\ell)} = D_\ell(\sigma_\Delta) \sum_{m=-\infty}^{\infty} A_{k-1}^{(m)} B_{f,k-1}^{(\ell-m)} = D_\ell(\sigma_\Delta)[A_{k-1}^{(\ell)} \otimes B_{f,k-1}^{(\ell)}] \tag{25}$$

where $\otimes$ denotes convolution of sequences. From condition (12), we derive the initial condition

$$B_{f,0}^{(\ell)} = \delta(\ell) \tag{26}$$

where $\delta(\ell)$ denotes the Kronecker delta. Similarly, the backward recursion to compute the coefficients $\{B_{b,k}^{(\ell)}\}$ is given by

$$B_{b,k}^{(\ell)} = D_\ell(\sigma_\Delta)[A_{k+1}^{(\ell)} \otimes B_{b,k+1}^{(\ell)}] \tag{27}$$

with initial condition

$$B_{b,K-1}^{(\ell)} = \delta(\ell)\,. \tag{28}$$

Note that the computation of these coefficients can be simplified taking into account the symmetries $A_k^{(-\ell)} = A_k^{(\ell)*}$, $B_{f,k}^{(-\ell)} = B_{f,k}^{(\ell)*}$, and $B_{b,k}^{(-\ell)} = B_{b,k}^{(\ell)*}$. Finally, substituting (15), (19), and (20) into eqn. (13) and defining

$$E_k^{(\ell)} \triangleq e^{-\frac{|c_k|^2}{2\sigma^2}} \left\{ B_{f,k}^{(\ell)} \otimes B_{b,k}^{(\ell)} \otimes \left[\mathrm{I}_\ell \left(\frac{|r_k||c_k|}{\sigma^2} \right) e^{-j\ell\phi(r_k c_k^*)} \right] \right\} \tag{29}$$

we have

$$P_u(c_k) \propto \sum_{\ell=-\infty}^{\infty} E_k^{(\ell)} \int_0^{2\pi} e^{j\ell\theta_k}\, d\theta_k = E_k^{(0)}\,. \tag{30}$$

Remark: Truncation of the Fourier coefficients. The convolution of the infinite-duration Fourier coefficients can be effectively implemented by truncation. Hence, a reduced number N of coefficients must be taken into account due to the fact that, for a given x, functions $\mathrm{I}_\ell(x)$ are monotonically decreasing for increasing values of ℓ. Standard smoothed truncation methods (windowing) can be applied [28]. In particular, by means of computer simulations, we found that the Kaiser window with an optimized parameter β [28] yields good results, as it will be demonstrated in Section IV.

C. Tikhonov Parameterization

Let us consider eqn. (9). If the messages $P_d(c_k)$ were the exact a posteriori probabilities of the code symbols, it would be

$$p_d(\theta_k) \propto \sum_{x\in\mathcal{X}} P_d(c_k = x) f_k(c_k = x, \theta_k) \propto p(r_k|\theta_k)\,. \tag{31}$$

We use Fact 1 in Appendix A and approximate $p(r_k|\theta_k)$ by the Gaussian pdf at minimum *divergence*, given by $g_{\mathbb{C}}(\alpha_k e^{j\theta_k}, 2\sigma^2 + \beta_k - |\alpha_k|^2, r_k)$, where α_k and β_k are the first and second-order moments of $c_k \sim P_d(c_k)$, given by

$$\alpha_k \triangleq \sum_{x\in\mathcal{X}} x P_d(c_k = x) \tag{32}$$

$$\beta_k \triangleq \sum_{x\in\mathcal{X}} |x|^2 P_d(c_k = x)\,. \tag{33}$$

Under the above min-divergence Gaussian approximation, we obtain

$$\begin{aligned} p_d(\theta_k) &\propto p(r_k|\theta_k) \\ &\simeq g_{\mathbb{C}}\left(\alpha_k e^{j\theta_k}, 2\sigma^2 + \beta_k - |\alpha_k|^2, r_k\right) \\ &\propto \exp\left\{2\frac{\text{Re}[r_k \alpha_k^* e^{-j\theta_k}]}{2\sigma^2 + \beta_k - |\alpha_k|^2}\right\}\,. \end{aligned} \tag{34}$$

Substituting (34) in the forward recursion (10), we obtain

$$p_f(\theta_k) \simeq \int_0^{2\pi} \exp\left\{2\frac{\text{Re}[r_{k-1}\alpha_{k-1}^* e^{-j\theta_{k-1}}]}{2\sigma^2 + \beta_{k-1} - |\alpha_{k-1}|^2}\right\} p_f(\theta_{k-1}) p_\Delta(\theta_k - \theta_{k-1})\, d\theta_{k-1}\,. \tag{35}$$

When the channel phase is slowly-varying, i.e., for $\sigma_\Delta \to 0$, we have $p_\Delta(\theta_k - \theta_{k-1}) = \delta(\theta_k - \theta_{k-1})$. In this case, the solution of the recursion given by eqn. (35) with initial condition (12) is a sequence of Tikhonov pdfs, given by

$$p_f(\theta_k) \propto \exp\left\{\text{Re}[a_{f,k} e^{-j\theta_k}]\right\} \tag{36}$$

where $a_{f,k}$ can be recursively computed as

$$a_{f,k} = a_{f,k-1} + 2\frac{r_{k-1}\alpha_{k-1}^*}{2\sigma^2 + \beta_{k-1} - |\alpha_{k-1}|^2} \tag{37}$$

with the initial condition $a_{f,0} = 0$. Similarly, the solution of the backward recursion (11) under the above approximations is the sequence of Tikhonov pdfs

$$p_b(\theta_k) \propto \exp\left\{\mathrm{Re}[a_{b,k} e^{-j\theta_k}]\right\} \tag{38}$$

where $a_{b,k}$ can be recursively computed as

$$a_{b,k} = a_{b,k+1} + 2\frac{r_{k+1}\alpha^*_{k+1}}{2\sigma^2 + \beta_{k+1} - |\alpha_{k+1}|^2} \tag{39}$$

with the initial condition $a_{b,K-1} = 0$. From (36), (38) and (13) we obtain

$$P_u(c_k) \propto \exp\left\{-\frac{|c_k|^2}{2\sigma^2}\right\} \mathrm{I}_0\left(\left|a_{f,k} + a_{b,k} + \frac{r_k c^*_k}{\sigma^2}\right|\right) . \tag{40}$$

When the phase is rapidly-varying, the approximation $p_\Delta(\theta_k - \theta_{k-1}) \simeq \delta(\theta_k - \theta_{k-1})$ does not hold any longer. However, we found that good approximations of functions $p_f(\theta_k)$ and $p_b(\theta_k)$ are still given in the form (36) and (38) where the coefficients $a_{f,k}$ and $a_{b,k}$ are updated by the modified forward and backward recursions

$$a_{f,k} = \left[a_{f,k-1} + 2\frac{r_{k-1}\alpha^*_{k-1}}{2\sigma^2 + \beta_{k-1} - |\alpha_{k-1}|^2}\right] \cdot \gamma\left(\sigma^2_\Delta, \left|a_{f,k-1} + 2\frac{r_{k-1}\alpha^*_{k-1}}{2\sigma^2 + \beta_{k-1} - |\alpha_{k-1}|^2}\right|\right) \tag{41}$$

$$a_{b,k} = \left[a_{b,k+1} + 2\frac{r_{k+1}\alpha^*_{k+1}}{2\sigma^2 + \beta_{k+1} - |\alpha_{k+1}|^2}\right] \cdot \gamma\left(\sigma^2_\Delta, \left|a_{f,k+1} + 2\frac{r_{k+1}\alpha^*_{k+1}}{2\sigma^2 + \beta_{k+1} - |\alpha_{k+1}|^2}\right|\right) . \tag{42}$$

The real function $\gamma(x_1, x_2)$, of real arguments x_1 and x_2 can be numerically computed and stored in a lookup table. The motivation of (41) and (42) and a closed-form approximated expression of the correction factor γ is provided in Appendix B.

Remark: Modification in the case of long pilot fields. When the pilot symbols are arranged in bursts (training sequences) separated by long blocks of code symbols, as in the case of the DVB-S2 system [29], it is necessary to slightly modify the algorithm in order to speed-up the convergence process and to avoid the risk of a phase ambiguity. In fact, consider the recursive integral equation (10) from the second iteration on. If the product

$$p_d(\theta_k)p_f(\theta_k) = \left(\sum_{x\in\mathcal{X}} P_d(c_k = x) f_k(c_k = x, \theta_k)\right) p_f(\theta_k)$$

contains a dominant exponential term, i.e., if there exists a value $\overline{x} \in \mathcal{X}$ such that

$$\ln P_d(c_k = \overline{x}) + \left| a_{f,k} + \frac{r_k \overline{x}^*}{\sigma^2} \right| > \delta + \ln P_d(c_k = x) + \left| a_{f,k} + \frac{r_k x^*}{\sigma^2} \right| , \forall x \in \mathcal{X}, x \neq \overline{x} \quad (43)$$

where δ is a real parameter to be optimized by computer simulation, it is preferable to let $\alpha_k = \overline{x}$ and $\beta_k = |\overline{x}|^2$. Otherwise, we choose α_k and β_k as in (32) and (33). This corresponds to using a decision-aided scheme based on hard decisions for some symbols c_k. Similar considerations also hold for the recursive integral equation (11). In the numerical results related to the DVB-S2 system, we found that $\delta = 1.5$ yields satisfactory results.

D. Gaussian parameterization

Another exemplification of the canonical distribution approach consists of modeling the phasor process $h_k \triangleq e^{j\theta_k}$ as a complex circularly symmetric Gauss-Markov process and treating $\mathbf{h} = (h_0, \ldots, h_{K-1})$ and $\mathbf{r}$ as jointly Gaussian. This assumption yields the forward and backward recursions (10) and (11) in the form of a Kalman smoother (non-causal linear MMSE estimator).

As for the Tikhonov parameterization of the previous section, we impose a jointly Gaussian structure on the observation $\{r_k\}$ and the phasor process $\{h_k\}$ by using the minimum Divergence approximation of the pdf of r_k given h_k, i.e., we let $p(r_k|h_k) \simeq g_{\mathbb{C}}\left(\alpha_k h_k, 2\sigma^2 + \beta_k - |\alpha_k|^2, r_k\right)$ where we have used Fact 1 of Appendix A, and where the conditional mean and variance of r_k given h_k are given by $E[r_k|h_k] = \alpha_k h_k$ and by $\mathrm{Var}(r_k|h_k) = 2\sigma^2 + \beta_k - |\alpha_k|^2$, with α_k and β_k given in (32) and in (33), respectively.

Under the Gauss-Markov assumption for $\{h_k\}$ and the above joint Gaussianity, we can define the "state" and "observation" equations by

$$h_{k+1} = \rho h_k + v_k \quad (44)$$

$$r_k = \alpha_k h_k + w_k \quad (45)$$

where $\{v_k\}$ and $\{w_k\}$ are independent Gaussian processes with independent components such that $v_k \sim \mathcal{N}_{\mathbb{C}}(0, 1-\rho^2)$ and $w_k \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma^2 + \beta_k - |\alpha_k|^2)$. For the Wiener phase noise model, we obtain explicitly $\rho = e^{-\sigma_\Delta^2/2}$. The time-reversal process $(h_{K-1}, \ldots, h_1, h_0)$ is also Gauss-Markov [30] with state and observation equations given by $h_{k-1} = \rho h_k + v'_k$ and by $r_k = \alpha_k h_k + w'_k$, respectively, where $\{v'_k\}$ and $\{w'_k\}$ have the same statistics of $\{v_k\}$ and $\{w_k\}$.

Under this model, we have $p(h_k | \{r_j : j \neq k\}) = g_{\mathbb{C}}(m_k, \Sigma_k, h_k)$, where the conditional mean and variance can be computed iteratively using the Kalman smoother [30], [31]. The derivation of the Kalman filter via the SPA is given in [1]. The forward and backward recursions (10) and (11) are evaluated explicitly by a repeated application of Facts 2 and 3 in Appendix A (details are omitted for the sake of brevity). Let $m_{k|k-1}, \Sigma_{k|k-1}$ be the conditional mean and variance of h_k given $\{r_j : j = 0, \ldots, k-1\}$ (prediction) and $m_{k|k}$ and $\Sigma_{k|k}$ be the conditional mean and variance of h_k given $\{r_j : j = 0, \ldots, k\}$ (filtering). Similarly, Let $\mu_{k|k+1}, \Xi_{k|k-1}$ be the conditional mean and variance of h_k given $\{r_j : j = k+1, \ldots, K-1\}$ (backward prediction) and $\mu_{k|k}$ and $\Xi_{k|k}$ be the conditional mean and variance of h_k given $\{r_j : j = k, \ldots, K-1\}$ (backward filtering). The resulting recursions are given by

$$\begin{aligned}
m_{k|k} &= m_{k|k-1} + \frac{\Sigma_{k|k-1}\alpha_k^*}{|\alpha_k|^2(\Sigma_{k|k-1} - 1) + 2\sigma^2 + \beta_k}\left(r_k - \alpha_k m_{k|k-1}\right) \\
\Sigma_{k|k} &= \frac{2\sigma^2 + \beta_k - |\alpha_k|^2}{|\alpha_k|^2(\Sigma_{k|k-1} - 1) + 2\sigma^2 + \beta_k}\Sigma_{k|k-1} \\
m_{k+1|k} &= \rho m_{k|k} \\
\Sigma_{k+1|k} &= \rho^2 \Sigma_{k|k} + 1 - \rho^2
\end{aligned} \tag{46}$$

for $k = 0, \ldots, K-1$, with initial conditions $\Sigma_{0|-1} = 1$ and $m_{0|-1} = 0$, and by

$$\begin{aligned}
\mu_{k|k} &= \mu_{k|k+1} + \frac{\Xi_{k|k+1}\alpha_k^*}{|\alpha_k|^2(\Xi_{k|k+1} - 1) + 2\sigma^2 + \beta_k}\left(r_k - \alpha_k \mu_{k|k+1}\right) \\
\Xi_{k|k} &= \frac{2\sigma^2 + \beta_k - |\alpha_k|^2}{|\alpha_k|^2(\Xi_{k|k+1} - 1) + 2\sigma^2 + \beta_k}\Xi_{k|k+1} \\
\mu_{k-1|k} &= \rho \mu_{k|k} \\
\Xi_{k-1|k} &= \rho^2 \Xi_{k|k} + 1 - \rho^2
\end{aligned} \tag{47}$$

for $k = K-1, \ldots, 0$, with initial conditions $\Xi_{K-1|K} = 1$ and $\mu_{K-1|K} = 0$.

Finally, for each k we obtain

$$\begin{aligned} m_k &= \frac{\Xi_{k|k+1}}{\Sigma_{k|k-1}+\Xi_{k|k+1}} m_{k|k-1} + \frac{\Sigma_{k|k-1}}{\Sigma_{k|k-1}+\Xi_{k|k+1}} \mu_{k|k+1} \\ \Sigma_k &= \frac{\Sigma_{k|k-1}\Xi_{k|k+1}}{\Sigma_{k|k-1}+\Xi_{k|k+1}} . \end{aligned} \tag{48}$$

It remains to find an expression for the message $P_u(c_k)$, that is, the probability of the code symbol c_k given the observation r_k and the phasor estimate $h_k \sim \mathcal{N}_{\mathbb{C}}(m_k, \Sigma_k)$. We let $h_k = R_k e^{j\theta_k}$ and, by using Facts 4 and 5 in Appendix A, after some manipulations, we obtain

$$\begin{aligned} P_u(c_k) &\propto \int \exp\left(-\frac{1}{2\sigma^2}|r_k - c_k e^{j\theta_k}|^2\right) g_{\mathbb{C}}\left(m_k, \Sigma_k, R_k e^{j\theta_k}\right) \; R_k dR_k d\theta_k \\ &\propto e^{-\frac{|c_k|^2}{2\sigma^2}} \int_0^\infty e^{-z} I_0\left(\sqrt{\frac{|r_k c^*|^2}{\sigma^4} + 4\frac{z|m_k|^2}{\Sigma_k} + 4\frac{\sqrt{z}}{\sqrt{\Sigma_k}\sigma^2}\mathrm{Re}\{r_k c^* m_k^*\}}\right) dz . \end{aligned} \tag{49}$$

The above integral can be easily computed by using Gauss-Laguerre quadrature rules.

IV. Numerical Results

In this Section, the performance of the proposed schemes is assessed by computer simulations in terms of bit error rate (BER) versus E_b/N_0, E_b being the received signal energy per information bit and N_0 the one-sided noise power spectral density. Unless otherwise stated, the considered code is a (3,6)-regular LDPC code with codewords of length 4000 [32], a maximum of 200 iterations of the SPA on the overall graph is allowed, and the BPSK modulation is adopted. Pilot symbols are inserted in the transmitted codeword in order to make the iterative decoding algorithms bootstrap. This corresponds to a decrease in the effective transmission rate, resulting in an increase in the required signal-to-noise ratio which has been introduced artificially in the curve labeled "known phase" for the sake of comparison. Hence, the gap between the "known phase" curve and the others is uniquely due to the need for phase estimation/compensation, and not to the rate decrease due to the insertion of pilot symbols.

In Fig. 3, the algorithms described in Sections III-A and III-B (discretization of channel parameters and Fourier parameterization) have been considered, assuming a pilot symbol in every block of 20 transmitted symbols. The ESA phase noise model and a more severe Wiener model (2) with $\sigma_\Delta = 6$ degrees, have been considered. This latter case has been used to stress the robustness of the described schemes to a strong phase noise and to select the best algorithm, from a performance-complexity trade-off point of view, to be used for high-order modulations. In the case of the ESA model, all the receivers were designed by assuming a Wiener phase noise model with $\sigma_\Delta = 0.3$ degrees.

In the case of the Wiener model, different values L of discretization levels and different values of the number N of considered Fourier coefficients have been considered. No improvement has been observed for values of $L > 16$ and this is in agreement with a result in [6]. Similarly, values of $N > 17$ are not considered since they do not produce any performance improvement. Therefore, the value of $N = 17$ (i.e., $-8 \leq \ell \leq 8$ in all the equations of Section III-B) can be considered as optimal for $\sigma_\Delta = 6$ degrees. Hence, the gap of about 0.2 dB with respect to the curve labeled "known phase" is only due to the loss in channel capacity for a time-varying channel phase.

In the binary case considered in the previous figure, the proposed algorithms have a practically optimal performance and a similar complexity. However, for a modulation format characterized by a more dense constellation, if for the discretization-based algorithm the optimal number of discretization levels, and thus the complexity, must be increased, it can be expected that the number N of considered Fourier coefficients in the proposed algorithm remains practically the same. This aspect is shown in Fig. 4 where a QPSK modulation is considered. The phase noise has $\sigma_\Delta = 6$ degrees and even in this case we have a pilot symbol in every block of 20 transmitted symbols. For the discretization-based algorithm $L = 8M = 32$ quantization levels are considered whereas for the algorithm based on Fourier parameterization, the number of Fourier coefficients is still $N = 17$.

In Fig. 5, the performance of the algorithms based on Tikhonov and Gaussian parameterizations is shown in the same conditions of Fig. 3. We observe that, despite the very low complexity, these algorithms have practically the same performance of more computationally demanding algorithms based on discretization and Fourier parameterization. This fact can be also observed from Fig. 6 where all the considered algorithms are compared for a Wiener phase model with $\sigma_\Delta = 6$ degrees. In this figure, the performance of two other algorithms described in the literature is also shown for the sake of comparison. The first one is the "ultra fast" algorithm with overlapped windows described in [23], with the value of N optimized by computer simulation. The second one is based on the EM algorithm [17]–[21]. In order to adapt the algorithm to a time-varying channel phase, different phase estimates are computed for each code symbol, taking into account the contribution of the adjacent symbols belonging to a window whose dimension is optimized by computer simulation. For this reason the algorithm is denoted by EM with sliding window (EM-SW). We found that the optimal window has width of 60 symbols for the considered phase noise. In both cases, the performance loss is due to the fact that these two algorithms are designed for a different phase model, i.e., a block-constant phase. Based on the above experiments and on extensive numerical evidence (not shown for the sake of space limitation) we conclude that all the proposed algorithms exhibit a practically optimal performance (i.e., they perform as well as the discretization approach). Among them, those based on Tikhonov and Gaussian parameterizations, because of their low complexity (roughly equivalent to that of the EM-SW algorithm), represent the best candidates for this detection scenario. For this reason, these two algorithms will be considered in the remaining results.

The sensitivity to distributions of the pilot symbols is considered in Fig. 7. In the case of the Wiener model with $\sigma_\Delta = 6$ degrees, two different distributions are considered, namely 1 pilot simbol in each block of 20 consecutive bits and 20 pilots in each block of 400 consecutive bits (hence the effective transmission rate is the same). We may observe that the algorithm based on Tikhonov parameterization is almost insensitive thanks to the algorithm modification described in Section III-C. A similar modification is not possible in the case of the algorithm based on Gaussian parameterization, since it can be shown that the choice of a dominant term corresponds to a hard-decision based uniquely on the decoder outcome $P_d(c_k)$. We verified that this modification of the algorithm described in Section III-D does not provide any performance improvement. Note that, in general, the distribution of the pilots has to be optimized for the specific detection algorithm employed.

Finally, we consider the application of our new algorithms to the DVB-S2 system. We consider two standardized LDPC codes with codewords of length 64800 [29]. The first one has rate 2/3 and is mapped onto a 8-PSK modulation. The second one has rate 4/5 and is mapped onto a 32-APSK modulation. A maximum number of 50 iterations is considered and 36 pilot symbols every 1476 symbols are included, as prescribed by the existing standard [29]. The above mentioned phase noise ESA model is considered. The performance is shown in Fig. 8. For the algorithm based on Tikhonov parameterization, the loss due to phase noise is less than 0.1 dB in both cases. Notice that a further improvement in performance may be obtained if the maximum number of iterations is not limited to 50. The Kalman smoother (Gaussian parameterization) does not perform as well mainly because of the bursty allocation of pilot symbols.

V. CONCLUSIONS

In this paper, the problem of joint detection and decoding of coded signals transmitted over an AWGN channel affected by strong phase noise has been considered. We obtained a number of new algorithms based on the direct application of the sum-product algorithm over the factor-graph representing a suitable factorization of the posterior joint probability mass function of the information bits given the received signal. To overcome the problem of computing and propagating messages in the graph corresponding to probability density functions (associated to continuous random variables), we used the method of canonical distributions.

Different parameterizations have been considered. Among all considered schemes, the novel algorithm based on Tikhonov parameterization exhibits practically optimal performance and very low complexity, and represents an attractive solution for systems where powerful LDPC-coded modulations are transmitted in the presence of phase noise, such as in next-generation satellite Digital Video Broadcasting.

APPENDIX

A. Some facts about Gaussian distributions

The following facts can be easily proved by direct calculation.

Fact 1. Let $p(x)$ and $q(x)$ be two probability distribution functions. The divergence $D(p\|q)$ (also known as cross-entropy, or Kullbach-Leibler distance [33]) is defined by

$$D(p\|q) = \int p(x) \log \frac{p(x)}{q(x)} dx \,.$$

Let $f(x)$ be a probability distribution function of a real random variable with mean μ_f and variance σ_f^2. The solution to the divergence minimization problem $\min_{g \in \mathcal{G}} D(f\|g)$, where $\mathcal{G}$ is the set of all real Gaussian pdfs, is given by $g(\mu_f, \sigma_f^2, x)$, i.e., it is the Gaussian pdf with the same mean and variance.

Let $f(x)$ be the probability distribution function of a complex random variable with mean μ_f and variance σ_f^2. The solution to the divergence minimization problem $\min_{g \in \mathcal{G}'} D(f\|g)$, where $\mathcal{G}'$ is the set of all complex circularly-symmetric Gaussian pdfs, is given by $g_{\mathbb{C}}(\mu_f, \sigma_f^2, x)$, i.e., it is the complex circularly-symmetric Gaussian pdf with the same mean and variance. □

Fact 2.

$$g_{\mathbb{C}}(A_1, \Sigma_1, x) g_{\mathbb{C}}(A_2, \Sigma_2, x) \propto g_{\mathbb{C}} \left(\frac{\Sigma_2}{\Sigma_1 + \Sigma_2} A_1 + \frac{\Sigma_1}{\Sigma_1 + \Sigma_2} A_2, \frac{\Sigma_1 \Sigma_2}{\Sigma_1 + \Sigma_2}, x \right) . \quad (50)$$

The proportionality constant is trivially given by

$$\kappa = \int g_{\mathbb{C}}(A_1, \Sigma_1, x) g_{\mathbb{C}}(A_2, \Sigma_2, x) dx \,.$$

The same relationship holds for real Gaussian pdfs. □

Fact 3.

$$\int g_{\mathbb{C}}(A_1, \Sigma_1, x) g_{\mathbb{C}}(A_2 x, \Sigma_2, y) dx \propto g_{\mathbb{C}}\left(A_1 A_2, \Sigma_2 + |A_2|^2 \Sigma_1, y\right) \,. \tag{51}$$

The same relationship holds for real Gaussian pdfs. □

Fact 4. Let $X \sim \mathcal{N}_{\mathbb{C}}(A, \Sigma)$ be expressed in magnitude and phase as $X = Re^{j\theta}$. Hence, the joint distribution of R and θ is given by

$$f(R, \theta) = \frac{1}{\pi\Sigma} R \exp\left(-\frac{R^2 + |A|^2}{\Sigma}\right) \exp\left(\frac{2R|A|\cos(\theta - \phi)}{\Sigma}\right) \tag{52}$$

where we have defined $A = |A|e^{j\phi}$. □

Fact 5. Let $a, b \in \mathbb{R}_+$. We have,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(a\cos(\theta - \alpha) + b\cos(\theta - \beta)\right) d\theta = I_0\left(\sqrt{a^2 + b^2 + 2ab\cos(\alpha - \beta)}\right) \tag{53}$$

where $I_0(z)$ is the modified Bessel function of the first kind and order zero. □

B. Modified Tikhonov parameterization

Consider the function

$$f(y) = \frac{1}{\sqrt{2\pi\sigma_\Delta^2}} \int_0^{2\pi} e^{\mathrm{Re}[ze^{-jx}]} e^{-\frac{(x-y)^2}{2\sigma_\Delta^2}} dx = \frac{1}{\sqrt{2\pi\sigma_\Delta^2}} \int_{-\pi}^{\pi} e^{\mathrm{Re}[ze^{-j(x+y)}]} e^{-\frac{x^2}{2\sigma_\Delta^2}} dx$$

where z is a complex number and x and y are real numbers. By discarding irrelevant multiplicative factors, we shall show that $f(y) \simeq e^{\gamma(\sigma_\Delta^2, |z|)\mathrm{Re}[ze^{-jy}]}$, where $\gamma(\sigma_\Delta^2, |z|)$ is a real function of $|z|$ and σ_Δ^2. This can be seen by using the following approximation which holds for large values of $a \in \mathbb{R}^+$ (in practice $a > 5$)

$$\frac{e^{a\cos(x-y)}}{2\pi \mathrm{I}_0(a)} \simeq \frac{1}{\sqrt{2\pi/a}} e^{-\frac{a}{2}(x-y)^2} = g(y, \frac{1}{a}, x) \,. \tag{54}$$

In fact, for sufficiently large values of a, the Tikhonov pdf $\frac{e^{a\cos(x-y)}}{2\pi \mathrm{I}_0(a)}$ has its support in a small interval around y. Hence, by using a second-order Taylor expansion, we have $\cos(x-y) \simeq 1 - \frac{(x-y)^2}{2}$. A normalization constant has been further added to obtain a pdf.

The correction term γ in (41) and in (42) can be derived by using the approximation (54). In fact, we let

$$\begin{aligned}
f(y) &= \frac{1}{\sqrt{2\pi\sigma_\Delta^2}} \int_0^{2\pi} e^{\mathrm{Re}[ze^{-jx}]} e^{-\frac{(x-y)^2}{2\sigma_\Delta^2}}\, dx \\
&\overset{(a)}{\simeq} \int_{-\infty}^{\infty} e^{\mathrm{Re}[ze^{-jx}]} g(x, \sigma_\Delta^2, y)\, dx \\
&\overset{(b)}{\simeq} 2\pi \mathrm{I}_0(|z|) \int_{-\infty}^{\infty} g(\phi(z), \frac{1}{|z|}, x) g(x, \sigma_\Delta^2, y)\, dx \\
&\overset{(c)}{\propto} g(\phi(z), \frac{1}{|z|} + \sigma_\Delta^2, y) \\
&\overset{(d)}{\simeq} \frac{1}{2\pi \mathrm{I}_0(\frac{|z|}{1+\sigma_\Delta^2|z|})} \exp\left\{ \frac{1}{1+\sigma_\Delta^2|z|} \mathrm{Re}[ze^{-jy}] \right\} \\
&\propto \exp\left\{ \frac{1}{1+\sigma_\Delta^2|z|} \mathrm{Re}[ze^{-jy}] \right\}
\end{aligned} \tag{55}$$

where (a) follows from the observation that, for $\sigma_\Delta \ll 1$, the function $e^{-\frac{(x-y)^2}{2\sigma_\Delta^2}}$ has its support in a small interval around y, (b) and (d) follow from the approximation (54) and (c) follows from Fact 3 of Appendix A.

Hence

$$\gamma(\sigma_\Delta^2, |z|) = \frac{1}{1+\sigma_\Delta^2|z|} .$$

REFERENCES

[1] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, pp. 498–519, Feb. 2001.

[2] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, Mar. 1973.

[3] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284–287, Mar. 1974.

[4] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. on Commun.*, vol. 44, no. 10, pp. 1261–1271, October 1996.

[5] R. G. Gallager, *Low-Density Parity-Check Codes*, MIT Press, Cambridge, MA, 1963.

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

[7] I. D. Marsland and P. T. Mathiopoulos, "On the performance of iterative noncoherent detection of coded M-PSK signals," *IEEE Trans. on Commun.*, vol. 48, no. 4, pp. 588–596, Apr. 2000.

[8] G. Colavolpe, G. Ferrari, and R. Raheli, "Noncoherent iterative (turbo) detection," *IEEE Trans. on Commun.*, vol. 48, no. 9, pp. 1488–1498, Sept. 2000.

[9] G. Ferrari, G. Colavolpe, and R. Raheli, "Noncoherent iterative decoding of spectrally efficient coded modulations," *Annals of Telecommun.*, vol. 56, pp. 409–421, July/August 2001.

[10] A. Anastasopoulos and K. M. Chugg, "Adaptive iterative detection for phase tracking in turbo coded systems," *IEEE Trans. on Commun.*, vol. 49, Dec. 2001.

[11] A. P. Worthen and W. E Stark, "Unified design of iterative receivers using factor graphs," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 843–849, Feb. 2001.

[12] H. Jin and T. Richardson, "Design of low-density parity-check codes for noncoherent MPSK communication," in *Proc. IEEE symposium on information theory*, Lausanne, Switzerland, June-July 2002, p. 169.

[13] H. Jin and T. Richardson, "On iterative joint decoding and demodulation," in *Proc. 41st Allerton Conference*, Monticello, IL, USA, October 2003.

[14] J. Dauwels and H.-A. Loeliger, "Joint decoding and phase estimation: an exercise in factor graphs," in *Proc. IEEE Symposium on Information Theory*, Yokohama, Japan, July 2003, p. 231.

[15] J. Dauwels and H.-A. Loeliger, "Phase estimation by message passing," in *Proc. IEEE International Conf. Commun.*, Paris, France, June 2004.

[16] R. Nuriyev and A. Anastasopoulos, "Pilot-symbol-assisted coded transmission over the block-noncoherent AWGN channel," *IEEE Trans. on Commun.*, vol. 51, pp. 953–963, June 2003.

[17] V. Lottici and M. Luise, "Carrier phase recovery for turbo-coded linear modulations," in *Proc. IEEE International Conf. Commun.*, Apr. 2002, pp. 1541–1545.

[18] V. Lottici and M. Luise, "Embedding carrier phase recovery into iterative decoding of turbo-coded linear modulations," submitted to IEEE Trans. on Commun., 2002.

[19] N. Noels, C. Herzet, A. Dejonghe, V. Lottici, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, "Turbo synchronization: an EM algorithm interpretation," in *Proc. IEEE International Conf. Commun.*, Anchorage, AK, U.S.A., June 2003, pp. 2933–2937.

[20] N. Noels, V. Lottici, A. Dejonghe, H. Steendam, M. Moeneclaey, M. Luise, and L. Vandendorpe, "A theoretical framework for soft information based synchronization in iterative (turbo) receivers," submitted to IEEE Trans. on Commun., 2003.

[21] H. Steendam, N. Noels, and M. Moeneclaey, "Iterative carrier phase synchronization for low-density parity-check coded systems," in *Proc. IEEE International Conf. Commun.*, Anchorage, AK, U.S.A., June 2003, pp. 3120–3124.

[22] L. Zhang and A. Burr, "Application of turbo principle to carrier phase recovery in turbo encoded bit-interleaved coded modulation system," in *Proc. Intern. Symp. on Turbo Codes & Relat. Topics*, Brest, France, Sept. 2003, pp. 87–90.

[23] I. Motedayen-Aval and A. Anastasopoulos, "Polynomial-complexity noncoherent symbol-by-symbol detection with application to adaptive iterative decoding of turbo-like codes," *IEEE Trans. on Commun.*, vol. 51, pp. 197–207, Feb. 2003.

[24] A. Ginesi, D. Fittipaldi, A. Bigi, and R. De Gaudenzi, "Pilot-aided carrier synchronization techniques for broadband satellite transmissions," Tech. Rep., ESA-ESTEC, Sept. 2003.

[25] L. Benvenuti, L. Giugno, V. Lottici, and M. Luise, "Code-aware carrier phase noise compensation on turbo-coded spectrally-efficient high-order modulations," in *8-th Intern. Work. on Signal Processing for Space Commun.*, Catania, Italy, Sept. 2003, pp. 177–184.

[26] O. Macchi and L.L. Scharf, "A dynamic programming algorithm for phase estimation and data decoding on random phase channels," *IEEE Trans. Inform. Theory*, pp. 581–595, September 1981.

[27] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*, Dover, 1972.

[28] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, New Jersey, 1989.

[29] ETSI, "ETSI - DVBS2 74r13, Digital Video Broadcasting (DVB): Second generation framing structure, channel coding and modulation systems for Braoadcasting, Interactive Services, News Gathering and other broadband satellite applications," 2003.

[30] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*, Prentice-Hall, 2000.

[31] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, 1994.

[32] D. J. C. MacKay, "Regular LDPC online database," available at the url *http://www.inference.phy.cam.ac.uk/mackay/*.

[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, Inc., New York, 1991.

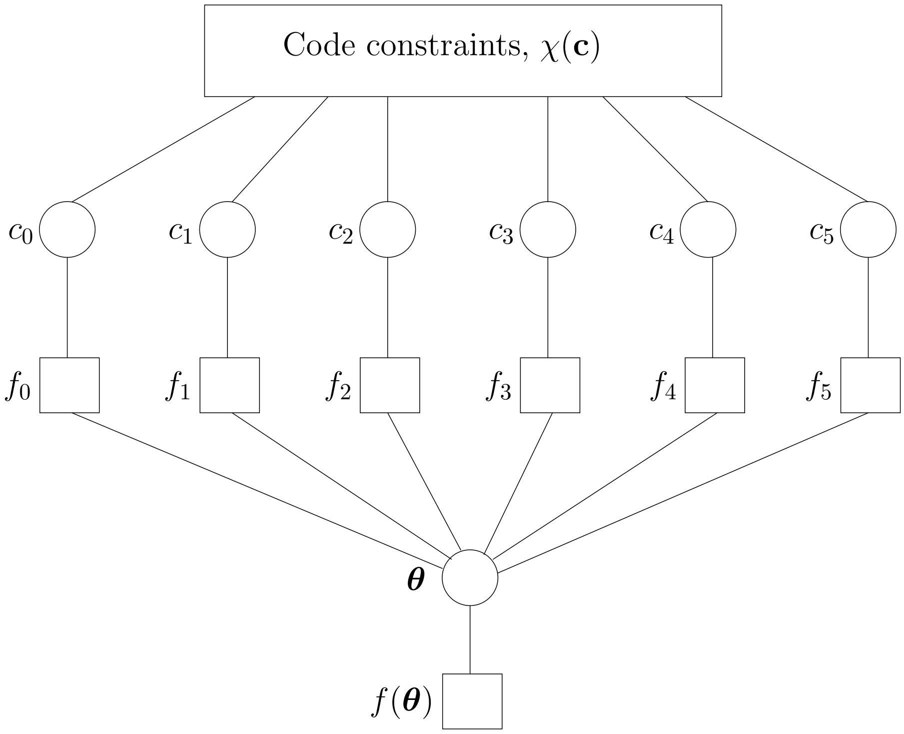

Fig. 1. Factor graph corresponding to eqn. (6).

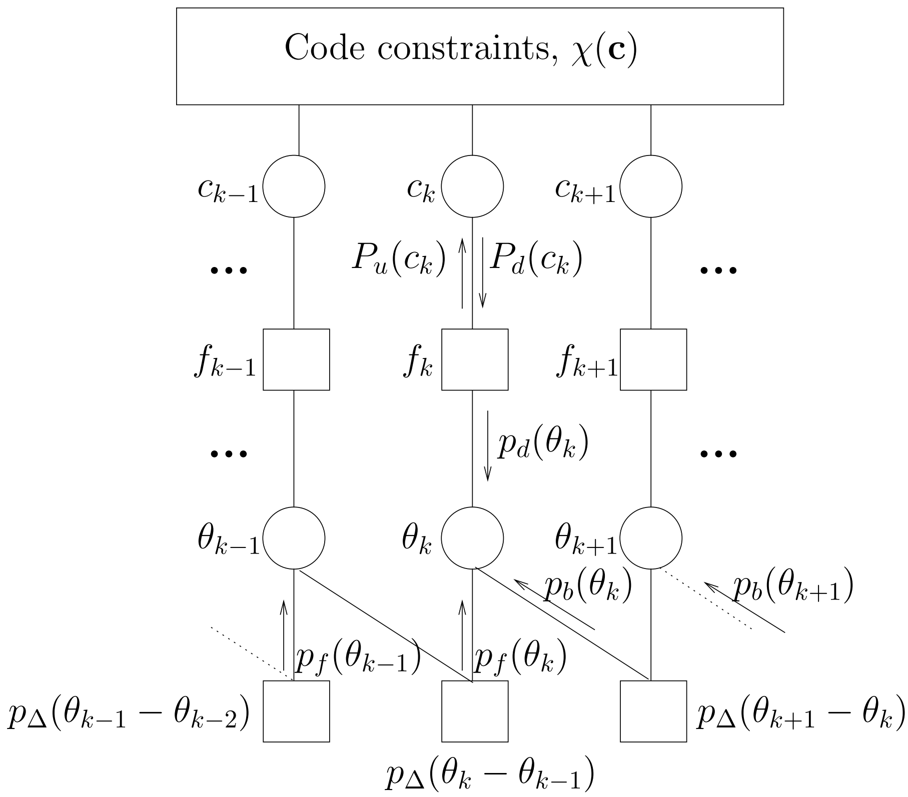

Fig. 2. Factor graph corresponding to eqn. (8).

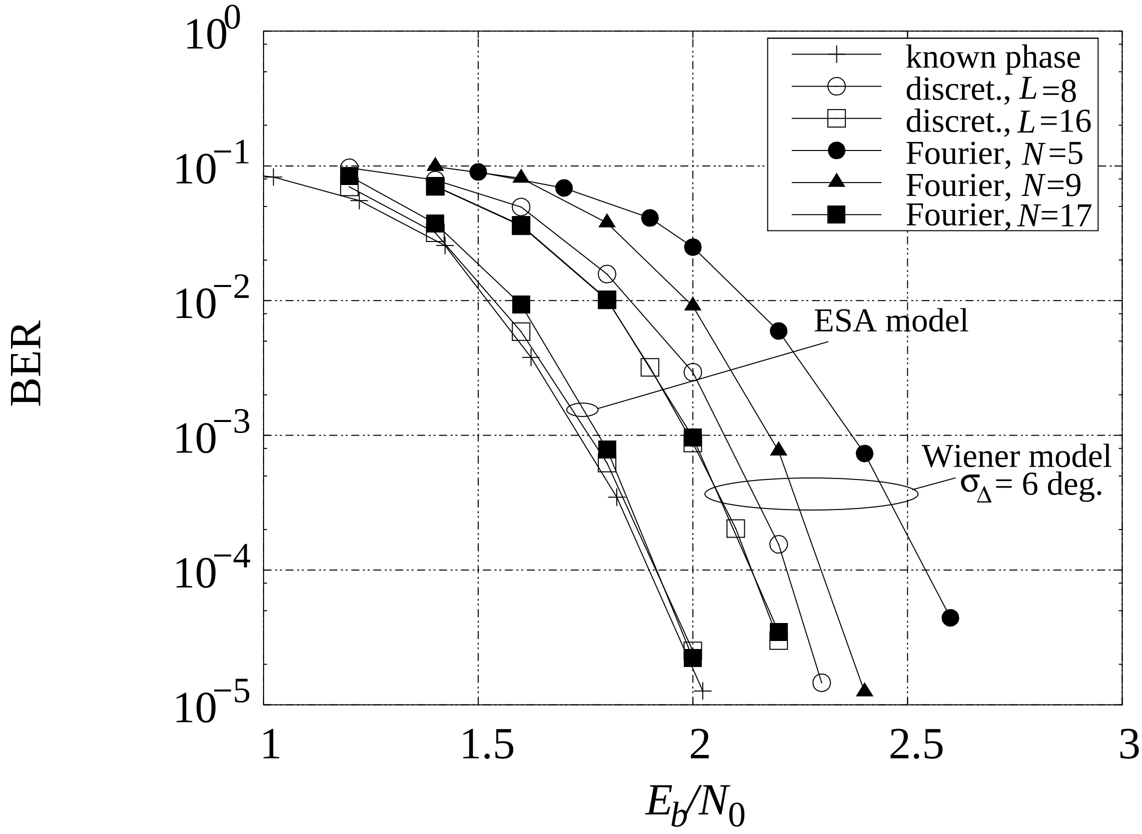

Fig. 3. Performance of the algorithms based on discretization of channel parameters and Fourier parameterization. BPSK and two different phase models are considered.

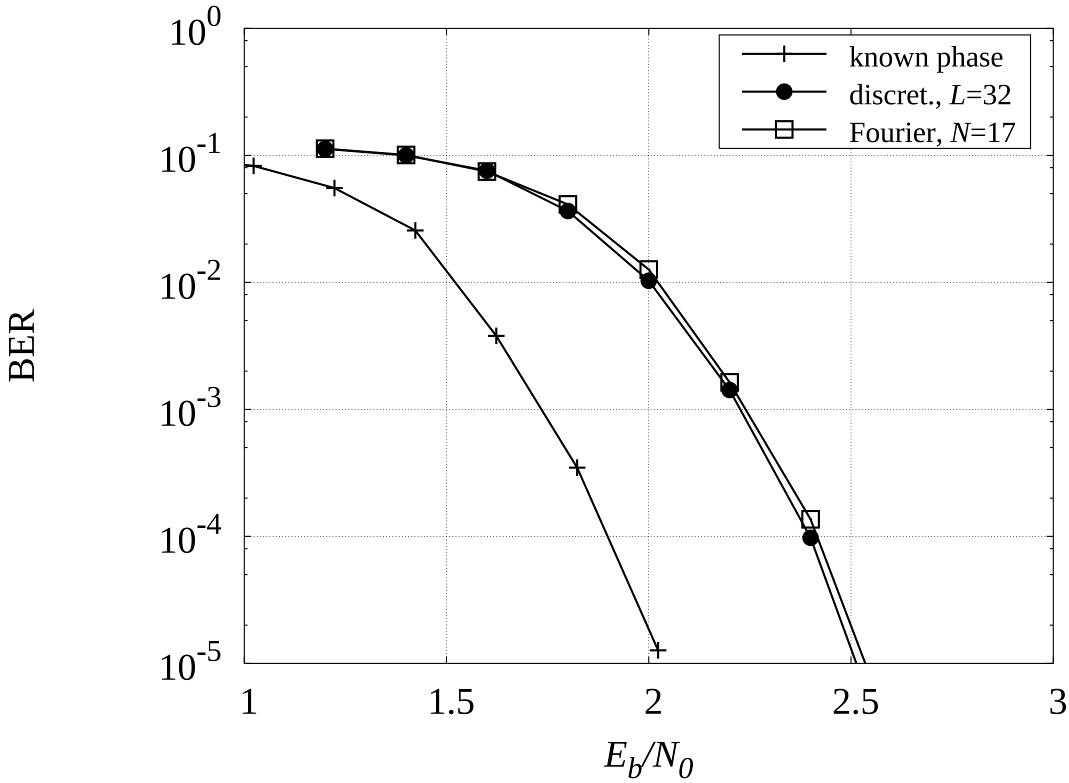

Fig. 4. Performance of the algorithms based on discretization of channel parameters and Fourier parameterization. QPSK and the Wiener model with $\sigma_\Delta = 6$ degrees are considered.

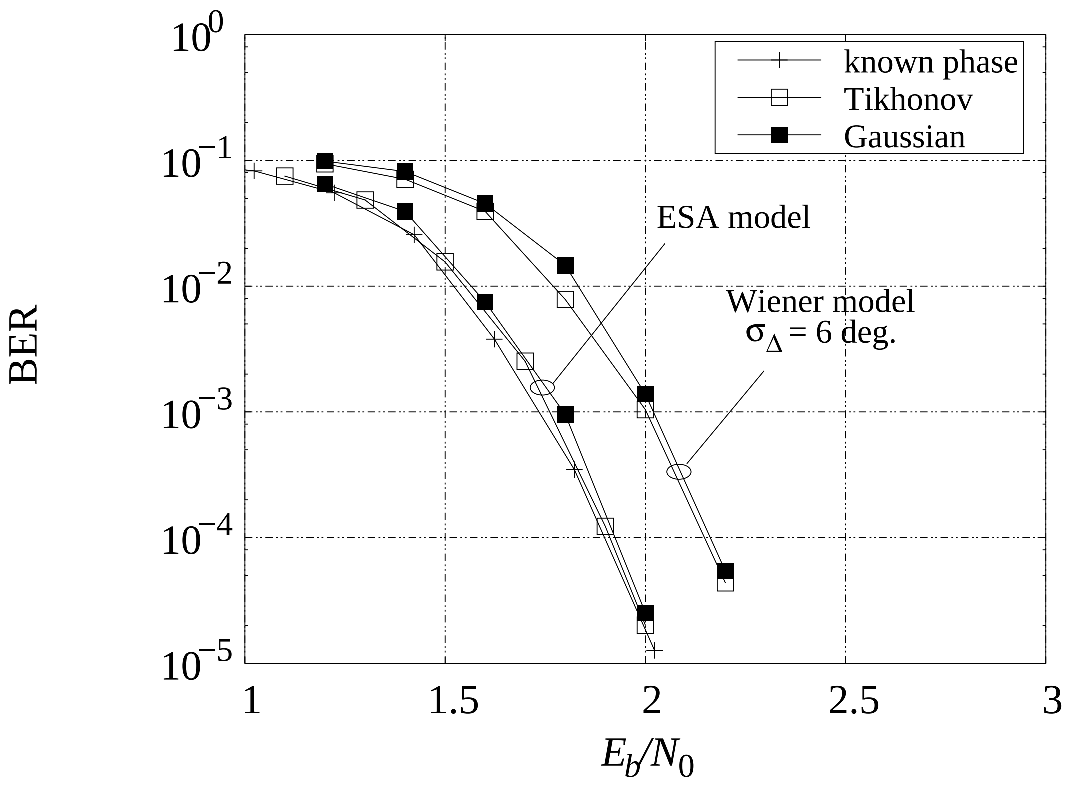

Fig. 5. Performance of the algorithms based on Tikhonov and Gaussian parameterizations. BPSK and two different phase models are considered.

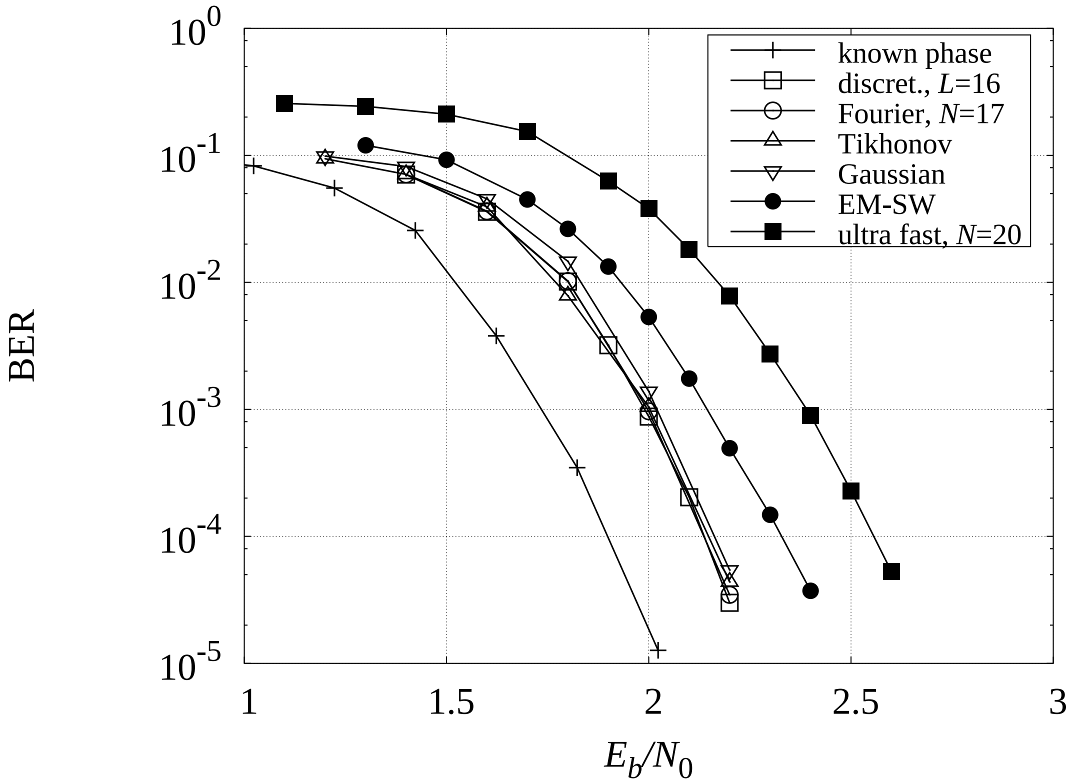

Fig. 6. Performance of all the proposed algorithms and comparison with other algorithms proposed in the literature. BPSK and the Wiener phase model with $\sigma_\Delta = 6$ degrees are considered.

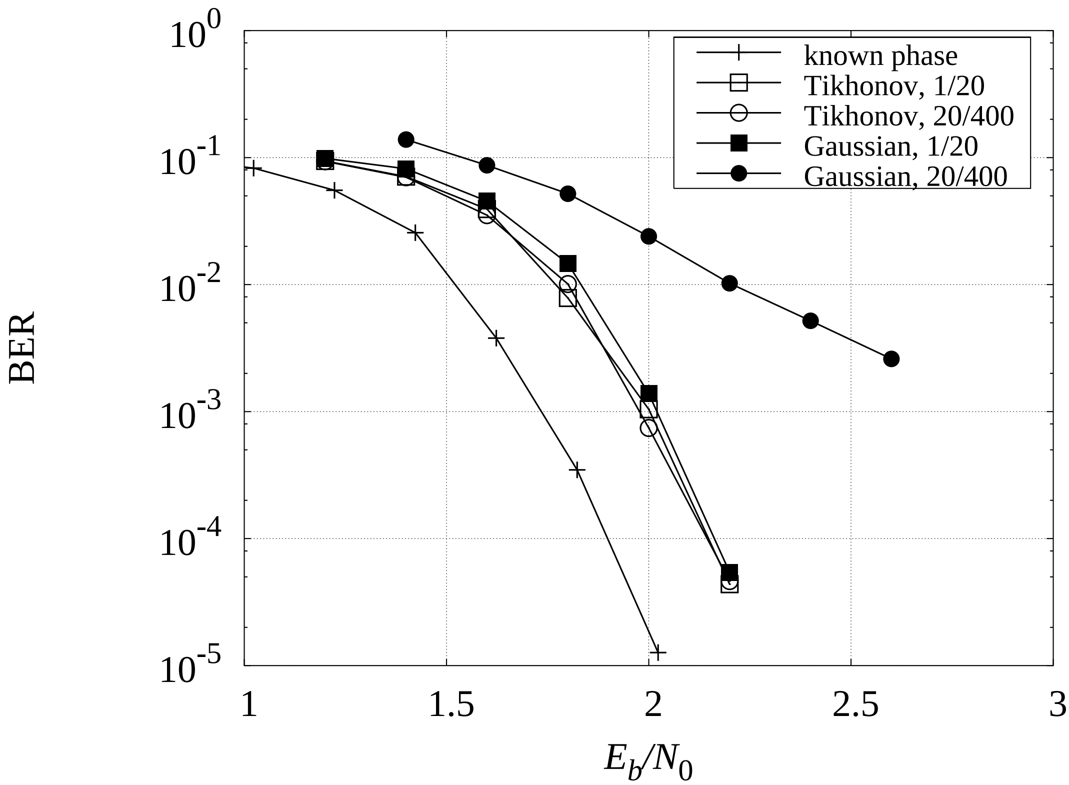

Fig. 7. Performance of the algorithms based on Tikhonov and Gaussian parameterizations. BPSK and two different pilot distributions are considered.

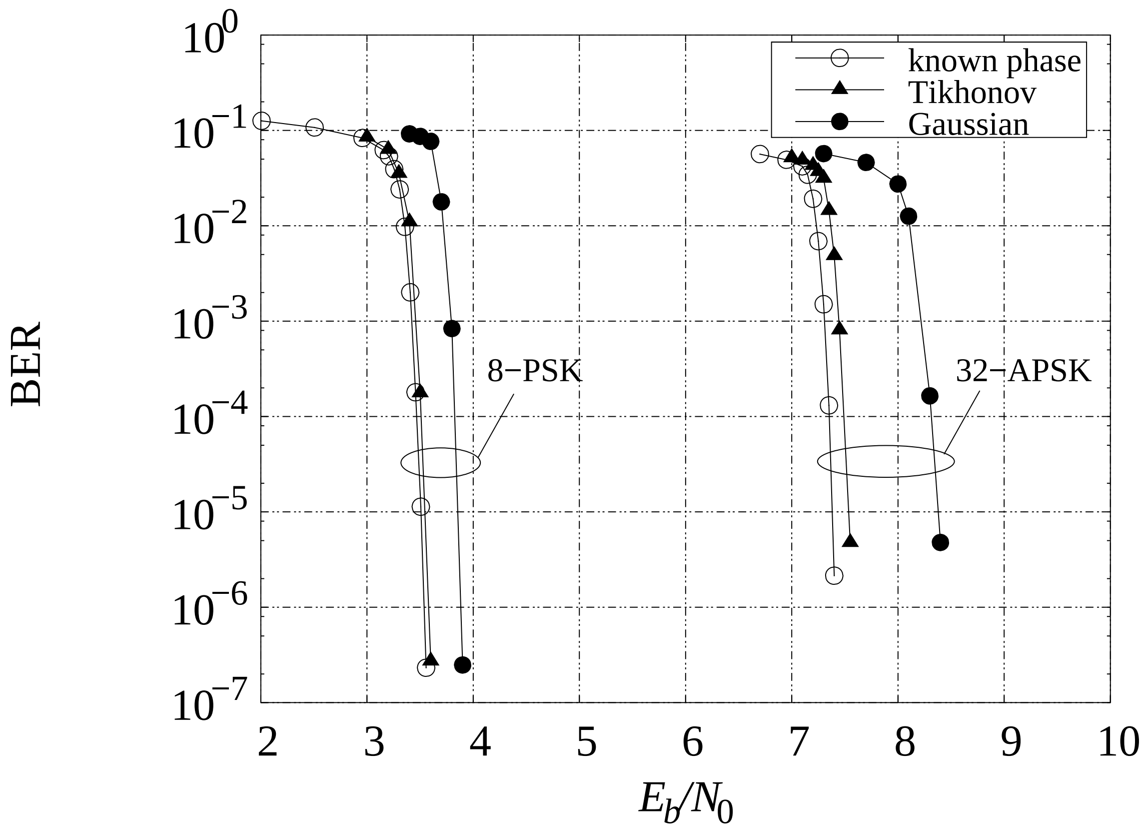

Fig. 8. Performance of the algorithms based on Tikhonov and Gaussian parameterizations. The ESA phase model is considered along with 8-PSK and 32-APSK modulations.