The Difficult Road of Expectation Propagation Towards Phase Noise Detection

Giulio Colavolpe, Elisa Conti, Amina Piemontese, and Armando Vannucci
University of Parma and CNIT Research Unit, Dept. of Engineering and Architecture, 43124 Parma, Italy
ORCIDs 0000-0002-0577-8626, 0000-0003-2365-136X, 0000-0002-7802-9634, and 0000-0002-5939-4134

ICC 2023 - IEEE International Conference on Communications | 978-1-5386-7462-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICC45041.2023.10278761

***Abstract*—Expectation Propagation (EP) is a promising framework in message-passing algorithms based on factor-graphs. The inherent ability to combine prior (partial) knowledge of system variables with channel observations suggests that an effective estimation of random channel parameters can be achieved even with a very limited number of pilot symbols, thus increasing the payload efficiency. Yet, the way in which the probability distributions of latent variables (both data and parameters) are combined and projected often requires ad-hoc adjustments to reach satisfactory performance. Here, we apply EP to a classical problem of LDPC-coded transmission on a strong Wiener phase noise channel and discuss how and why, even in the simple case of binary modulation, EP can fail or succeed.**

***Index Terms*—expectation propagation (EP), phase-noise, iterative detection/decoding, factor graphs (FGs), sum-product algorithm (SPA), low-density parity-check (LDPC) codes.**

I. Introduction

Detection algorithms for channels affected by a time-varying phase noise have received a lot of attention in the literature considering either linear or continuous-phase modulations and different scenarios (see, e.g., [1]–[5] and references therein). This is because in many communication links phase noise must be considered one of the major impairments. Examples are represented by scenarios employing a high carrier frequency, such as (coherent) optical communications, communications from geostationary satellites, etc. In particular, iterative soft-output detection and decoding has been a widely investigated subject.

One of the most effective algorithms in this class is that proposed in [1]. It is designed using the framework based on factor graphs (FGs) and the sum-product algorithm (SPA). In a scenario like the one at hand, where continuous random variables (the channel phase) appear in the FG, a common approach to implement the SPA is to resort to the use of canonical distributions [6]. In particular, in [1], the messages representing the a-posteriori probability density functions (pdfs) of the channel phase are represented through Tikhonov distributions, which can be described by a single complex parameter.

Although suboptimal, the resulting algorithm exhibits an excellent trade-off between performance and complexity. The suboptimality is related to the presence of both discrete (the code symbols) and continuous random variables in the graph, which brings up mixture pdfs with exponential proliferation, approximated in [1] with unimodal distributions. The presence of distributed pilot symbols is thus required to make the iterative joint detection and decoding algorithm bootstrap.

As an extension of [3], the results of Raphaeli and coworkers [4], [5] are obtained by letting mixture messages propagate one step further into the FG; their exponential proliferation being limited by an appropriate pruning of the mixture components, performed at the level of the Markov chain governing phase noise. This way, the phase uncertainty inherent in the observation of channel output is free to interact with the provisional estimation of previous phase samples (or with the following ones, in backward block processing). Such a mixture message reduction approach, however, is characterized by a considerable complexity.

The same objective of letting channel observations interact with provisional estimates of adjacent phase samples is achieved by the expectation propagation (EP) algorithm [7]–[10], where similar mixture reduction techniques are applied to the marginal distributions of the variables of interest, rather than to individual messages. Despite the projection of messages or marginals do have some features in common, and can even reduce to the same algorithm in some cases [11]–[13], there is a profound conceptual difference between them. In fact, only in the latter case, a message coming from a channel observation is merged, i.e., multiplied, with the (provisional) *prior belief* on the destination variable that comes from the rest of the FG. This is the key to the potential success of EP in many similar applications, including, e.g., transmission over fading channels [14]–[16].

The EP framework has already been successfully applied to phase noise channels, with either distributed or concentrated pilots [17], [18], resorting to a joint detection and decoding, as implemented by EP message passing on the overall FG. This is not, however, a practical solution, since a separate (and sequential) detection of channel parameters and decoding is desirable in the design of digital receivers.

In this work, we discuss why an effective phase detection is not achieved by EP before any data information is provided by the decoder. We propose a modification of EP that overcomes its limitations and remarkably reduces the algorithmic complexity. The proposed algorithm reaches the performance benchmark even in the challenging scenario where pilot symbols are concentrated in the preamble/postamble of the data packet, as done in [17], to solve the initial phase ambiguity. Concentrating pilots is consistent, for instance, with the CCSDS (Consultative Committee for Space Data Systems)

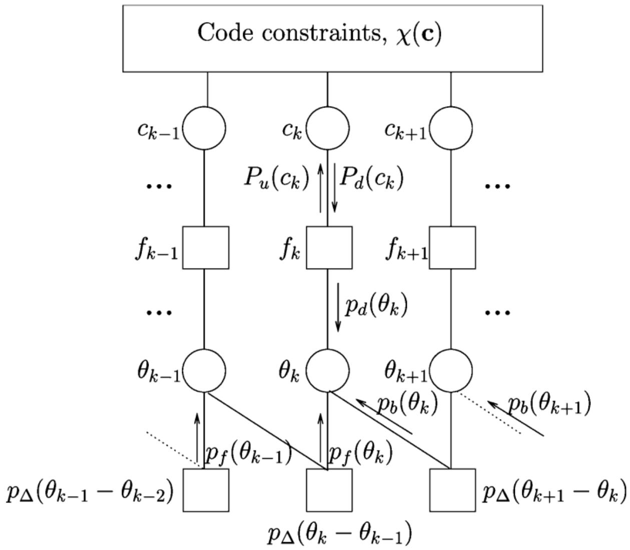

Fig. 1. Factor graph representing the joint probability distribution of phase noise samples and coded symbols.

standards for deep space telemetry and telecommand. In a scenario like that of deep space communications, where low signal power levels and low baud rates are employed, one often resorts to a binary modulation format, which is (also for the sake of simplicity) the one adopted in the investigated system.

II. System Model and Related Factor Graph

In the system we wish to investigate, an LDPC codeword $\boldsymbol{c} = [c_0, c_1, \cdots, c_{K-1}]$, with length K, is transmitted over an additive white Gaussian noise (AWGN) channel also affected by Wiener phase noise, so that the received samples are

$$r_k = c_k e^{j\theta_k} + n_k \quad , \tag{1}$$

where $\{n_k\}$ is a sequence of independent (complex and circularly symmetric) Gaussian noise samples, i.e., $n_k \sim \mathcal{N}_{\mathbb{C}}(0, 2\sigma^2)$, with zero mean and given variance per component σ^2, while the phase noise sequence $\boldsymbol{\theta} = [\theta_0, \theta_1, \cdots, \theta_{K-1}]$ follows the Wiener model, where each sample

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

results from the previous one plus a zero-mean (real) Gaussian increment $\Delta_k \sim \mathcal{N}(0, \sigma_\Delta^2)$ whose variance σ_Δ^2 dictates the severity of phase noise.

Since the system model is the same as the one discussed in [1], the FG representing the joint distribution of the latent variables $\boldsymbol{\theta}$ and $\boldsymbol{c}$, conditioned on the value of the observed variables $\boldsymbol{r} = [r_0, r_1, \cdots, r_{K-1}]$, is the same as the one shown therein and is reported in Fig. 1. The subscripts of messages identify their direction in the FG (*u*p, *d*own, *f*orward, *b*ackward), whereas the subscript of variable and factor nodes coincides with a time index.

Before observing the channel output, the phase noise sequence is independent of that of coded symbols, so that the joint posterior distribution[1] of all latent system variables is

$$\begin{aligned} P(\boldsymbol{c}, \boldsymbol{\theta} | \boldsymbol{r}) &\propto P(\boldsymbol{c}) p(\boldsymbol{\theta}) p(\boldsymbol{r} \mid \boldsymbol{c}, \boldsymbol{\theta}) \\ &\propto \chi(\boldsymbol{c}) p(\theta_0) \prod_{k=1}^{K-1} p(\theta_k \mid \theta_{k-1}) \prod_{k=0}^{K-1} f_k(c_k, \theta_k) \end{aligned} \tag{3}$$

where $\propto$ is the proportionality symbol, $\chi(\boldsymbol{c})$ is an indicator function implementing the code constraints and we assumed that all the allowed codewords in the employed LDPC codebook are equally likely.

The factor node f_k implements the observation of the received samples in (1), hence

$$\begin{aligned} f_k(c_k, \theta_k) = p(r_k \mid c_k, \theta_k) &= g_{\mathbb{C}}(r_k - c_k e^{j\theta_k}; \sigma^2) \\ &\propto \exp\left(-\frac{1}{2\sigma^2}\left|r_k - c_k e^{j\theta_k}\right|^2\right) \end{aligned} \tag{4}$$

is a complex Gaussian distribution that we denote by $g_{\mathbb{C}}(x - \eta_X; \sigma_X^2)$, for a general complex vector X with mean η_X and covariance matrix σ_X^2.

Besides f_k, the only other factor nodes in the FG of Fig. 1 (if we disregard the check nodes included in the code subgraph) are the ones related to the conditional pdfs in (3) that implement the Markov chain governing the Wiener phase noise,

$$p(\theta_k \mid \theta_{k-1}) \triangleq p_\Delta(\theta_k - \theta_{k-1}) = g(\theta_k - \theta_{k-1}; \sigma_\Delta^2) \quad , \tag{5}$$

which are Gaussian too, as per (2). In (5), we denote by $g(y - \eta_Y; \sigma_Y^2)$ the pdf of a real Gaussian variable Y with mean η_Y and covariance matrix σ_Y^2.

Owing to the fact that the additive sources of randomness in (1) and (2) are Gaussian, the FG in Fig. 1 seems to represent a Gaussian belief network, where all messages are Gaussian [19] and the message passing procedure resembles the simple operations of a Kalman smoother. This is clearly not true, for two reasons. First, the problem at hand entails a "mixed model", where continuous ($\boldsymbol{\theta}$) and discrete ($\boldsymbol{c}$) variables coexist, which produces probabilistic mixtures in the FG, specifically arising from the messages $p_d(\theta_k)$ which are linear combinations of simpler component distributions. In addition, (4) is a Gaussian pdf only if seen as a function of r_k whereas the factor nodes f_k send downward messages $p_d(\theta_k)$ to the variable nodes θ_k, hence (4) must be seen as a function of θ_k and is thus proportional to a Tikhonov distribution [1],

$$\begin{aligned} f_k(c_k, \theta_k) &\propto \exp\left(-\frac{|c_k|^2}{2\sigma^2} + \frac{1}{\sigma^2}\Re\left[r_k c_k^* e^{-j\theta_k}\right]\right) \\ &= \exp\left(-\frac{|c_k|^2}{2\sigma^2}\right) 2\pi I_0(|z_k|) t(\theta_k; z_k) \end{aligned} \tag{6}$$

where $z_k = r_k c_k^* / \sigma^2$ is the complex parameter characterizing the general Tikhonov distribution

$$t(\theta_k; z_k) = \frac{1}{2\pi I_0(|z_k|)} \exp\left(\Re\left[z_k e^{-j\theta_k}\right]\right) \quad . \tag{7}$$

[1]We denote by $p(\cdot)$ the probability density function of continuous variables (vectors) and with a capital $P(\cdot)$ the probability mass function (pmf) of discrete variables as well as the pdf of mixed random vectors.

In directional statistics [20], the phase $\angle z_k$ corresponds to the *circular mean* of the circular random variable θ_k in (7), while the magnitude $|z_k|$ is a measure of its precision, i.e., of the inverse of the variance $\sigma^2_{\theta_k} = 1 - \frac{I_1(|z_k|)}{I_0(|z_k|)} \in [0; 1]$, where $I_p(x)$ is the modified Bessel function of the first kind of order p.[2]

The Bayesian inference problem for the estimation of symbols and channel parameters cannot be solved by the plain SPA since the message

$$p_d(\theta_k) = \sum_{m=0}^{M-1} P_d(c_k^m) f_k(c_k^m, \theta_k) \propto \sum_{m=0}^{M-1} \alpha_k^m t(\theta_k; z_k^m) \\ \left(\alpha_k^m = P_d(c_k^m) \exp\left(-\frac{|c_k^m|^2}{2\sigma^2}\right) 2\pi I_0(|z_k^m|)\right) \tag{8}$$

is a Tikhonov mixture, whose coefficients α_k^m depend on the extrinsic information on each symbol (i.e., on the values of its pmf for each of the possible M symbols c_k^m), as provided by the LDPC decoding part of the FG (top of Fig. 1), as well as on the magnitude of the symbols. In the SPA, the mixture messages (8) would propagate through the bottom half of the FG and eventually proliferate to produce untractable mixtures with an exponentially increasing number of components. Approximate inference is thus demanded, which can be performed in a variety of ways.

III. Transparent Propagation Algorithms and Expectation Propagation

The first and most straightforward approach to approximate message passing is to discretize the distribution of θ_k, with a given number of samples N_θ, so that all the latent system variables appear to be discrete and modelled by their pmf. This produces an algorithm called discrete-phase BCJR (dp-BCJR) in [1], since the forward-backward message passing procedure, along the Markov chain in the bottom half of the FG, is the same as that of the celebrated BCJR algorithm. Its computational complexity is very high, thus we shall implement this algorithm, denoted as *dp-BCJR* in the simulation results, as a practical benchmark, assuming that its performance gets close to that of an ideal SPA when N_θ is sufficiently large.

A totally different approach relies on projecting messages and/or marginals onto a selected family of approximating distributions, usually in the exponential form, $q(x) = \exp(\sum_i \eta_i g_i(x))$ (so that the family is closed under the multiplication operation), where η_i are called natural parameters and the functions $g_i(x)$ are the features of the family. This way, only the natural parameters η_i are updated, for each distribution, resulting in a parametric message passing procedure. In [1], the Tikhonov approximating family is selected quite naturally, since it is known to be the marginal distribution of θ_k, conditioned on the values taken by the corresponding symbol and by the corresponding channel output, i.e., $p(\theta_k \mid c_k, r_k) \propto f_k(c_k, \theta_k)$, as per (4). Curiously, the way in which message $p_d(\theta_k)$ has been projected onto a Tikhonov pdf in [1] relies on the Gaussian expression (4) of the factor node, so that (8) is approximated by the Gaussian pdf with minimum Kullback-Leibler (KL) divergence from the mixture, which is further interpreted as a Tikhonov message towards θ_k. We shall denote this algorithm as *TP Gauss*, in the simulation results.

A more natural solution would have been to project the Tikhonov mixture (8) onto a Tikhonov pdf with minimum KL divergence, which is achieved (for this as well as for any other exponential approximating family) by matching the expectations of the features [8]. The features of a Tikhonov pdf like (7) are $\cos(\theta_k)$ and $\sin(\theta_k)$, associated with the natural parameters $\Re[z_k]$ and $\Im[z_k]$ respectively, so that their expectations can be jointly computed as the $\Re/\Im$ parts of the first trigonometric moment $E[\exp(j\theta_k)]$. Exploiting known results [20] and the linearity of expectation, the mixture in (8) is approximated by the Tikhonov pdf $p_d^{TP}(\theta_k) = t(\theta_k; z_k^{TP})$, which achieves the minimum KL divergence when z_k^{TP} obeys the following moment matching equation

$$\frac{I_1(|z_k^{TP}|)}{I_0(|z_k^{TP}|)} e^{j\angle z_k^{TP}} = \sum_{m=0}^{M-1} \overline{\alpha}_k^m \frac{I_1(|z_k^m|)}{I_0(|z_k^m|)} e^{j\angle z_k^m} \tag{9}$$

where $\overline{\alpha}_k^m$ stands for the normalized version of the coefficients in (8), i.e., $\sum_{m=0}^{M-1} \overline{\alpha}_k^m = 1$. The complex equation (9) yields the circular mean $\angle z_k^{TP}$ as well as the variance of phase noise, related to the Bessel ratio of the magnitude $|z_k^{TP}|$ [4]. We shall denote this variation of the algorithm in [1] as *TP Tikhonov*, in the simulation results.

In the EP algorithmic framework [7], denoted as *EP* in the simulation results, it is the entire marginal of each latent variable that is approximated/projected; in our case, $p(\theta_k) = p_d(\theta_k)p_f(\theta_k)p_b(\theta_k)$, of which the message $p_d(\theta_k)$ only represents one factor. Denoting the projection operation by $proj[p(\cdot)] = \arg\min_{q(\cdot)\in\mathcal{F}} KL\left[p(\cdot) \parallel q(\cdot)\right]$, where $\mathcal{F}$ is the approximating family, the approximating message

$$p_d^{EP}(\theta_k) = \frac{proj[p_d(\theta_k)p_f(\theta_k)p_b(\theta_k)]}{p_f(\theta_k)p_b(\theta_k)} \tag{10}$$

is computed and sent in place of the mixture $p_d(\theta_k)$. In (10), the product of forward and backward messages, $p_f(\theta_k)p_b(\theta_k)$ plays the role of a temporary prior for the variable θ_k, whose estimated marginal is updated, after the observation $f_k(c_k, \theta_k)$. Hence, the EP approximation for the marginal $p(\theta_k)$ is $p^{EP}(\theta_k) = p_d^{EP}(\theta_k)p_f(\theta_k)p_b(\theta_k)$, which is equal to the numerator of (10) and thus belongs, by construction, to the approximating family (Tikhonov, in our case).

The projection operation in (10) still amounts to a moment matching operation like (9), despite a marginal is projected, in EP, rather than a message, as in TP. Of course, in the case of EP, the temporary prior $p_f(\theta_k)p_b(\theta_k)$ must be accounted for in the right hand side of (9), by simply adding the (complex) parameters $z_{k,f}$ and $z_{k,b}$ of the forward and backward messages to those of each mixture component, i.e., to z_k^m (hence the coefficients α_k^m must be calculated accordingly).

[2] More in general, the p-th trigonometric moment of the Tikhonov variable θ_k is $E_t[\exp(jp\theta_k)] = \exp(jp\angle z_k)\frac{I_p(|z_k|)}{I_0(|z_k|)}$, where $E_t[\cdot]$ denotes expectation under the distribution $t(\cdot)$, so that the phase of the first trigonometric moment is the circular mean.

The general approach to approximate message passing can be either a projection of the marginal pdf of each variable node (onto the selected approximating family), as prescribed by EP, or a projection of individual mixture messages onto the same family, as described above. If, for the sake of argument, the EP algorithm did not produce any information about the temporary prior, then we would assume $p_f(\theta_k)p_b(\theta_k)$ to be uniform in (10) and hence it would *transparently shift out* of the projection operation, to be simplified with the denominator, so that (10) would reduce to the simple projection of message $p_d(\theta_k)$. This is the reason for which we denote the message projection approach as transparent propagation (TP) [12], hence the superscript TP. It can be demonstrated that in some specific problems, the TP and EP approaches lead to the same solution [13], [21], [22], although in general they differ.

No matter if one of the TP algorithms or if EP is adopted, the forward and backward messages $p_f(\theta_k)$ and $p_b(\theta_k)$ are assumed to belong to the Tikhonov family too, despite the factor nodes in the Markov chain of the FG, e.g., $p(\theta_{k+1} \mid \theta_k)$, introduce a convolution between the Tikhonov message $p_d^{TP/EP}(\theta_k)p_f(\theta_k)$ and a Gaussian like (5), for the computation of $p_f(\theta_{k+1})$. As it is shown in the Appendix of [1], if the phase noise variance σ_Δ^2 is not exaggeratedly large then $p_f(\theta_{k+1})$ is very well approximated by a Tikhonov pdf whose parameter $|z_{f,k+1}|$ is less than that of the incoming message, $|z_{f,k} + z_k^{TP/EP}|$, while the circular mean is the same (see (36)-(38) in [1], for details). This corresponds conceptually to the fact that the Wiener phase noise update in (2) does not bias the temporary estimate for θ_{k+1} but decreases its precision, according to its Gaussian variance. A similar conclusion holds for the backward messages too.

IV. The Failure of EP and its Modification

The EP algorithm has been already applied to transmission over Wiener phase noise channels in [17] (and later in [18]), in the absence of distributed pilot symbols, achieving a good performance when global detector-decoder iterations are employed (in [17], $N_{DD} = 5$). Albeit, the challenge of making EP work with separate phase detection and decoding has not been solved. In fact, simulation results, presented in Sec. VI, reveal that, with distributed pilots, the performance of EP after the first phase-detecting iteration is limited and TP algorithms can perform better, especially in the large signal to noise ratio (SNR) regime. Its difficulties become even more evident when moving to a different scenario, where pilot symbols are concentrated, as a preamble/postamble (half and half) at both ends of the codeword. In the concentrated pilots scenario, however, EP is the only choice, since both algorithms based on the Tikhonov parametrization (*TP Gauss* and *TP Tikhonov*) are known to fail completely in this case [1], [17]. In fact, they cannot achieve any effective phase detection, at the first iteration, without the aid of distributed pilots. On the contrary, the interest in EP with concentrated pilots lies in its ability to self-sustain the process of phase estimation across a long block of payload coded symbols, once it is bootstrapped by a proper block of preamble/postamble pilots.

Through a detailed numerical analysis of EP (on which we shall not go deeper in this paper), we found that its failure is confined to some critical data packets that, due to a long sequence of noisy observations, bring the sequence of precision values $|z_{k,f/b}^{EP}|$ to a collapse, while propagating forward or backward messages, $p_f(\theta_k)$ and $p_b(\theta_k)$. Once a critical lower threshold is exceeded, the precision approaches zero and is rarely able to recover, even if the channel in (1) outputs reliable observations, i.e., samples r_k with little noise. Since $|z_{k,f/b}^{EP}|$ result from the EP message (10), computed by applying the moment matching (9), we verified that the above phenomenon is largely due to a numerical instability entailed in the approximation of the Bessel ratio $I_1(x)/I_0(x)$ in (9) whose inverse, not available in closed form, is necessary to obtain $|z_{k,f/b}^{EP}|$. In fact, despite the commonly adopted approximation $I_1(|z|)/I_0(|z|) \simeq \exp(-0.5/|z|)$ [17], [18] is very accurate when $|z| \gg 0$, the saturating shape of $I_1(|z|)/I_0(|z|) < 1$ (whose value switches from 0 to 0.9 when $|z|$ increases from 0 to 5) tends to compress the precision $|z|$, hence to depress the level of confidence of the estimation embedded in the transmitted message.

As a countermeasure, we introduce two modifications in the EP algorithm. First, we very roughly substitute $I_1(|z|)/I_0(|z|)$ with $|z|$, in (9), so as to avoid a saturation of the precision values. Remarkably, with this substitution, the moment matching operation takes the form of a linear combination of the Tikhonov parameters of the mixture components, so that the corresponding parameter for the marginal of θ_k results: $z_k^{EP} = \sum_{m=0}^{M-1} \overline{\alpha}_k^m z_k^m$. In the second place, we introduce a monitoring of the precision values $|z_{k,f/b}^{EP}|$ along the forward and backward propagation of $p_{f/b}(\theta_k)$. This revealed that every error in phase detection, i.e., every *phase slip*, is associated with a sudden drop of the corresponding precision values. It is then sufficient to reject those messages $p_f(\theta_k)$ or $p_b(\theta_k)$ whose precision decreases abruptly, and thus rely only on the opposite message ($p_b(\theta_k)$ or $p_f(\theta_k)$), to estimate θ_k.

V. Message Scheduling

Regarding message scheduling, for any algorithm, we take advantage of the structure of the FG, which evidences the logical separation of the decoding part from the phase detection part, corresponding to the upper and lower halves of Fig. 1. Indeed, although it is known that messages in an FG do not represent probability distributions,[3] $P_d(c_k)$ is the extrinsic information on the k-th code symbol, sent from the decoder to the phase detector, as well as $p_u(\theta_k) \triangleq p_f(\theta_k)p_b(\theta_k)$ (not specified in Fig. 1), which is seen as a temporary estimate of phase noise, sent towards the decoder.

We shall assume that the phase-detector subgraph operates first, by exchanging *horizontal* messages along the Markov chain in the lower part of Fig. 1, so as to update $p_f(\theta_k)$ and $p_b(\theta_k)$ by message passing (see [1], to recall the rules of computation). After that, *vertical messages* $p_u(\theta_k)$ are

[3]Hence, a so-called improper distribution, e.g., a Gaussian with negative variance, could arise from (10) and it is allowed to propagate it along the graph, although certain EP variants employ a rejection of such messages [19].

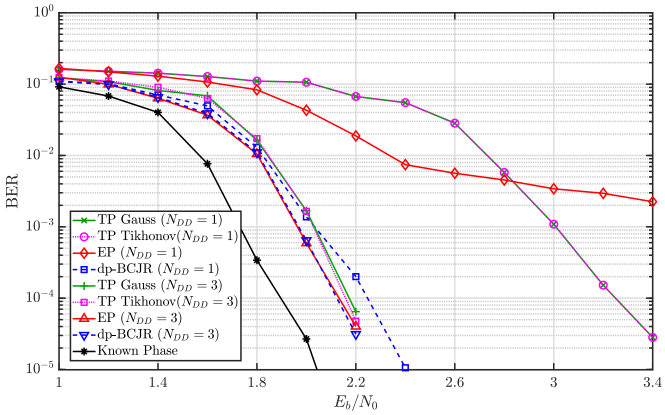

Fig. 2. Bit error rate curves for the tested algorithms, without ($N_{DD} = 1$) or with ($N_{DD} = 3$) detector-decoder iterations. The phase noise standard deviation is $\sigma_\Delta = 6$ [deg] and $(1/20)$ pilot symbols are distributed among the payload. In dp-BCJR, phase is discretized using $N_\theta = 64$ levels.

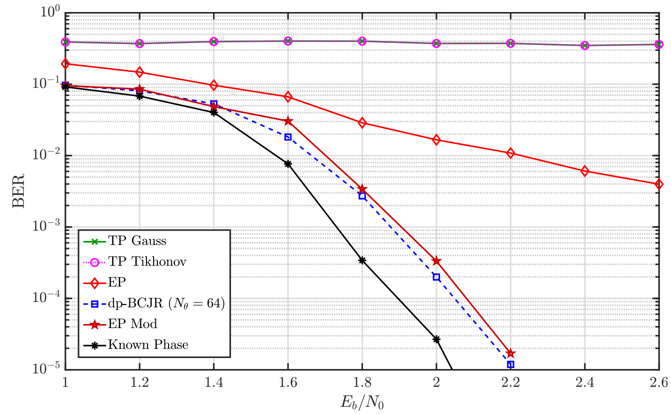

Fig. 3. Bit error rate curves for the tested algorithms, with separate detection and decoding ($N_{DD} = 1$). The phase noise standard deviation is $\sigma_\Delta = 3[deg]$ and $(1/20)$ pilot symbols are concentrated in the preamble and postamble (half and half) of the transmitted packet. In dp-BCJR, phase is discretized using $N_\theta = 64$ levels.

sent from the phase detector to the upper part of the FG, implementing the decoder, where an inner message passing procedure takes place, until convergence or until a maximum number of decoding iterations is reached (we fix this maximum to 200).

The exchange of information, through vertical messages, between detector and decoder can be iterated for N_{DD} times, so as to yield a refinement of both the symbols' and the phase samples' estimates, from which the other receiver half can benefit. On the other hand, detector-decoder iterations are considered impractical and are usually avoided in the implementation of digital receivers.

The horizontal message passing occurring in the phase detector subgraph could be itself iterated, with multiple forward-backward passes, before sending vertical messages to the decoder subgraph. Despite such multiple "inner" iterations could in principle refine the phase estimate, this is hardly ever the case. In fact, for this as well as for other similar problems that we analyzed [11], no real improvement was observed, hence we limited the phase detecting (horizontal) message passing to a single (inner) iteration.

VI. Simulation Results

We report in Fig. 2 the performance of different receivers for a block coded transmission through the system under investigation. Bit error rate (BER) is plotted versus the SNR E_b/N_0, where $N_0 = \sigma^2$ is the variance per component of noise samples in (1) and E_b is the average energy per information (payload) bit. We transmitted a block of 4000 coded symbols punctured by distributed pilot symbols, which are known at the receiver, amounting to a 5.3% overhead. Pilots were distributed across the block of payload symbols by regularly alternating one pilot symbol and 19 payload symbols, so that the transmitted sequence started with a pilot. The payload symbols were the output of a $(3, 6)$-regular low-density-parity-check (LDPC) code with rate-$1/2$ and length 4000 [23] and the binary phase-shift keying (BPSK) modulation format ($M = 2$) was adopted, since it is robust to the strong phase noise that was assumed with $\sigma_\Delta = 6°$.

For the known phase scenario, there was no need for phase estimation, hence the factor nodes f_k were *leaves* of the FG in Fig. 1, thus reduced to a tree. The f_k nodes sent the conditional probability of the observation $P(r_k \mid c_k)$ upwards (in our case, the log-likelihood ratio for the possible binary values of c_k) and the whole system reduced to a coded transmission over an AWGN channel. Thus, the *known phase* curve in Fig. 2 shows the performance of the adopted LDPC code. In order to make a fair comparison, we accounted for the presence of pilot symbols, when computing the average energy per bit E_b, even though they were useless in this case.

In the implementation of dp-BCJR, we found that the rule of thumb $N_\theta = 8M$ [1] (where M is the cardinality of a PSK constellation) was not sufficient for an accurate quantization of the phase noise variables, that required at least 32 levels in the BPSK case [17]; we used $N_\theta = 64$ as reported in the legends of Figs. 2,3. In the absence of detector-decoder iterations ($N_{DD} = 1$), dp-BCJR is less than half a dB away from the theoretical known phase reference, while multiple iterations bring its performance a little closer, practically reaching a limit after three iterations (reported in Fig. 2 with $N_{DD} = 3$).

The two TP algorithms described in Sec. III, with the Gaussian projection of [1] or with the Tikhonov projection in (9), performed identically, both showing a 1 dB penalty, compared to the practical dp-BCJR benchmark, in the absence of detector-decoder iterations. The overlapped curves of *TP Gauss* and *TP Tikhonov* in Fig. 2 demonstrate that the procedure in [1] of matching of Gaussian moments practically achieves the same results as that of matching the circular moments (9), when projecting the mixture message $p_d(\theta_k)$ onto the Tikhonov family. Remarkably, both TP algorithms get very close to the dp-BCJR benchmark after $N_{DD} = 3$ detector-decoder iterations.

Even better was the performance for the EP algorithm when $N_{DD} = 3$, since its BER curve overlaps the dp-BCJR benchmark. Despite EP is thus the best algorithm with affordable complexity, we notice that this is not the case if we constrain $N_{DD} = 1$. In fact, in the absence of detector-decoder iterations, EP seems to outperform the TP algorithms only for low to moderate E_b/N_0 values, while a sudden change of slope limits the achievable BER values in the large SNR regime. This was a clear sign of weakness, for the EP algorithm, that could be overtaken only by letting the phase-detecting and the decoding part of the FG exchange (vertical) messages, i.e., by allowing $N_{DD} > 1$ iterations.

The partial failure of EP when $N_{DD} = 1$ could not be attributed to its classical tendency to generate improper distributions, since we tested and implemented a number of strategies, taken from the literature, to cope with this issue. Unfortunately, improper message *rejection* strategies [19] were not able to solve the problem while message *damping* [10] only mitigated but did not remove the apparent error floor of EP in Fig 2.

The difficulties of EP became even more evident when referring to the concentrated pilots scenario as in Fig. 3, where we report the performance of receivers using the algorithms described above while constraining $N_{DD} = 1$, hence without any detector-decoder iteration. In this case, we assumed $\sigma_\Delta = 3$ [deg.] for the phase noise, since none of the tested receivers could achieve a good performance when $\sigma_\Delta = 6$ [deg.]. As expected, both TP algorithms failed completely, due to the concentrated pilots, while the performance of EP is still unsatisfactory, as evidenced by the BER curves, and would only marginally improve by letting $N_{DD} > 1$, as we tested. On the contrary, the Modified EP algorithm discussed in Sec. IV and labelled *EP Mod* in Fig. 3, almost coincides with the practical dp-BCJR benchmark, and is less than 0.2 dB away from the theoretical known phase limit. For this algorithm, no further iteration $N_{DD} > 1$ is needed and the receiver achieves an excellent performance with separate phase detection and decoding.

VII. Conclusions

We demonstrated, for the first time to our knowledge, the successful operation of an algorithm derived by expectation propagation in a digital receiver for transmissions over channels affected by Wiener phase noise, when no distributed pilot symbols are employed and the receiver implements a practical separation of phase detection and decoding.

By carefully analyzing the message passing procedure in a simple scenario employing LDPC and binary modulation, we found that the EP rules of computation for the messages can sometimes induce a collapse of the level of confidence (precision) in the resulting estimates, hence lead to phase slips and subsequent errors. We introduced a modification of the message projection rule of EP that, remarkably, simplifies the computational complexity of the resulting algorithm and makes it possible to detect and avoid the above spurious phase detection errors so as to reach the performance of the reference benchmark.

References

[1] G. Colavolpe, A. Barbieri, and G. Caire, "Algorithms for iterative decoding in the presence of strong phase noise," *IEEE J. Select. Areas Commun.*, vol. 23, no. 9, pp. 1748–1757, Sep. 2005.

[2] A. Barbieri and G. Colavolpe, "Simplified soft-output detection of CPM signals over coherent and phase noise channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2486–2496, Jul. 2007.

[3] G. Colavolpe, "Communications over phase noise channels: a tutorial review," *International Journal of Satellite Communications and Networking*, vol. 32, pp. 167–185, May/June 2014.

[4] S. Shayovitz and D. Raphaeli, "Message passing algorithms for phase noise tracking using tikhonov mixtures," *IEEE Transactions on Communications*, vol. 64, no. 1, pp. 387–401, 2016.

[5] A. Kreimer and D. Raphaeli, "Efficient low-complexity phase noise resistant iterative joint phase estimation and decoding algorithm," *IEEE Transactions on Communications*, vol. 66, no. 9, pp. 4199–4210, 2018.

[6] B. D. Woerner and W. E. Stark, "Trellis-coded direct-sequence spread-spectrum communications," *IEEE Trans. Commun.*, vol. 42, pp. 3163–3170, Dec. 1994.

[7] T. P. Minka, "Expectation propagation for approximate Bayesian inference," in *Proc. Conf. on Uncertainty in Artificial Intelligence (UAI)*, Seattle, Washington, USA, Aug. 2001.

[8] ——, "Divergence measures and message passing," Microsoft Research Cambridge, Tech. Rep., 2005, mSR-TR-2005-173 (2005).

[9] T. Heskes and O. Zoeter, "Expectation propagation for approximate inference in dynamic Bayesian networks," in *Proc. Conf. on Uncertainty in Artificial Intelligence (UAI)*, Edmonton, Canada, Jul. 2002.

[10] A. Vehtari, A. Gelman, T. Sivula, P. Jylänki, D. Tran, S. Sahai, P. Blomstedt, J. P. Cunningham, D. Schiminovich, and C. Robert, "Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data," 2014, [Online] https://arxiv.org/abs/1412.4869.

[11] A. Vannucci, G. Colavolpe, and L. Veltri, "Estimation of correlated gaussian samples in impulsive noise," *IEEE Communications Letters*, vol. 24, no. 1, pp. 103–107, 2020.

[12] A. Mirbadin, A. Vannucci, and G. Colavolpe, "Expectation propagation and transparent propagation in iterative signal estimation in the presence of impulsive noise," in *2020 43rd International Conference on Telecommunications and Signal Processing (TSP)*, 2020, pp. 175–179.

[13] A. Mirbadin, A. Vannucci, G. Colavolpe, R. Pecori, and L. Veltri, "Iterative receiver design for the estimation of gaussian samples in impulsive noise," *Applied Sciences*, vol. 11, no. 2, 2021.

[14] E. Conti, A. Piemontese, G. Colavolpe, and A. Vannucci, "Detection for flat-fading channels based on expectation propagation," submitted to *International Symposium on Topics in Coding (ISTC 2023)*, Brest (France), September 2023.

[15] Y. Qi and T. P. Minka, "Window-based expectation propagation for adaptive signal detection in flat-fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 348–355, 2007.

[16] E. Conti, A. Piemontese, G. Colavolpe, and A. Vannucci, "Expectation propagation for flat-fading channels," submitted to *IEEE Wireless Communications Letters*, 2023.

[17] G. Colavolpe and A. Modenini, "Iterative carrier synchronization in the absence of distributed pilots for low SNR applications," in *6th ESA International Workshop on Tracking, Telemetry and Command Systems for Space Applications (TTC'13)*. European Space Agency, 2013.

[18] L. Szczecinski, H. Bouazizi, and A. Aharony, "Parametric phase tracking via expectation propagation," 2020, [Online] https://arxiv.org/abs/2005.01844.

[19] M. Seeger, "Bayesian Gaussian process models: PAC-bayesian generalisation error bounds and sparse approximations," *PhD thesis, University of Edinburgh*, 2003.

[20] K. V. Mardia and P. E. Jupp, *Directional statistics*. Wiley Online Library, 2000, vol. 2.

[21] A. Vannucci, G. Colavolpe, R. Pecori, and L. Veltri, "Estimation of a gaussian source with memory in bursty impulsive noise," in *2019 IEEE International Symposium on Power Line Communications and its Applications (ISPLC)*, 2019, pp. 1–6.

[22] A. Mirbadin, E. Kiani, A. Vannucci, and G. Colavolpe, "Estimation of gaussian processes in markov-middleton impulsive noise," in *2019 1st Global Power, Energy and Communication Conference (GPECOM)*, 2019, pp. 68–73.

[23] D. MacKay, "http://www.inference.org.uk/mackay/," web page.