

Expectation Propagation for Flat-Fading Channels

Elisa Conti¹, Graduate Student Member, IEEE, Amina Piemontese², Giulio Colavolpe³,
and Armando Vannucci⁴

Abstract—This letter addresses the problem of signal detection in flat-fading channels. In this context, receivers based on the expectation propagation framework appear to be very promising although presenting some critical issues. We develop a new algorithm based on this framework where, unlike previous works, convergence is achieved after a single forward-backward pass, without additional inner detector iterations. The proposed message scheduling, together with novel adjustments of the approximating distributions' parameters, allows to obtain significant performance advantages with respect to the state-of-the-art solution. Simulation results show the applicability of this algorithm when sparser pilot configurations have to be adopted and a considerable gain compared to the current available strategies.

Index Terms—Expectation propagation, fading channels, parameter estimation, single-input single-output (SISO).

I. INTRODUCTION

DETECTORS for flat Rayleigh fading channels have been widely studied in the literature (e.g., see [1], [2], [3] and references therein). This kind of channels is commonly adopted because it properly represents real scenarios, such as narrowband communications over multipath wireless channels. The framework based on factor graphs (FGs) and the sum-product algorithm (SPA) provides a powerful tool for the accomplishment of both detection and decoding [4]. However, when this framework is applied jointly to detection and channel estimation, discrete and continuous random variables appear in the FG and the SPA becomes impractical.

A common approach foresees the use of canonical distributions [4], e.g., the most suitable for the analyzed scenario is the Gaussian one. A possibility is to directly reduce complicated functions to a unique Gaussian through a message projection, as performed in the Kalman smoother [5]. Although this approach is effective in pilot-aided transmission, it eventually fails when the spacing between pilot blocks is longer than a maximum accepted distance, as we will demonstrate in the numerical results.

In order to overcome this limitation, in this letter we focus on a different kind of receiver, which is based on the expectation propagation (EP) framework [6]. This message-passing algorithm is based on approximating the distributions of the

variables of interest, rather than those of the exchanged messages. This way, a *prior belief* coming from the rest of the FG drives the approximation of intractable messages. The combination of the SPA rules with the approximation deriving from EP leads to a hybrid framework which, in the literature, is commonly referred to as EP-belief propagation (EP-BP). This framework has been deeply studied and analyzed in different scenarios, including, e.g., the transmission over phase noise channels [7], the MIMO sparse codes multiple access detection [8] or joint channel estimation and decoding for faster-than-Nyquist signaling [9]. As it is inherent to the EP update strategy, improper distributions, e.g., a Gaussian with a negative variance, could appear. In [2], where, unlike other approaches [10], [11], the presence of improper distributions is accepted, the EP framework has been applied to a flat-fading channel, modeled as a complex autoregressive moving-average process. The authors proposed an EP smoothing approach where inner detector iterations are required to achieve convergence and it is tested using differential encoding. However, when using a more powerful encoder, hence working in a lower signal-to-noise ratio (SNR) regime, the approach in [2] is not equally effective and this is the focus of our investigation.

The main contributions of this letter can be summarized as follows.

- An implementation of EP-BP that overcomes its limitations and reduces the algorithmic complexity is proposed.
- An improved scheduling which allows to achieve convergence without the aid of multiple inner detector iterations (*iterative refinement* [6]) is adopted.
- A new technique for balancing the variance of the involved Gaussian distributions is introduced.
- The problem of improper distributions is addressed, the scenarios where they can compromise the algorithm operations are identified and a novel strategy to handle these events is presented.

Finally, we demonstrate through simulations that the proposed techniques present a superior performance, in terms of bit-error rate (BER) and computational complexity, with respect to state-of-the-art EP implementations [2]. Moreover, the proposed solution ensures a higher robustness against concentrated pilot distributions, laying the foundation for overcoming the limitations related to the classical and well-established Kalman filter.

II. SYSTEM MODEL

We consider the transmission of a sequence of K complex coded symbols $\mathbf{c} = (c_0, c_1, \dots, c_{K-1})$ over a flat fading channel. The sequence \mathbf{c} is the result of the binary encoding, through the code \mathcal{C} , and modulation over an M -ary constellation ($\mathcal{X} = \{x_1, \dots, x_M\} \subset \mathbb{C}$) of the information bit

Manuscript received 1 June 2023; accepted 7 July 2023. Date of publication 17 July 2023; date of current version 9 November 2023. This work was supported in part by the Italian Ministry of Universities and Research (MUR) under the PRIN Liquid_Edge Contract and in part by the University of Parma through the Action Bando di Ateneo 2022 per la ricerca co-funded by MUR D.M. 737/2021 PNR PNRR NextGenerationEU. The associate editor coordinating the review of this article and approving it for publication was N. Wu. (Corresponding author: Elisa Conti.)

The authors are with the Department of Engineering and Architecture, University of Parma, 43124 Parma, Italy (e-mail: elisa.conti@unipr.it; amina.piemontese@unipr.it; giulio.colavolpe@unipr.it; armando.vannucci@unipr.it).

Digital Object Identifier 10.1109/LWC.2023.3295952

bottom of Fig. 1 belong to the same approximating family because of the closure of the Gaussian family under both the multiplication and the convolution operations [1], [4].

A different possibility consists in approximating the distributions of variable nodes, rather than the messages, and this is the key idea on which the EP framework is based. The approximation of $p_d(h_k)$ in (7) is thus performed in EP, under the influence of a *temporary prior* distribution $p^{(p)}(\mathbf{h}_k)$ which is the product of the incoming messages to the variable node \mathbf{h}_k from the lower part of the FG, i.e., $p_{k|k-1}^{(f)}(\mathbf{h}_k)$ and $p_{k|k+1}^{(b)}(\mathbf{h}_k)$.⁴ EP is a message-passing algorithm whose approximations are generated by projecting the variable distribution on a fully-factorized exponential family \mathcal{F} , where the projection is performed by minimizing the Kullback-Leibler (KL) divergence. These two features allow to compute the projection by solving a set of *moment matching* equations [6]. In the present case of a Gaussian approximating family, the distribution resulting from the projection is thus the unique Gaussian whose mean, variance and scale factor equal those of the distribution to be approximated.

The specific algorithmic steps of EP can be found in [13]. The aim is to approximate the marginal distribution of \mathbf{h}_k which, deriving from the combination of the prior and observation messages [13], is still multimodal. Denoting with $\boldsymbol{\eta}_{MIX,m}$ and $\mathbf{C}_{MIX,m}$ the mean and covariance matrix of the m -th mode, respectively, its expression is given by [4], [13]

$$\hat{p}(\mathbf{h}_k) = \sum_{m=1}^M P_m w_m g_{\mathcal{C}}(\mathbf{h}_k, \boldsymbol{\eta}_{MIX,m}, \mathbf{C}_{MIX,m}), \quad (8)$$

where⁵ $P_m = P_d(c_k = x_m)/|x_m|^2$ and w_m is the symbol-dependent coefficient that weights the m -th mixture component. The approximation of the mixture (8) with a Gaussian $p^{(EP)}(\mathbf{h}_k) = \text{proj}_{KL}[\hat{p}(\mathbf{h}_k)]$ is performed, as usual in EP, by *matching the moments* of the distributions. The zero-th order moment is the mass of the distribution, i.e., $Z = \sum_{m=1}^M P_m w_m$. However, the mass of the messages is not significant in message-passing algorithms like SPA or EP, and it is replaced with its normalized version $p_h(\mathbf{h}_k) = Z^{-1} \hat{p}(\mathbf{h}_k)$. The matching of the first and second order moments is performed by solving⁶ $\mathbb{E}_{p^{(EP)}}[\mathbf{h}_k] = \mathbb{E}_{p_h}[\mathbf{h}_k]$ and $\mathbb{E}_{p^{(EP)}}[|\mathbf{h}_k|^2] = \mathbb{E}_{p_h}[|\mathbf{h}_k|^2]$, respectively. After few calculations, we obtain

$$\boldsymbol{\eta}_{EP} = \sum_{m=1}^M W_m \boldsymbol{\eta}_{MIX,m}, \quad (9)$$

$$\mathbf{C}_{EP} = \sum_{m=1}^M W_m \left(\mathbf{C}_{MIX,m} + \boldsymbol{\eta}_{MIX,m} \boldsymbol{\eta}_{MIX,m}^H \right) - \boldsymbol{\eta}_{EP} \boldsymbol{\eta}_{EP}^H, \quad (10)$$

where $W_m = Z^{-1} P_m w_m$. Now, it is possible to compute the parameters of the updated message $\tilde{p}_d(\mathbf{h}_k)$

$$\boldsymbol{\Sigma}_d = \boldsymbol{\Sigma}_{EP} - \boldsymbol{\Sigma}_{p,k}, \quad (11)$$

$$\boldsymbol{\Sigma}_d \boldsymbol{\eta}_d = \boldsymbol{\Sigma}_{EP} \boldsymbol{\eta}_{EP} - \boldsymbol{\Sigma}_{p,k} \boldsymbol{\eta}_{p,k}, \quad (12)$$

⁴We will refer to the precision and mean of the forward and backward messages at k with $\boldsymbol{\Sigma}_{f,k}$, $\boldsymbol{\eta}_{k|k-1}$ and $\boldsymbol{\Sigma}_{b,k}$, $\boldsymbol{\eta}_{k|k+1}$ respectively.

⁵The dependence of P_m on k is omitted for simplicity of notation.

⁶We denote by $\mathbb{E}_p[x]$ the expectation of x under the distribution $p(\cdot)$.

where $\boldsymbol{\eta}_{p,k}$ and $\boldsymbol{\Sigma}_{p,k}$ are the mean and variance of the prior distribution. The approximated message is then propagated following the SPA rules [4].

IV. PROPOSED ALGORITHM

The classical EP implementation [2] presents some critical issues which do not lead to a satisfactory performance in the considered scenario. In the following, we present novel techniques which allow to overcome the limitations related to the standard EP structure.

Damping and Boosting Factors: In order to face the EP instabilities, as discussed in the literature, a number of solutions exist, including *damping* [10], [11]. However, unlike its common use, in the analyzed scenario the introduction of this correction factor is studied specifically to provide a performance improvement through the balancing of overly confident estimates without the involvement of old updates. In fact, in our implementation, the damping factor, ξ_d , which belongs to the interval $[0, 1]$, is applied to the precision of the approximating message $\tilde{p}_d(\mathbf{h}_k)$ only. In this way, the algorithm underestimates the updates reliability as commonly done in the literature when dealing with sub-optimal algorithms. We propose to set this factor to the maximum value of $P_d(c_k)$, i.e., the probability mass function of the symbols. The main advantage of this approach is that a lower number of turbo iterations is required for achieving convergence. At the first turbo iteration⁷ (which is the only one in the separate detection and decoding case) this quantity is equal to $1/M$. On the contrary, from the 2nd iteration ahead this value becomes a clearer and clearer representation of the decoder confidence about the transmitted symbols. It follows that the closer this value is to 1, the higher the weight given to the approximating message after the KL projection.

Moreover, we propose an additional technique which allows to obtain a further performance improvement: the use of a *boosting* factor for pilot symbols. Similarly to damping, the precision (11) is multiplied by a factor that, this time, is greater than 1, so that its effect is the opposite: it increases the confidence of the observation message in correspondence with a pilot symbol, i.e., when $p_d(h_k)$ is a unimodal Gaussian distribution and, therefore, no approximation is performed. The selected value for the boosting factor is $\xi_b = 2$.

Negative Variances: Due to the subtraction in (11), it may occur that the algorithm generates negative variances, so that the approximation of the observation message, $\tilde{p}_d(\mathbf{h}_k)$, is not necessarily a valid pdf [2]. This happens when the precision of the prior distribution is larger than the one of the marginal after the KL projection. Since the focus of EP is on the marginal distribution of the channel parameters, the algorithm could accept improper distributions as messages. However, when the scheduling described in the following is adopted for the separate detection and decoding case, the only message which can present such variances is the observation one.

⁷The detector subgraph operates first by exchanging *horizontal* messages along the Markov chain in the lower part of the FG shown in Fig. 1. This message passing inside the detector can be iterated with multiple forward-backward passes (N_d) before sending *vertical* messages to the upper part of the FG (decoder). An inner message passing procedure can take place also there until convergence or until a maximum number of decoder iterations (N_c) is reached. The exchange of information between detector and decoder can be iterated for N_t times. We will refer to these iterations as *turbo* iterations.

Algorithm 1 Proposed EP Algorithm

- 1: **Initialize** the messages $p_{0|-1}^{(f)}(\mathbf{h}_0)$ and $p_{K-1|K}^{(b)}(\mathbf{h}_{K-1})$ by setting $\Sigma_{f,0}, \Sigma_{b,K-1}$ to \mathbb{I}_N and $\eta_{0|-1}, \eta_{K-1|K}$ to $\mathbf{0}$
- 2: **if** $n_{iter,T} > 1$ **then**
- 3: Set $\alpha_k(\mathbf{h}_k)$ and $\beta_k(\mathbf{h}_k)$ to $p_{k|k-1}^{(f)}(\mathbf{h}_k)$ and $p_{k|k+1}^{(b)}(\mathbf{h}_k)$ computed at the $(n_{iter,T} - 1)$ -th turbo iteration, respectively, for any value of k
- 4: **else**
- 5: Set $\alpha_k(\mathbf{h}_k)$ and $\beta_k(\mathbf{h}_k)$ to unitary messages for any value of k
- 6: **end if**
- 7: **for** $k = 0 : K - 2$ **do**
- 8: Compute the *temporary prior* $(\eta_{p,k}$ and $\Sigma_{p,k})$ by multiplying $p_{k|k-1}^{(f)}(\mathbf{h}_k)$ with $\beta_k(\mathbf{h}_k)$
- 9: Compute the parameters of the *exact marginal* distribution (i.e., for each m , compute $W_m, \eta_{MIX,m}, \mathbf{C}_{MIX,m}$ in [13])
- 10: Perform the *moment matching*, using (9)-(10)
- 11: Compute $\tilde{p}_d(\mathbf{h}_k)$ via (11)-(12) and perform either *damping* or *boosting* on Σ_d depending on k
- 12: **if** $(\Sigma_d(1,1) < 0)$ & $(n_{iter,T} > 1)$ **then**
- 13: Perform the *observation message projection*
- 14: **end if**
- 15: Compute the forward messages $p_{k|k}^{(f)}(\mathbf{h}_k)$ and $p_{k+1|k}^{(f)}(\mathbf{h}_{k+1})$ through the SPA rules [4] (refer to Fig. 1)
- 16: **end for**
- 17: **for** $k = K - 1 : -1 : 1$ **do**
- 18: Compute $p^{(p)}(\mathbf{h}_k)$ by multiplying $p_{k|k+1}^{(b)}(\mathbf{h}_k)$ and $\alpha_k(\mathbf{h}_k)$. Then, repeat the steps 9-14
- 19: Compute the backward messages $p_{k|k}^{(b)}(\mathbf{h}_k)$ and $p_{k-1|k}^{(b)}(\mathbf{h}_{k-1})$ through the SPA rules [4]
- 20: **end for**
- 21: Multiply the forward message obtained through 7-16 with the backward one computed in 17-20 for each k
- 22: Compute for each k and each $c_k \in \mathcal{X}$ the probabilities $P_u(c_k)$ to be sent to the decoder

When turbo iterations are considered, improper distributions could instead propagate giving rise to numerical problems. Hence, in this case, we propose an alternative approach to deal with negative variances, where the algorithm performs a message projection instead of a variable distribution projection [13]. The proposed method leads to a performance improvement with respect to the propagation of improper distributions and consists in approximating the message $p_d(\mathbf{h}_k)$ locally with a Gaussian distribution whose mean and variance are obtained through moment matching without the influence of the rest of the network [14], [15].

Algorithm Summary: With reference to the generic $n_{iter,T}$ turbo iteration, the proposed algorithm structure is outlined in Algorithm 1.

The forward and backward recursions lead to two different approximations for $p_d(\mathbf{h}_k)$. The main advantage of the adopted scheduling is that a satisfying performance is already achieved after a single forward-backward sweep inside the detector.

V. SIMULATION RESULTS

The performance of the proposed algorithm was analyzed in terms of 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. We considered a (3, 6)-regular rate-1/2 LDPC code with length 4000 and QPSK modulation ($M = 4$). In all simulations, in order to allow the convergence of the algorithm, pilot symbols were inserted in the transmitted

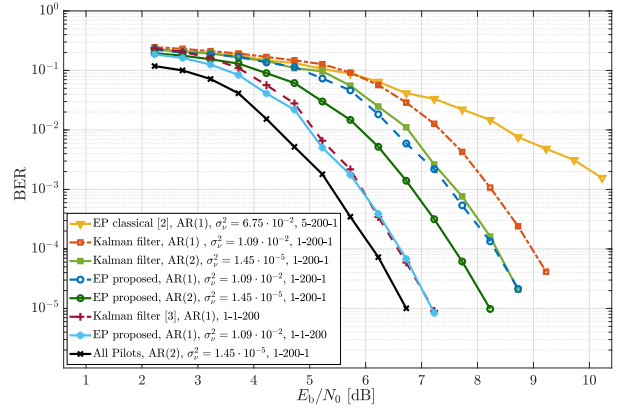


Fig. 2. Performance for the 1/20 pilots distribution with $f_D T = 10^{-2}$.

codeword.⁸ The presence of pilots involves a slight decrease of the effective information rate, resulting in an increase in the required SNR. This increase has been artificially introduced in the curve labeled “All Pilots”, inserted as ideal bound, for the sake of comparison. The curves labeled “Kalman filter” in Figs. 2 and 4 refer to the algorithm proposed in [3]. The algorithms were tested for both the AR(1) and AR(2) fading models. The values of the parameter ρ were set according to the YW equations, whereas the increment variance σ_v^2 was set to the one minimizing the BER, found through computer simulations.

In Fig. 2, we compared the performance of the proposed algorithm with both the Kalman filter and the “batch-EP” of [2], labeled with “classical EP”, when the normalized Doppler bandwidth of the fading process $f_D T$ was set to 10^{-2} and one pilot symbol was inserted in every block of 20 transmitted symbols. Firstly, we analyzed the practical case of separate detection and decoding.⁹ Figure 2 shows that the proposed EP algorithm can achieve a performance gain of approximately 0.9 dB with respect to the Kalman filter for both the AR(1) and the AR(2) models. Moreover, a significant gain is observed comparing the proposed technique with the classical EP of [2]. Regarding the latter algorithm, the authors proposed an EP smoothing approach where an iterative refinement of the observation approximations is foreseen. Consistently with that used in [2], a maximum number (N_d) of 5 inner detector iterations was allowed, and it resulted to be sufficient for achieving convergence. For this algorithm, the curve shown in Fig. 2 was obtained using the AR(1) model since the AR(2) model did not yield further improvement. The complexity of the considered algorithms was evaluated with reference to one inner detector iteration for a single time epoch k and it is reported in Fig. 3. They were compared considering the total number of required operations, including two-terms multiplications and additions. It can be noticed that the additional operations involved by the EP framework lead inevitably to a complexity increase which, relying on our implementation, is still limited and reasonably comparable to the one of the Kalman filter. On the other hand, a steeper increase is

⁸We use the notation {number of pilots in a block}/{blocks distance} to describe a pilot distribution.

⁹In order to indicate the number of inner detector, inner decoder and turbo iterations, the notation $N_d - N_c - N_t$ is used.

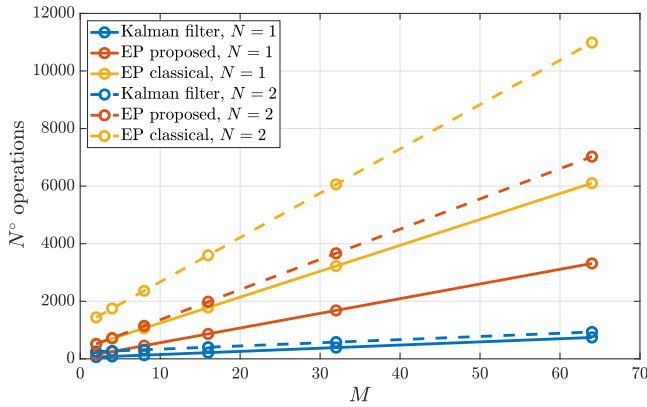


Fig. 3. Complexity comparison.

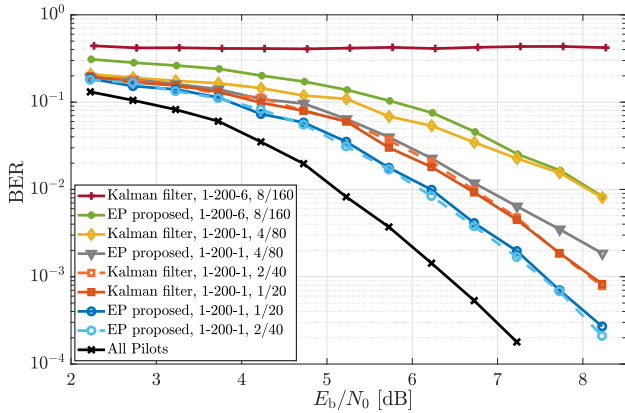


Fig. 4. Performance for different pilots distribution with $f_D T = 5 \cdot 10^{-3}$ and $\sigma_v^2 = 8 \cdot 10^{-3}$.

registered for the classical strategy [2]. In Fig. 2, the BER curves related to the joint iterative detection and decoding, where a maximum of 200 turbo iterations is allowed, are also reported with reference to the AR(1) fading model. It is possible to notice that the curves related to the proposed EP algorithm and the Kalman filter (optimized according to [3]) are approximately overlapped. In fact, when a sufficient number of turbo iterations is performed, the decoder provides symbol probabilities approaching the ones passed in correspondence with a known symbol. Therefore, the observation message is composed by a single Gaussian and the operations performed by the two detectors coincide.

In Fig. 4, keeping fixed the effective information rate, different pilot distributions were considered in order to show the higher robustness of EP with respect to the Kalman filter against pilot symbols arrangements in sequences separated by longer blocks of code symbols. The normalized Doppler bandwidth was set to $5 \cdot 10^{-3}$. The AR(1) fading model was adopted for the sake of simplicity, however, similar conclusions can be drawn for AR(2). It is possible to notice that both the proposed algorithm and the Kalman filter are practically insensitive to a change in the pilots blocks distance from 20 to 40 symbols. Differently from the EP-based algorithm, increasing the gap to 80 symbols leads to a significant loss for the Kalman filtering approach. Finally, considering a pilots distance of 160 code symbols, the EP detector is still able to provide reasonable fading estimates and, with the aid of 6 turbo iterations, it

loses approximately 2 dB from the case where the 1/20 pilots distribution is adopted. On the contrary, in the same conditions, the Kalman filter does not work because of its lower robustness against longer pilots spacing.

VI. CONCLUSION

In this letter, we proposed a new solution for signal detection over channels with flat-fading through EP. The presented method allows to solve the numerical instabilities related to the traditional procedure provided in the literature. We showed through numerical results the advantages in terms of BER that can be achieved in the separate detection and decoding case with respect to the classical Kalman filter approach and we demonstrated the higher robustness of this algorithm against more concentrated pilots distributions.

The advantages of this framework offer a variety of possibilities for new research directions, such as its application to the multiple-input multiple-output channel model and the study of more severe fading conditions, for instance frequency selectivity. Moreover, we could extend our analysis to other kinds of divergence measures for the involved approximations.

REFERENCES

- [1] H. Niu, M. Shen, J. A. Ritcey, and H. Liu, "A factor graph approach to iterative channel estimation and LDPC decoding over fading channels," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1345–1350, Jul. 2005.
- [2] 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, Jan. 2007.
- [3] A. Barbieri, A. Piemontese, and G. Colavolpe, "On the ARMA approximation for fading channels described by the Clarke model with applications to Kalman-based receivers," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 535–540, Feb. 2009.
- [4] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.
- [5] M. I. Jordan and C. Bishop, *An Introduction to Graphical Models*. Cambridge, MA, USA: MIT Press, 2002.
- [6] T. P. Minka, "Divergence measures and message passing," Microsoft Research Cambridge, Cambridge, U.K., Rep. mSR-TR-2005-173, 2005.
- [7] G. Colavolpe, E. Conti, A. Piemontese, and A. Vannucci, "The difficult road of expectation propagation towards phase noise detection," in *Proc. IEEE Intern. Conf. Commun.*, Rome, Italy, 2023, pp. 4684–4689.
- [8] W. Yuan, N. Wu, Q. Guo, Y. Li, C. Xing, and J. Kuang, "Iterative receivers for downlink MIMO-SCMA: Message passing and distributed cooperative detection," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3444–3458, May 2018.
- [9] N. Wu, W. Yuan, Q. Guo, and J. Kuang, "A hybrid BP-EP-VMP approach to joint channel estimation and decoding for FTN signaling over frequency selective fading channels," *IEEE Access*, vol. 5, pp. 6849–6858, 2017.
- [10] A. Gelman et al., "Expectation propagation as a way of life: A framework for Bayesian inference on partitioned data," 2014, *arXiv:1412.4869*.
- [11] M. Seeger, "Bayesian Gaussian process models: PAC-Bayesian generalisation error bounds and sparse approximations," Ph.D. dissertation, Inst. Adapt. Neural Comput. Divis. Inform., Univ. Edinburgh, Edinburgh, U.K., 2003.
- [12] J. Heo and K. M. Chugg, "Adaptive iterative detection for turbo codes on flat fading channels," in *Proc. WCNC*, Chicago, IL, USA, Sep. 2000, pp. 134–139.
- [13] E. Conti, A. Piemontese, G. Colavolpe, and A. Vannucci, "Detection for flat-fading channels based on expectation propagation," 2023, *arXiv:2305.19439*.
- [14] A. Vannucci, G. Colavolpe, R. Pecori, and L. Veltri, "Estimation of a Gaussian source with memory in bursty impulsive noise," in *Proc. 2019 IEEE ISPLC*, Prague, Czech Republic, 2019, pp. 1–6.
- [15] A. Vannucci, G. Colavolpe, and L. Veltri, "Estimation of correlated Gaussian samples in impulsive noise," *IEEE Commun. Letters*, vol. 24, no. 1, pp. 103–107, Jan. 2020.