Open access • Proceedings Article • DOI:10.1109/PIMRC.2019.8904409

# Doubly Iterative Turbo Equalization: Optimization through Deep Unfolding

— Source link

Serdar Sahin, Charly Poulliat, Antonio Maria Cipriano, Marie-Laure Boucheret

**Institutions:** University of Toulouse

**Published on:** 08 Sep 2019 - Personal, Indoor and Mobile Radio Communications

**Topics:** Expectation propagation, SC-FDE and Decoding methods

# Doubly Iterative Turbo Equalization: Optimization through Deep Unfolding

Serdar Şahin*†, Charly Poulliat†, Antonio Maria Cipriano* and Marie-Laure Boucheret†

* THALES, Gennevilliers, 92230, France, Email: name.surname@thalesgroup.com

† Univ. of Toulouse, IRIT-INPT, CNRS, Toulouse, 31000, France, Email: name.surname@enseeiht.fr

***Abstract*—This paper analyzes some emerging techniques from the broad area of Bayesian learning for the design of iterative receivers for single-carrier transmissions using bit-interleaved coded-modulation (BICM) in wideband channels. In particular, approximate Bayesian inference methods, such as expectation propagation (EP), and iterative signal-recovery methods, such as approximate message passing (AMP) algorithms are evaluated as frequency domain equalizers (FDE). These algorithms show that decoding performance can be improved by going beyond the established turbo-detection principles, by iterating over inner detection loops before decoding. A comparative analysis is performed for the case of quasi-static wideband communications channels, showing that the EP-based approach is more advantageous. Moreover, recent advances in structured learning are revisited for the iterative EP-based receiver by unfolding the inner detection loop, and obtaining a deep detection network with learnable parameters. To this end, a novel, mutual-information dependent learning cost function is proposed, suited to turbo detectors, and through learning, the detection performance of the deep EP network is optimized.**

***Index Terms*—SC-FDE, expectation propagation, approximate message passing, deep unfolding.**

## I. Introduction

Single-carrier frequency domain equalization (SC-FDE) is a fundamental technique for wireless communications over frequency selective channels, like in current 4G cellular systems. When considered with a bit-interleaved coded modulation (BICM) scheme, equalization can be seen as a Bayesian linear regression problem with asymptotically independent priors which benefits from soft channel decoder updates.

There is a long research track addressing this problem with turbo equalizers [1]–[3], which are derived using approximate Bayesian inference methods, such as belief propagation (BP), often with Gaussian-approximation constraint (GABP), to reduce receiver complexity. Recently, more advanced methods such as expectation propagation (EP) [4] gained interest in this context [5], by enabling the use of inner detection loops.

Besides, the need for low cost, sparsity-aware detectors for emerging compressed sensing and massive multiple-input multiple-output (MIMO) applications has led to many iterative signal-recovery (i.e. detection) methods [6]–[9]. Starting with iterative thresholding, an approximate message passing (AMP) framework was established, by iteratively performing linear vector estimation and scalar denoising, with variants such as generalized AMP (GAMP), orthogonal AMP (OAMP) and vector AMP (VAMP). There has been many theoretical analyses on these algorithms [9], [10], and some links to approximate inference algorithms (BP/GABP/EP) [11] were established. However relative behaviour of these algorithms in the turbo detection problem considered in this paper has not been fully investigated.

Moreover, recent advances in deep learning for structured models allows "unfolding" (also called "unrolling") inference loops into deep feedforward neural networks [12]. Improved signal-recovery algorithms were investigated by unfolding AMP-like algorithms [13]. It has been shown that unfolded BP [14] can improve inference by applying exponential smoothing over messages and learning its weights.

Considering these developments, this paper investigates turbo estimation with iterative approximate inference algorithms for the SC-FDE BICM context. Indeed, AMP-like algorithms are known for reducing the complexity of approximate inference methods, but in the considered SC-FDE systems, the approximate inference techniques can readily be instantiated with quasi-linear complexity. EP and some AMP-like methods are shown to have algorithmic equivalences, with differences caused by implementation heuristics, but the EP-based receiver achieves a more attractive detection performance, when calibrated optimally. Finally, deep unfolding is applied to the EP-based receiver to obtain a multi-layer detection network. A turbo-receiver oriented training cost function is proposed to optimize this receiver's parameters, with small training complexity.

This paper's contributions are as follows

1) instantiation of approximate inference and AMP-like algorithms for SC-FDE BICM,
2) theoretical and numerical comparison of these algorithms for wideband channel equalization,
3) learning for unfolded EP-based turbo-detection to achieve improved detection performance.

This paper is organized as follows. The system model is given in section II. Approximate inference algorithms are given in section III, and their links to AMP-based receivers is in section IV. A structured learning methodology is proposed in section V to optimize inference.

## II. System Model

This paper considers SC block transmissions using cyclic prefix. Using a BICM scheme, a $K_b$-bits information block $\mathbf{b}$ is encoded and then interleaved into a binary sequence $\mathbf{d}$ of length $K_d$. A memoryless modulator $\varphi$ maps this sequence to $\mathbf{x} \in \mathcal{X}^K$, with $|\mathcal{X}| = M$, $Q = \log_2 M$ and $K = K_d/Q$. This operation maps the $Q$-word $\mathbf{d}_k \triangleq [d_{Qk}, \ldots, d_{Q(k+1)-1}]$ to the symbol $x_k$, and $\varphi_q^{-1}(x_k)$ or $d_{k,q}$ are used to refer to $d_{kQ+q}$.

Assuming perfect synchronization in both time and frequency with the transmitter, and ideal channel state information, the received baseband observations are $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, with, $\mathbf{H}$ the channel matrix and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}_K, \sigma_w^2 \mathbf{I}_K)$ the additive complex circular white Gaussian noise (AWGN). $\mathbf{H}$ is a circulant matrix, whose first column is $\mathbf{h} = [h_0, \ldots, h_{L-1}, \mathbf{0}_{1,K-L}]$, $L < K$ being the channel spread. In the frequency domain,

$$\underline{\mathbf{y}} = \mathcal{F}_K \mathbf{y} = \underline{\mathbf{H}}\underline{\mathbf{x}} + \underline{\mathbf{w}}, \tag{1}$$

where $\underline{\mathbf{x}} = \mathcal{F}_K \mathbf{x}$, $\underline{\mathbf{w}} = \mathcal{F}_K \mathbf{w}$ and $\underline{\mathbf{H}} = \mathcal{F}_K \mathbf{H} \mathcal{F}_K^H$. $\mathcal{F}_K$ is the normalized $K$-DFT matrix whose elements are $[\mathcal{F}_K]_{k,l} = \exp(-2j\pi kl/K)/\sqrt{K}$, and such that $\mathcal{F}_K \mathcal{F}_K^H = \mathbf{I}_K$. Thanks to DFT properties, $\underline{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}_K, \sigma_w^2 \mathbf{I}_K)$, $\underline{\mathbf{H}} = \mathbf{Diag}(\underline{\mathbf{h}})$ with $\underline{\mathbf{h}} = \sqrt{K}\mathcal{F}_K \mathbf{h}$.

## III. Approximate Bayesian Inference

Optimal detection performance is achieved through joint maximum a posteriori (MAP) estimation of transmitted bits, $\hat{\mathbf{b}} = \arg\max_{\mathbf{b}} p(\mathbf{b}|\mathbf{y}, \mathbf{H}, \sigma_w^2)$ at the expense of a prohibitive computational cost of $\mathcal{O}(2^{K_b})$. In this section, MAP estimation is simplified with deterministic approximate inference methods such as BP or EP [15].

### *A. Maximum a posteriori detection with BP (BCJR)*

Belief propagation (BP) achieves exact MAP performance on computational graphs without loops, and otherwise, loopy BP provides sub-optimal iterative approximations, which, for BICM yields turbo detection algorithms, i.e. iterative joint detection and decoding [2]. Through the bit-wise asymptotic independence, MAP detector estimates the posterior PDF

$$p_{\text{DET}}(\mathbf{d}_k) \propto p(\mathbf{y}|\mathbf{H}, \sigma_w^2, \mathbf{d}) p_a(\mathbf{d}) \downarrow \mathbf{d}_k, \tag{2}$$

where $p_a(\mathbf{d}) = \prod_k p_a(\mathbf{d}_k) = \prod_k p_{\text{DEC}}(\mathbf{d}_k)/p_e(\mathbf{d}_k)$ is the fully-factorized extrinsic PDF, estimated at the MAP decoder, and $\downarrow \mathbf{d}_k$ denotes marginalization on $\mathbf{d}_k$. The posterior estimate at the MAP decoder is

$$p_{\text{DEC}}(\mathbf{d}_k) \propto p(\mathbf{b}|\mathbf{d}) p_e(\mathbf{d}) \downarrow \mathbf{d}_k, \tag{3}$$

with $p_e(\mathbf{d}) = \prod_k p_e(\mathbf{d}_k) = \prod_k p_{\text{DET}}(\mathbf{d}_k)/p_a(\mathbf{d}_k)$ being the fully-factorized extrinsic PDF of the MAP detector. After iterations of this turbo process, the MAP decoder estimates the posterior PDF $p_{\text{DEC}}(\mathbf{b}) \propto p(\mathbf{b}|\mathbf{d}) p_e(\mathbf{d}) \downarrow \mathbf{b}$ on $\mathbf{b}$, yielding an estimate $\hat{\mathbf{b}}$.

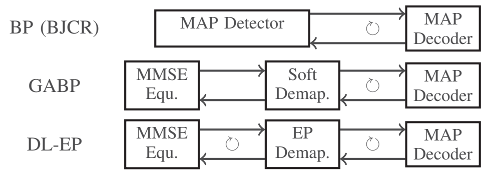


Fig. 1. Turbo-equalization: BP and GABP vs. double-loop EP.

In practice, the decoder and the detector exchange extrinsic information through log likelihood ratios (LLRs) of coded bits $\mathbf{d}$. A priori, extrinsic and a posteriori LLRs are respectively denoted $L_a(\cdot)$, $L_e(\cdot)$ and $L(\cdot)$, with respect to the detector. Prior LLRs characterize the prior PDF $p_a(\mathbf{d}_k)$, as follows

$$\begin{aligned} p_a^{(\tau)}(\mathbf{d}_k) &\propto \textstyle\sum_{\alpha \in \mathcal{X}} \mathcal{P}_k^{(\tau)}(\alpha) \delta(\varphi(\mathbf{d}_k) - \alpha), \\ \mathcal{P}_k^{(\tau)}(\alpha) &\propto \textstyle\prod_{q=0}^{Q-1} \exp(-\varphi_q^{-1}(\alpha) L_a^{(\tau)}(d_{k,q})), \end{aligned} \tag{4}$$

where $\mathcal{P}_k^{(\tau)}$ is a probability mass function (PMF) on $x_k$ and $\tau = 0, \ldots, \mathcal{T}$ is the ongoing turbo-iteration index.

The MAP detector in eq. (2) can be implemented with a BCJR algorithm with a computational complexity of $\mathcal{O}(KM^L)$. This is prohibitive for high order constellations or for channels with large delay spread. In the following, the use of a MAP decoder is maintained.

### *B. Expectation Propagation (EP)*

To alleviate the computational costs of MAP detection, filter-based turbo equalizers were explored, with minimum mean squared-error (MMSE)-like criteria and interference cancellation, using GABP [1], [3]. Nevertheless, the decoding threshold of the GABP is far from those of joint MAP detector, especially in highly selective channels. Non-linear approaches were evaluated to improve turbo equalization, by using detection information ("decision feedback") on symbols, on top of decoder information.

Among those, interesting structures were given by the EP framework, a more general approximate inference algorithm [4], compared to GABP. EP attributes a family of probability distributions to each factor involved in the inference. The posterior estimates of these factors are approximated by the reverse information projection [15] onto their respective family of distributions, to enable computing closed-form marginals.

When working with variables that follow Gaussian distribution, EP yields MMSE-like filtering and interference cancellation structure. Moreover, the structure of the covariance matrix of these PDFs has an impact on the detection complexity and performance. One practical example is the scalar EP (SEP), which considers variable PDFs to be fully-factorized with a scalar covariance (i.e. white estimation noise). Scalar EP (SEP) is used for low-complexity turbo equalization in [5], as it can be implemented with fast Fourier transforms (FFT).

This receiver performs inner self-iterations $s = 0, \ldots, \mathcal{S}$ over filtering and demapping. In this loop, first, the demapper estimates posterior symbols distribution

$$\mathcal{D}_k^{(\tau,s)}(\alpha) \propto \exp\left(-|x_k^{e(\tau,s)} - \alpha|^2/v^{e(\tau,s)}\right)\mathcal{P}_k^{(\tau,s)}(\alpha),$$

with the mean and variance of the PMF $\mathcal{D}_k^{(\tau,s)}$ being

$$\begin{aligned} \mu_k^{d(\tau,s)} &\triangleq \mathbb{E}_{\mathcal{D}}[x_k] = \textstyle\sum_{\alpha\in\mathcal{X}} \alpha\mathcal{D}_k^{(\tau,s)}(\alpha), \\ \gamma^{d(\tau,s)} &\triangleq K^{-1}\textstyle\sum_k \mathrm{Var}_{\mathcal{D}}[x_k]. \end{aligned} \tag{5}$$

The soft feedback to the receiver is obtained in this case by the division of the PDF $\mathcal{CN}(\mu_{p,k}^d, \gamma_p^d)$, by the equalized symbol PDF $\mathcal{CN}(x_{p,k}^e, v_p^e)$, yielding

$$\frac{x_k^{\star(\tau,s+1)}}{v^{\star(\tau,s+1)}} \triangleq \frac{\mu_k^{d(\tau,s)}}{\gamma^{d(\tau,s)}} - \frac{x_k^{e(\tau,s)}}{v^{e(\tau,s)}}, \tag{6}$$

$$1/v^{\star(\tau,s+1)} \triangleq 1/\gamma^{d(\tau,s)} - 1/v^{e(\tau,s)}. \tag{7}$$

However, using these raw estimates may lead to undesirable local extrema [15], hence smoothing is used

$$x_k^{d(\tau,s)} = (1-\beta)x_k^{\star(\tau,s)} + \beta x_k^{d(\tau,s-1)}, \tag{8}$$

$$v^{d(\tau,s)} = (1-\beta)v^{\star(\tau,s)} + \beta v^{d(\tau,s-1)}, \tag{9}$$

with $0 \le \beta \le 1$. Next, MMSE-like filtering is used

$$\begin{aligned} \xi^{(\tau,s)} &= K^{-1}\textstyle\sum_k |\underline{h}_k|^2/(\sigma_w^2 + v^{d(\tau,s)}|\underline{h}_k|^2), \\ \underline{f}_k^{(\tau,s)} &= \underline{h}_k/[\xi^{(\tau,s)}(\sigma_w^2 + v^{d(\tau,s)}|\underline{h}_k|^2)], \\ \underline{x}_k^{e(\tau,s)} &= \underline{x}_k^{d(\tau,s)} + \underline{f}_k^{(\tau,s)*}(\underline{y}_k - \underline{h}_k\underline{x}_k^{d(\tau,s)}), \\ v^{e(\tau,s)} &= 1/\xi^{(\tau,s)} - v^{d(\tau,s)}, \end{aligned} \tag{10}$$

which completes an inner iteration.

At the final self-iteration, extrinsic LLRs are

$$L_e^{(\tau)}(d_{k,q}) = \ln\frac{\sum_{\alpha\in\mathcal{X}_q^0}\mathcal{D}_k^{(\tau,\mathcal{S})}(\alpha)}{\sum_{\alpha\in\mathcal{X}_q^1}\mathcal{D}_k^{(\tau,\mathcal{S})}(\alpha)} - L_a^{(\tau)}(d_{k,q}), \tag{11}$$

with $\mathcal{X}_q^b = \{\alpha\in\mathcal{X} : \varphi_q^{-1}(x) = b\}$, $b \in \mathbb{F}_2$. These LLRs are then processed by the MAP decoder to produce the next turbo iteration's prior LLRs $L_a^{(\tau+1)}(d_{k,q})$.

This double-loop scalar EP (DL-SEP) receiver has a complexity of $\mathcal{O}(\mathcal{S}K\log_2 K)$, and interference cancellation with extrinsic feedback instead of APP significantly improves the convergence speed and the asymptotic performance [5]. When $\mathcal{S} = 0$, DL-SEP coincides with GABP [3] and the structural differences between such structures is shown on Fig. 1.

## IV. Links with Approximate Message Passing

There is a great number of contributions on iterative message-passing algorithms for low complexity parsimonious detection. This section discusses the extension of these algorithms for SC-FDE BICM detection, and establishes their theoretical links to the inference methods. These algorithms have roughly the same computational cost order of $\mathcal{O}(\mathcal{S}K\log_2 K)$.

AMP algorithms do not address (de)mapping aspects of turbo detection, and they output APP symbol estimates. If these are directly fed to a soft-demapper, the decoding performance is significantly degraded [5], hence here, the final APP estimation step is replaced with extrinsic bit LLR demapper.

### *A. Generalized Approximate Message Passing (GAMP)*

Early AMP-like techniques are built around iterative thresholding which successively applies non-linear estimators and interference cancellers. AMP is derived by applying central-limit theorem to BP messages and keeping first order terms of its Taylor series expansion, and it was later extended to GAMP [7] to handle more general inference models. It is also possible to derive GAMP from EP with similar approximations [11] .

A GAMP based FDE has been formulated in [16],

$$\begin{aligned} \xi^{(\tau,s)} &= K^{-1}\textstyle\sum_k |\underline{h}_k|^2/(\sigma_w^2 + \gamma^{d(\tau,s)}|\underline{h}_k|^2), \\ \underline{f}_k^{(\tau,s)} &= \underline{h}_k/[\xi^{(\tau,s)}(\sigma_w^2 + \gamma^{d(\tau,s)}|\underline{h}_k|^2)], \\ \underline{\epsilon}_k^{d(\tau,s)} &= \gamma^{d(\tau,s)}/v^{e(\tau,s-1)}(\underline{x}_k^{e(\tau,s-1)} - \underline{\mu}_k^{d(\tau,s-1)}), \\ \underline{x}_k^{e(\tau,s)} &= \underline{\mu}_k^{d(\tau,s)} + \underline{f}_k^{(\tau,s)*}(\underline{y}_k - \underline{h}_k(\underline{\mu}_k^{d(\tau,s)} - \underline{\epsilon}_k^{d(\tau,s)})), \\ v^{e(\tau,s)} &= 1/\xi^{(\tau,s)}. \end{aligned} \tag{12}$$

This algorithm uses interference cancellation with APP PMF statistics, but the Onsager reaction term [6] appears as a bias compensator on the feedback with $\underline{\epsilon}_k^d$. This quantity is proportional to the estimation error between the linear and the non-linear components of the previous iteration. Hence GAMP is an APP-based interference canceller that aims to decorrelate posterior estimates.

### *B. Orthogonal Approximate Message Passing (OAMP)*

OAMP [8] extends AMP, with a decorrelated linear component, and a divergence-free non-linear component (i.e. its derivative's expected-value is null). There are many estimators based on zero-forcing or matched-filtering which satisfy these conditions, but the optimal solution in the MMSE sense, coincides with a scalar EP-like algorithm. Here, OAMP is cast into SC FDE model

$$\begin{aligned} \xi^{(\tau,s)} &= K^{-1}\textstyle\sum_k |\underline{h}_k|^2/(\sigma_w^2 + \hat{v}^{\star(\tau,s)}|\underline{h}_k|^2), \\ \underline{f}_k^{(\tau,s)} &= \underline{h}_k/[\xi^{(\tau,s)}(\sigma_w^2 + \hat{v}^{\star(\tau,s)}|\underline{h}_k|^2)], \\ \underline{x}_k^{e(\tau,s)} &= \underline{x}_k^{\star(\tau,s)} + \underline{f}_k^{(\tau,s)*}(\underline{y}_k - \underline{h}_k\underline{x}_k^{\star(\tau,s)}), \\ v^{e(\tau,s)} &= 1/\xi^{(\tau,s)} - \hat{v}^{\star(\tau,s)}, \end{aligned} \tag{13}$$

with $\hat{v}^{\star(\tau,s)} \triangleq [\sum_k |\underline{y}_k - \underline{h}_k\underline{x}_k^{\star(\tau,s)}|^2 - K\sigma_w^2]/\sum_k |\underline{h}_k|^2$.

The major differences of OAMP to DL-SEP are the lack of damping and the use of an ML estimator to estimate feedback's variance.

### *C. Vector Approximate Message Passing (VAMP)*

Another EP-related AMP derivation is given by VAMP [9], with a general MMSE implementation (similar to

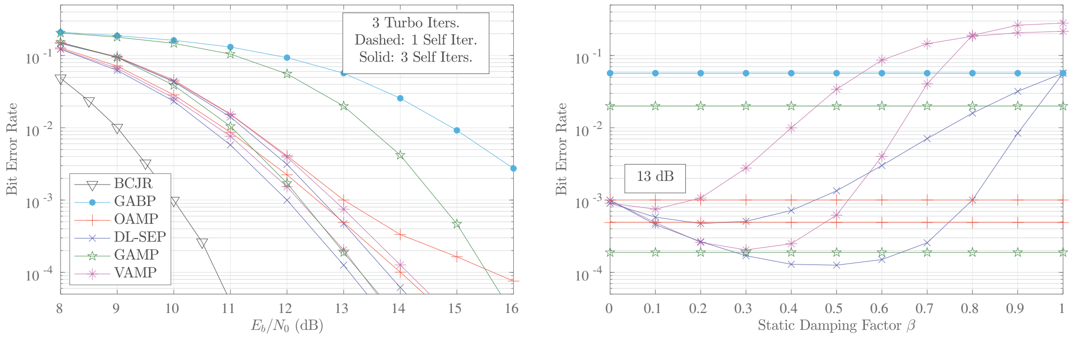


Fig. 2. BER for coded 8-PSK, with static BER-optimized damping (at left). Impact of damping on BER (at right).

OAMP) directly derived from EP. The main distinguishing aspect of VAMP is its singular value decomposition (SVD) implementation and its damping heuristics [9].

MMSE-based VAMP, implemented in the frequency-domain model is equivalent to the DL-SEP except for the damping. VAMP damping procedure (eqs. (26)-(27) in [9]) applies exponential smoothing on the non-linear APP estimate $\mu_k^d$, and on the linear estimate's precision $1/v^e$, unlike DL-SEP which smooths extrinsic non-linear estimate's mean and variance.

### D. Conclusions on AMP-like algorithms

AMP-like methods have been derived to reduce the complexity of original inference algorithms, but for the considered communications problem, inference methods have similar complexity. Bit error rate (BER) performance of turbo receivers based on AMP-like algorithms and inference methods are provided in Fig. 2, with a block length $K = 256$ and a recursive systematic convolutional (RSC) channel code $[1, 5/7]_8$, in the Proakis C channel $[0.23, 0.46, 0.69, 0.46, 0.23]$.

GABP and BCJR algorithms provide respectively an upper and a lower bound on achievable BER performance as conventional methods. Receivers that involve damping parameters are optimized, by brute-force, for each value of SNR, $\mathcal{S}$ and $\mathcal{T}$. Self-iterated receivers considerably improve the detection performance, and DL-SEP achieves the lowest error rates among alternatives. While GAMP, OAMP and VAMP approach DL-SEP performance as self-iterations increase, GAMP has slower convergence speed and OAMP has a diversity loss at high SNR, due to the sub-optimal feedback variance estimation. The right side of Fig. 2 shows the sensitivity of BER to the changes in a static $\beta$, showing that there are locally robust optimum damping values.

## V. Unfolding the Graph: Deep EP Network

### A. Motivations for deep learning at the physical layer

As physical layer emitters are man-made, model-based algorithms for receivers are expected to be robust if suitable physical channel models are available. Purely data-based deep learning strategies have been investigated for designing communication systems with auto-encoders, however practical interests remain limited, especially in the coding area, where training costs can be prohibitive [17]. Instead, model-oriented learning is more practical, when considered for optimizing existing algorithms' hyperparameters, or to account for poorly modelled channel, correlations, and system phenomena.

Deep unfolding [12] is one such strategy, which represents iterative algorithms as multi-layer deep feedforward networks, with parameters to be optimized. It has been applied to BP with exponential smoothing for improved channel decoding [14], and for unfolded-OAMP for MIMO detection [18], where the attenuation of non-linear estimations is trained. These works have shown performance benefits of learning damping parameters.

On the other hand, when the "raw" VAMP algorithm (no damping) is unfolded, the resulting deep network, where filters are replaced by fully connected layers, outperform conventional residual deep networks [13]. Moreover, analytically computed VAMP parameters yield the same performance as the trained VAMP network [13]. This suggests that raw VAMP/OAMP/SEP-like algorithms ($\beta = 0$) already yield near-optimal parameters for such structures, when channel parameters are available, and learning the involved filter/convolutional weights and non-linear parameters appears to be unnecessary.

However, as Fig. 2 attests, exponential smoothing improves these algorithms, and the BER sensitivity is a smooth function with a local extrema on $\beta$. Hence, unfolding the detection graph of DL-SEP is investigated, by considering each self-iteration as a neural layer, and learning parameters $\boldsymbol{\theta} = [\beta^{(1)}, \ldots, \beta^{(\mathcal{S})}]$ as shown in Figs. 3 and 4. The corresponding neural structure is akin to a network of convolutional layers, where layer outputs are linearly mixed with its inputs, with weight $\beta$.

### B. Learning for the deep EP network

To optimize DL-SEP, a loss function $\mathcal{L}$ is proposed, to track the turbo detection dynamics at the detector, considering the decoder outputs. The main idea is to use

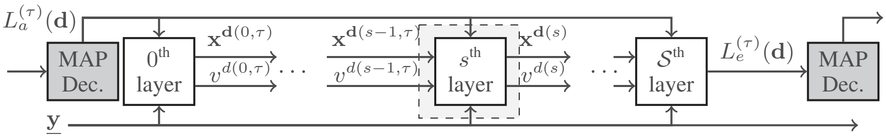


Fig. 3. “Learned-DL-SEP”: Unfolded deep EP network at the $\tau^{\text{th}}$ turbo iteration.

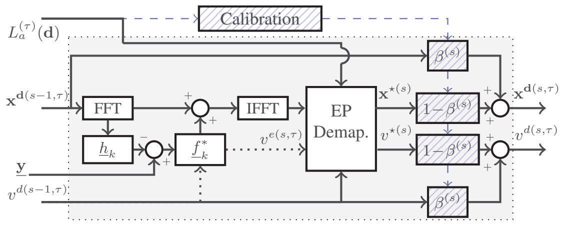


Fig. 4. “Learned-DL-SEP”: $s^{\text{th}}$ neural equalization layer.

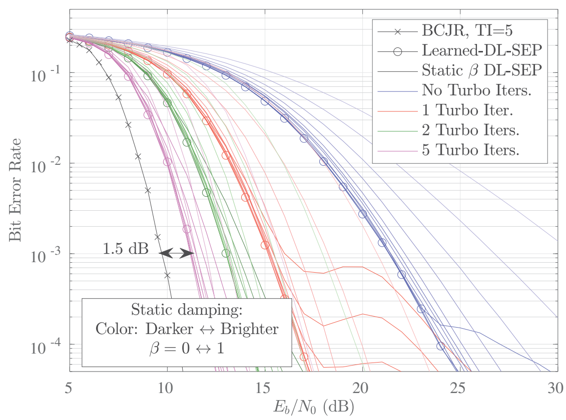


Fig. 5. Comparison, in Proakis C with coded 8-PSK, of DL-SEP with static damping and 3 self-iters. and Learned-DL-SEP with 3 layers.

a loss metric, correlated to the output BER, along with a quality indicator on a priori LLRs fed from the decoder.

The binary-cross entropy between a soft-bit outputs of DL-SEP with the transmitted bits is given as

$$\ell(d_{k,q}, \hat{d}_{k,q}) \triangleq -\log\left((\hat{d}_{k,q})^{(1-d_{k,q})}(1-\hat{d}_{k,q})^{d_{k,q}}\right),$$

where $\hat{d}_{k,q} \triangleq 1/(1+\exp(-L_e(d_{k,q}))$ is the soft bit, and this loss function corresponds to the Kullback-Leibler divergence between transmitted and estimated bits. Then, inspired from EXIT function synthesis methodology [19], the neural network is fed with a set of $N$ sample codewords of Gaussian-distributed prior LLRs, corresponding to a prior information $I_A$, $L_a(d_{k,q}, I_A)[n] \sim \mathcal{N}((1-2d_{k,q})\mu_a, 2\mu_a)$, $n = 1, \ldots, N$, with $\mu_a = J^{-1}(I_a)$ and $J(\mu) \triangleq 1 - \mathbb{E}_{L\sim\mathcal{N}(\mu,2\mu)}[\log_2(1+e^{-L})]$. Then the detector's extrinsic LLRs, for these samples, are $L_e(\mathbf{d})[I_A, n]$, and the learning cost function is

$$\mathcal{L}(\mathbf{d}, \hat{\mathbf{d}}, I_A) \triangleq \frac{1}{QNK}\sum_{k,q}\sum_{n}\ell(d_{k,q}, \hat{d}_{k,q}[I_A, n]), \quad (14)$$

where $\hat{d}_{k,q}[I_A, n]$ is the soft-bit related to $L_e(\mathbf{d})[I_A, n]$.

This loss function enables learning optimal values of $\boldsymbol{\theta}$ for a given $I_A$, and there is a bijective mapping between $I_A$ and the prior variance $v^{a(\tau)} \triangleq K^{-1}\sum_k \text{Var}_{\mathcal{P}_k^{(\tau)}}[x_k]$, where $\mathcal{P}_k^{(\tau)}$ is the PMF in eq. (4). Thus, trained parameters are tabulated as a function of $v^a \in [0, \sigma_x^2]$, and the receiver adjusts its weights $\boldsymbol{\theta}^{(\tau)}$, with the measured $v^{a(\tau)}$ and linear interpolation, at the ongoing turbo-iteration $\tau$.

Training is carried out with the ADAM optimizer [20], with an initial learning rate of 0.025 and mini-batches consisting of 200 samples of:

- a value of $\sigma_w^2$, from uniformly distributed $\text{SNR}_c = 20\log_{10}\sigma_x/\sigma_w$, on an interval of interest,
- a dummy codeword $\mathbf{d}$ from $2^{K_d}$ i.i.d. possibilities,
- a noise vector $\mathbf{w}$ and a channel realization $\mathbf{H}$,
- prior LLR codeword $L_a(\mathbf{d}, I_a)[n]$ realizations, with $n = 1, \ldots, N$ samples.

This learning strategy enables fine optimization of the DL-SEP algorithm, when considered as a deep network.

### C. Numerical results

Hence, we investigate the proposed unfolded DL-SEP, to check whether deep learning can automatically optimize DL-SEP parameters to predict its optimum behaviour. For considering a highly-selective situation, training is considered in the fixed Proakis C channel, with $K = 256$, RSC $[1, 5/7]_8$ and $\text{SNR}_c \in [5, 20]$ dB. Prior LLRs realization samples of $N = 25$ is found to be sufficient, with 150 training iterations, to have a precision within 0.05 on $\beta$. Weights are learned for $I_A \in \{0, 0.33, 0.67, 0.78, 0.89, 0.94, 0.99, 1\}$.

The performance of “Learned-DL-SEP” is shown in Fig. 5 along with DL-SEP with static damping $\beta$ (across self-iterations), with $\beta$ varying between 0 and 1 with 0.1 steps. While DL-SEP with low damping has good detection threshold, it suffers from error propagation at high SNR, oppositely high damping slows down convergence. “Learned-DL-SEP” manages thus to dynamically adapt to the situation, as deep learning allows us to find optimal values of $\beta^{(s,\tau)}$ as a function of $v^{a(\tau)}$ (dynamic damping). In Fig. 5, DL-SEP appears to reach the convex-hull of its feasible set of BER performance. In the end, the “learned-DL-SEP” with 3 layers, is within 1.5 dB of BCJR, at BER $= 10^{-3}$ for $\mathcal{T} = 5$.

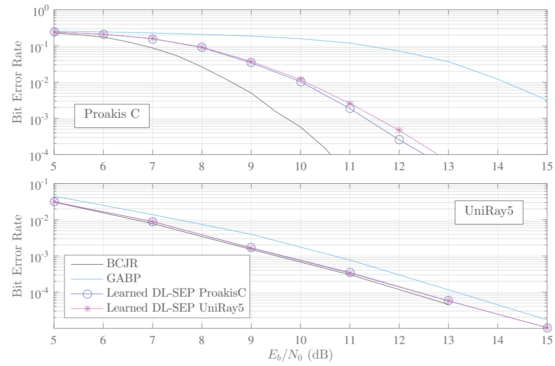


Fig. 6. Comparison of Learned-DL-SEP with 3 layers for 8-PSK, trained and evaluated in two different channels, at 5 turbo-iterations.

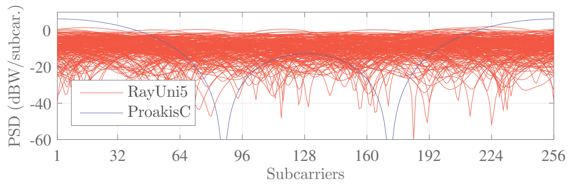


Fig. 7. Power spectral density for Proakis C and UniRay5 channels.

To pursue the analysis of the Deep EP network, training is now carried out on a Rayleigh fading channel with uniform power profile with $L = 5$ (denoted UniRay5), and hence, the learning process accounts for many different channel realizations. In previous works, it was noted that training in variable channels still ensures good detection performance for difficult channels (ill-conditioned channel matrix) [21]. In this work, we are also interested in seeing how would a network, trained in a difficult channel, would perform in random channels.

To evaluate this, in Fig. 6, the performance of Proakis-C-trained network and UniRay5-trained network are compared in both Proakis C and UniRay5 validation sets. It is shown that while UniRay5-trained network performs within 0.3-0.5 dB of the Proakis C trained network's BER, in the Proakis C channel, the Proakis-C-trained network performs identically to the UniRay5-trained network. Fig. 7 shows the power spectral density of 200 random UniRay5 channels, and the Proakis C channel; the latter has a significant spectral null region. This suggests that training sets with highly selective, difficult channels should possibly enable a learned receiver to perform near-optimally also in less selective channels.

## VI. Conclusion

Similarities between EP-based approximate-inference and AMP-like algorithms are laid out, and for the considered frequency domain equalization problem, EP-based inference is shown to reach lower error rates among other self-iterated turbo-receivers. These structures are applicable to many single-carrier communications systems, such as SC-FDE or SC-FDMA, and it can also be extended multi-user or MIMO systems [22].

Deep unfolding is shown to be a means to optimize the performance of this receiver with relative ease, and a reasonable complexity. This is enabled by the proposed turbo-oriented learning loss function, whose utility goes beyond the scope of this paper, to any soft-input soft-output detector. Finally, the impact of choosing the training set in more or less mild conditions is shown to impact the scope of optimality of such receivers.

## References

[1] J. Boutros and G. Caire, "Iterative multiuser joint decoding: unified framework and asymptotic analysis," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 1772–1793, Jul. 2002.

[2] C. Douillard, M. Jézéquel, C. Berrou, A. Picart *et al.*, "Iterative correction of intersymbol interference: Turbo-equalization," *Eur. Trans. Telecomm.*, vol. 6, no. 5, pp. 507–511, Sep. 1995.

[3] M. Tüchler and A. C. Singer, "Turbo equalization: An overview," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 920–952, Feb. 2011.

[4] T. P. Minka, "A family of algorithms for approximate bayesian inference," Ph.D. dissertation, M.I.T., Jan. 2001.

[5] S. Şahin, A. M. Cipriano, C. Poulliat *et al.*, "A framework for iterative frequency domain EP-based receiver design," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6478–6493, Dec. 2018.

[6] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[7] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE Proc. in Int. Symp. on Information Theory (ISIT)*, Jul. 2011, pp. 2168–2172.

[8] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, Jan. 2017.

[9] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *arXiv pre-print:1610.03082v2*, Jun. 2018.

[10] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," in *IEEE Proc. in ISIT'17*, Jun. 2017, pp. 501–505.

[11] X. Meng, S. Wu, L. Kuang, and J. Lu, "An expectation propagation perspective on approximate message passing," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1194–1197, Aug. 2015.

[12] J. L. J. R. Hershey and F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," Mitsubishi Electric Research Labs, Tech. Rep. TR2014-117, 2014, version 12.0.1.

[13] M. Borgerding, P. Schniter, and S. Rangan, "AMP-inspired deep networks for sparse linear inverse problems," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4293–4308, Aug. 2017.

[14] E. Nachmani, Y. Be'ery, and D. Burshtein, "Learning to decode linear codes using deep learning," in *54th Annual Allerton Conf. on Commun., Control, and Computing*, Sep. 2016, pp. 341–346.

[15] T. Minka *et al.*, "Divergence measures and message passing," Microsoft Research, Tech. Rep., 2005.

[16] Q. Guo, D. Huang, S. Nordholm *et al.*, "Iterative frequency domain equalization with generalized approximate message passing," *IEEE Sig. Proc. Lett.*, vol. 20, no. 6, pp. 559–562, Jun. 2013.

[17] T. OShea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[18] H. He *et al.*, "A model-driven deep learning network for MIMO detection," in *Proc. IEEE GlobalSIP*, Nov. 2018, pp. 584–588.

[19] S. Ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," in *AEU Int. J. Electron. Commun.*, vol. 54, no. 6, Jan. 2000.

[20] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, May 2015.

[21] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *Proc. IEEE SPAWC*, Jul. 2017, pp. 1–5.

[22] S. Şahin, C. Poulliat, A. M. Cipriano, and M.-L. Boucheret, "Spectrally efficient iterative MU-MIMO receiver for SC-FDMA based on EP," in *IEEE Proc. in PIMRC'18*, Sep. 2018.