JS 20

FUTURE NETWORKS: 5G AND BEYOND

Optimization and Analysis of Deep Unfolding Based Double Loop Turbo Equalizers

***Serdar Şahin*[1], *Antonio M. Cipriano*[1], *and Charly Poulliat*[2]**

[1] *Thales, {serdar.sahin, antonio.cipriano}@thalesgroup.com*
[2] *Toulouse INP, IRIT-ENSEEIHT, charly.poulliat@enseeiht.fr*

Keywords: inter-symbol interference, iterative equalization, expectation propagation, deep unfolding

Abstract:
This paper investigates the use of hybrid model-and-data-based deep learning on a recently proposed doubly-iterative turbo equalizer for handling inter-symbol interference (ISI) channel with single-carrier frequency domain equalization (SC-FDE). The receiver is obtained through a message-passing-based approximate Bayesian inference technique, known as expectation propagation (EP). Although this turbo-equalizer has been shown to behave asymptotically like maximum a posteriori (MAP) detection, finite-length numerical results suffer from drawbacks due to simplifying assumptions used during the modelling. Such limitations are partially mitigated by tuning heuristic hyper-parameters through robust learning algorithms. In this article, this strategy is further investigated with discussion on optimized parameters and with the use of an alternative loss function for training, or by adding further capabilities to adapt learned parameters to the channel state information.

1 Introduction

The potential impact of deep learning techniques for addressing interference mitigation problems at the physical layer is a hot topic. While data-driven artificial intelligence (AI) techniques with "black-box" neural networks have proven their worth in various signal processing fields with a modelling deficit, such as image or audio processing, their use for classical digital communications problems at the physical layer is questionable. The transmitter and channel models for most wireless systems are often accurate enough for the derivation of near-optimal signal processing algorithms, however resulting solutions may suffer from excessive computational complexity. In this context, hybrid AI techniques are of interest, by locally using learnable functions or parameters within a model-based algorithm, for handling modelling deficiency or for reducing algorithm complexity [1]. These techniques also have the advantage of requiring lower training complexity and their behaviour is more explainable than complex neural-network structures [2, 3, 4].
Here we study near-optimal detection and decoding of single-carrier transmissions with bit interleaved coded modulation (BICM). There is a long research track addressing this problem with iterative turbo detection and decoding techniques that provide reasonable performance and computational complexity trade-off [5, 6, 7]. More recently, doubly-iterative soft-input soft-output (SISO) detectors have gathered attention, and they are derived by exploiting Bayesian learning with expectation propagation (EP) [8].
Frequency domain equalizers (FDEs) obtained with such structures share similarities with approximate message passing (AMP) techniques [9], and, by using the deep unfolding approach [4], they can be seen as neural networks with some free hyper-parameters to be tuned.

2 System Model and EP-based Equalization

2.1 SC-FDE Transmission Model

This paper considers single-carrier block transmissions, circularized with 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 φ 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 $\mathbf{d}_k \triangleq [d_{Q(k-1)+1}, \ldots, d_{Qk}]$ to the symbol x_k, and $\varphi_q^{-1}(x_k)$ or $d_{k,q}$ are used to refer to d_{kQ+q}. $\mathcal{X}$ is such that independently and identically distributed (IID) data symbols have a zero-mean and unit variance, i.e. $\sigma_x^2 = 1$.
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, $\mathbf{H}$ being the channel matrix and $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}_K, \sigma_w^2\mathbf{I}_K)$ is the additive complex circularly symmetric white Gaussian noise (AWGN). $\mathbf{H}$ is a circulant

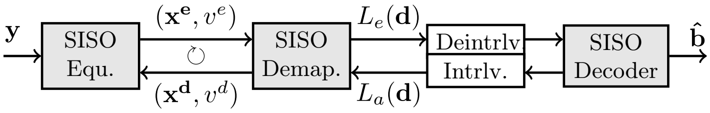

Figure 1 – Doubly-iterative EP-based receiver structure.

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, the observation are

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

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-point discrete Fourier transform (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, we have $\underline{\mathbf{w}} \sim \mathcal{CN}(\mathbf{0}_K, \sigma_w^2 \mathbf{I}_K)$ and $\underline{\mathbf{H}} = \mathbf{Diag}(\underline{\mathbf{h}})$ with $\underline{\mathbf{h}} = \sqrt{K}\mathcal{F}_K \mathbf{h}$.
At the receiver, we consider the use of an iterative detection and decoding scheme, where a SISO decoder exchanges extrinsic information with a SISO detector, through log-likelihood ratios (LLRs) on coded bits, thanks to BICM. From the SISO detector's point of view, prior and extrinsic LLRs are respectively denoted as $L_a(d_{k,q})$ and $L_e(d_{k,q})$, and the prior probability mass function (PMF) on a symbol x_k is denoted as $\mathcal{P}_k(\alpha)$, $\forall \alpha \in \mathcal{X}$. Iterations between these two modules will be denoted with $\tau = 0, \ldots, \mathcal{T}$ in the superscript, with $\mathcal{T}$ being the fixed total number of turbo-iterations.

2.2 EP-based Iterative Frequency Domain Equalizer

The optimal detection and decoding solution for the considered communication system is well-known to be given by the maximum a posteriori (MAP) criterion over the posterior probability density function (PDF) $p(\mathbf{b}, \mathbf{d}, \mathbf{x}|\underline{\mathbf{y}})$ marginalized on $\mathbf{b}$. The use of turbo detection through BICM facilitates this problem by decoupling the problem of detection and decoding thanks to SISO modules, which carry out MAP inference to estimate the coded block $\mathbf{d}$, under the assumption of using a sufficiently long interleaver along with memoryless mapping.
Nevertheless the computational and memory complexity of SISO MAP detection alone remains too high for the problem of equalization, due to exponentially growing number of states of the optimal BCJR algorithm. To alleviate this approximate Bayesian inference is performed with message passing algorithms such as Gaussian-approximated belief propagation (GaBP) or expectation propagation (EP), which yield filter-based turbo equalizers with lower complexity. In particular it is possible to derive low-complexity detectors (e.g. complexity scaling log-linearly in K) in the FD through the use of fast Fourier transform (FFT) algorithm. In the remainder of this paper we will discuss a recently proposed EP-based double-loop frequency domain equalizer who has attractive properties compared to alternative AMP techniques in the same category [9].
The operating principle of this SISO detector is given in Figure 1, where the detector itself consists of a SISO equalizer, which is an minimum-mean square error-like (MMSE-like) filter, exchanges soft estimates on data symbols with a SISO demapper. This technique major advantage compared to alternative iterative FDE is that soft data estimates $(\mathbf{x^e}, v^e)$ and $(\mathbf{x^d}, v^d)$ correspond to extrinsic information in the conventional turbo principle, and their correlation and orthogonality properties ensure good and predictable convergence properties for the iterations between the equalizer and the demapper [10]. Soft estimates from the demapper enable the equalizer to more adequately remove interference from the observations, while limiting the possible bias to its own outputs, and thus improving the equalization performance. The detailed description of this detection algorithm, called double-loop scalar EP (DL-SEP), is provided in Algorithm 1.

3 Unfolded Doubly-Iterative Turbo Equalization

Deep unfolding is a machine learning technique which enables deriving a multi-layer neural structure based on an iterative algorithm. The parameters of the original algorithm can become learnable parameters, or additional learnable parameters can be added based on expert insight on the involved heuristics and loose approximations on the original algorithm. This approach has been successfully applied to belief propagation (BP) based decoding algorithms by applying smoothing weights on exchanged messages [12]. On the other hand, in [13], Vector AMP (VAMP) algorithm has been unfolded (without any additional smoothing), and all its linear filtering parameters and non-linear decision parameters has been left free for optimization through learning. As a result, the optimized parameters behave identically to the analytical MMSE-optimal solution of VAMP, thus showing that among iterative detection structures that consist of successive linear and non-linear estimators, EP-based algorithms such as VAMP/OAMP (without smoothing, i.e. $\beta = 0$) are already optimal with respect to their structures, and exhaustive learning of all parameters is not necessary.
The unfolding concept has been applied to SISO detector design in [9], by optimizing some hyper-parameters of the DL-SEP algorithm. As this algorithm belongs to the category of EP-based MMSE-optimal VAMP/OAMP-like techniques, relearning the filters and the decision function is not expected to bring any significant gain.

Algorithm 1 Double-Loop Scalar EP Equalizer (DL-SEP) [11]

Input $\underline{\mathbf{y}}$, $\underline{\mathbf{H}}$, σ_w^2

Initialize the decoder with $L_a^{(0)}(\mathbf{d}_k) = 0, \forall k$.

for $\tau = 0$ to $\mathcal{T}$ **do**

Update the prior PMF on x_k, with $\forall k$, $\forall \alpha \in \mathcal{X}$,
$\mathcal{P}_k^{(\tau)}(\alpha) \propto \exp\Big(-\sum_{q=1}^{Q} \varphi_q^{-1}(\alpha) L_a(d_{k,q})\Big)$.

Compute the prior mean and variance of x_k, $\forall k$
$x_k^{p(\tau)} \triangleq \mathbb{E}_{\mathcal{P}}[x_k] = \sum_{\alpha\in\mathcal{X}} \alpha \mathcal{P}_k^{(\tau)}(\alpha)$, $\quad v^{p(\tau)} \triangleq \frac{1}{K}\sum_k \mathrm{Var}_{\mathcal{P}}[x_k] = \frac{1}{K}\sum_k \Big(\sum_{\alpha\in\mathcal{X}} |\alpha|^2 \mathcal{P}_k^{(\tau)}(\alpha) - |x_k^{p(\tau)}|^2\Big)$.

Initialize the soft feedback on x_k with the priors, $\forall k$, $x_k^{d(\tau,0)} = x_k^{p(\tau)}$ and $v^{d(\tau,0)} = v^{p(\tau)}$.

for $s = 0$ to $\mathcal{S}$ **do**

Perform FFT on $\mathbf{x}^{\mathbf{d}(\tau,s)}$ to get $\underline{\mathbf{x}}^{\mathbf{d}(\tau,s)}$.

MMSE filtering with interference cancellation is performed for $k = 1, \dots, K$,
$\xi^{(\tau,s)} = K^{-1}\sum_k |\underline{h}_k|^2/(\sigma_w^2 + v^{d(\tau,s)}|\underline{h}_k|^2)$,
$\underline{f}_k^{(\tau,s)} = \underline{h}_k/[\xi^{(\tau,s)}(\sigma_w^2 + v^{d(\tau,s)}|\underline{h}_k|^2)]$,
$\underline{x}_k^{e(\tau,s)} = \underline{x}_k^{d(\tau,s)} + \underline{f}_k^{(\tau,s)*}(\underline{y}_k - \underline{h}_k \underline{x}_k^{d(\tau,s)})$,
$v^{e(\tau,s)} = 1/\xi^{(\tau,s)} - v^{d(\tau,s)}$.

Perform IFFT on $\underline{\mathbf{x}}^{d(\tau,s)}$ to get $\mathbf{x}^{d(\tau,s)}$.

Update a posteriori PMF of x_k, with $\forall k$, $\forall \alpha \in \mathcal{X}$,
$\mathcal{D}_k^{(\tau,s)}(\alpha) \propto \exp\Big(-|x_k^{e(\tau,s)} - \alpha|^2/v^{e(\tau,s)}\Big)\mathcal{P}_k^{(\tau)}(\alpha)$.

if $s < \mathcal{S}$ **then**

Compute the mean and the variance of APP PMF of x_k, $\forall k$
$\mu_k^{d(\tau,s)} \triangleq \mathbb{E}_{\mathcal{D}}[x_k] = \sum_{\alpha\in\mathcal{X}} \alpha \mathcal{D}_k^{(\tau,s)}(\alpha)$, $\quad \gamma^{d(\tau,s)} \triangleq K^{-1}\sum_k \mathrm{Var}_{\mathcal{D}}[x_k]$,.

The soft data feedback towards the equalizer is computed through the division of the PDF $\mathcal{CN}(\mu_{p,k}^d, \gamma_p^d)$, by the PDF $\mathcal{CN}(x_{p,k}^e, v_p^e)$
$v^{\star(\tau,s)} \triangleq 1/(1/\gamma^{d(\tau,s)} - 1/v^{e(\tau,s)})$,
$x_k^{\star(\tau,s)} \triangleq v^{\star(\tau,s)}\Big(\mu_k^{d(\tau,s)}/\gamma^{d(\tau,s)} - x_k^{e(\tau,s)}/v^{e(\tau,s)}\Big)$.

Avoid undesirable fixed points of EP (local extrema) through smoothing ($0 \le \beta \le 1$),
$x_k^{d(\tau,s+1)} \triangleq (1-\beta)x_k^{\star(\tau,s)} + \beta x_k^{d(\tau,s)}$,
$v^{d(\tau,s+1)} \triangleq (1-\beta)v^{\star(\tau,s)} + \beta v^{d(\tau,s)}$.

end if

end for

Compute extrinsic LLRs, $\forall k, q$
$L_e^{(\tau)}(d_{k,q}) \triangleq \ln\sum_{\alpha\in\mathcal{X}_q^0} \mathcal{D}_k^{(\tau,\mathcal{S})}(\alpha) - \ln\sum_{\alpha\in\mathcal{X}_q^1} \mathcal{D}_k^{(\tau,\mathcal{S})}(\alpha) - L_a^{(\tau)}(d_{k,q})$,
with $\mathcal{X}_q^b = \{\alpha \in \mathcal{X} : \varphi_q^{-1}(x) = b\}$, $b \in \mathbb{F}_2$, then feed them to the decoder to get next priors $L_a^{(\tau+1)}(\mathbf{d}_k), \forall k$.

end for

However, smoothing parameters can be learned in order to alleviate the message passing algorithm's weaknesses to short cycles on the factor graph. By considering a different damping value at each self-iteration, the learnable parameters are $\{\beta^{(0)}, \beta^{(1)}, \dots, \beta^{(\mathcal{S})}\}$, and to render them independent of the selected channel code, a mutual-information based loss function has been used in [9].

3.1 Mutual Information based Loss Functions for Learning SISO Detectors

In order to optimize the SISO detector with no dependence on the channel code, we propose to use a loss function that is directly inspired from the extrinsic information transfer (EXIT) analysis of turbo iterative systems [14]. The behaviour of the loss function is characterized by a single scalar that corresponds to the average mutual information (AMI) I_A between $L_a(\mathbf{d})$ and $\mathbf{d}$. This approach can be effectively applied by considering the prior LLRs to be consistently Gaussian-distributed, i.e. $L_a(\mathbf{d}) \sim \mathcal{N}((1-2d_{k,q})\mu_a, 2\mu_a)$, with $\mu_a = J^{-1}(I_A)$ and $J(\mu) \triangleq 1 - \mathbb{E}_{L\sim\mathcal{N}(\mu,2\mu)}[\log_2(1+e^{-L})]$. Next, in order to optimize the iterative convergence capabilities of the detector, the loss function should be directly dependent on the extrinsic LLRs $L_e(\mathbf{d})$, such that the optimized receiver also maximizes the area under the EXIT chart of the detector and thus enhance the achievable rates.

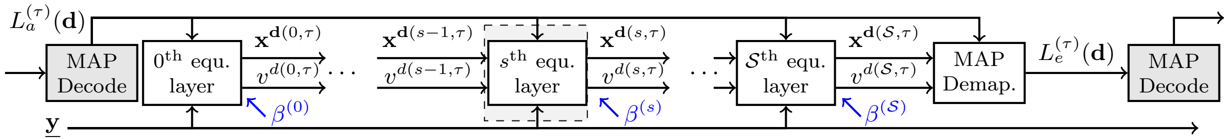

Figure 2 – "Learned-DL-SEP": Unfolded deep EP network (with $\mathcal{S}$ layers) at the τ^{th} turbo iteration.

Hence the proposed training loss function $\mathcal{L}$ is given as

$$\mathcal{L}(\mathbf{d}, \hat{\mathbf{d}}, I_A) \triangleq -\frac{1}{QNK}\sum_{k,q}\sum_{n}\ell(d_{k,q}, \hat{d}_{k,q}[n](I_A)), \tag{2}$$

$$\hat{d}_{k,q}[n](I_A) \triangleq \frac{1}{1+\exp(L_e(d_{k,q}, I_A)[n]}, \tag{3}$$

where I_A is the AMI between artificially generated $L_a(\mathbf{d})$ and $\mathbf{d}$, $L_e(\mathbf{d}, I_A)[n]$ is the extrinsic output of the unfolded network, when $L_a(d_{k,q}, I_A)[n]$ is provided as priors, and $n = 1, \ldots, N$ denotes the index of prior LLR samples. $\hat{d}_{k,q}[n](I_A)$ is a soft bit equivalent of the extrinsic messages, and $\ell(d_{k,q}, \hat{d}_{k,q}[n](I_A))$ denotes a bit-wise loss metric, which will be specified in Section 3.2.

The training process is carried out with the ADAM optimizer [15], which is a stochastic gradient descent technique where the learning rate is automatically adjusted depending on the first and second order statistics' estimates of the changes in the gradient of the loss function. By initializing the learning rate to 0.025, and by using mini-batches with 200 samples of

- noise variance realizations σ_w^2, with $\text{SNR}_c = 20\log_{10}\sigma_x/\sigma_w$ following uniform distribution over the interval of interest (e.g. 5 to 30 dB).
- a pseudo-codeword realization $\mathbf{d}$, uniformly selected among 2^{K_d} possibilites,
- realizations of the noise vector $\mathbf{w}$ and of the channel $\mathbf{H}$, according to σ_w^2 and channel power-delay profile,
- a set of prior LLR realizations $L_a(\mathbf{d}, I_a)[n]$, for $n = 1, \ldots, N$.

Learning process carries out an averaging over N prior codeword LLRs, for each sample of the mini-batch, in order to learn a value of the desired parameters $\boldsymbol{\beta} = [\beta^{(0)}, \beta^{(1)}, \ldots, \beta^{(\mathcal{S})}]$ as a function of I_A. Moreover, as there is a bijective between I_A and the prior variance of soft symbol estimates, $v^{p(\tau)}$ (see line 4 of Algorithm 1), $\boldsymbol{\beta}$ can be tabulated as a function of $v^{p(\tau)}$. Finally, during the practical deployment of the optimized receiver, instantaneous value of $v^{p(\tau)}$ at the ongoing turbo-iteration τ would enable the receiver to select its optimal weights $\boldsymbol{\beta}$.

This framework has enabled us to optimize DL-SEP by considering it as a deep neural network, and overcome somme error-floor issues at high SNR operating points in [9]. In the remainder of this paper we further investigate the capabilities and the extensions of this hybrid-AI based FDE structure.

3.2 The Impact of the Selected Loss Function

In our previous work, we have used the negative cross-entropy (CE) loss function, given by

$$\ell_{\text{CE}}(d_{k,q}, \hat{d}_{k,q}) = -\left[d_{k,q}\log\left(\hat{d}_{k,q}\right) + (1-d_{k,q})\log\left(1-\hat{d}_{k,q}\right)\right], \tag{4}$$

which was selected due to its strong link to the mutual information of extrinsic LLRs, which enables the learning framework to optimize the extrinsic information of the SISO module.

In the context of channel decoding with unfolded BP, the use of an alternative metric, called the soft-bit-error-rate (SBER), given by

$$\ell_{\text{SBER}}(d_{k,q}, \hat{d}_{k,q}) = \hat{d}_{k,q}{}^{1-d_{k,q}}\left(1-\hat{d}_{k,q}\right)^{d_{k,q}}, \tag{5}$$

yields better results for optimizing the bit error rate (BER) and the packet error rate (PER) of the system [16]. Hence, we are interest to test its pertinence in the equalization application under investigation.

Finally, it is also possible to use a symbol-wise metric to optimize this receivers performance, by using the mean-squared error (MSE) between the equalized estimate x_k^e and the transmitted symbols x_k in order to focus the training on optimizing the filtering and interference cancellation stage.

First, in order to illustrate the capacity of each one of these loss functions to capture the behaviour of BER, or PER, we consider a simplified neural equalizer with up to two layers, where the same fixed parameter β is

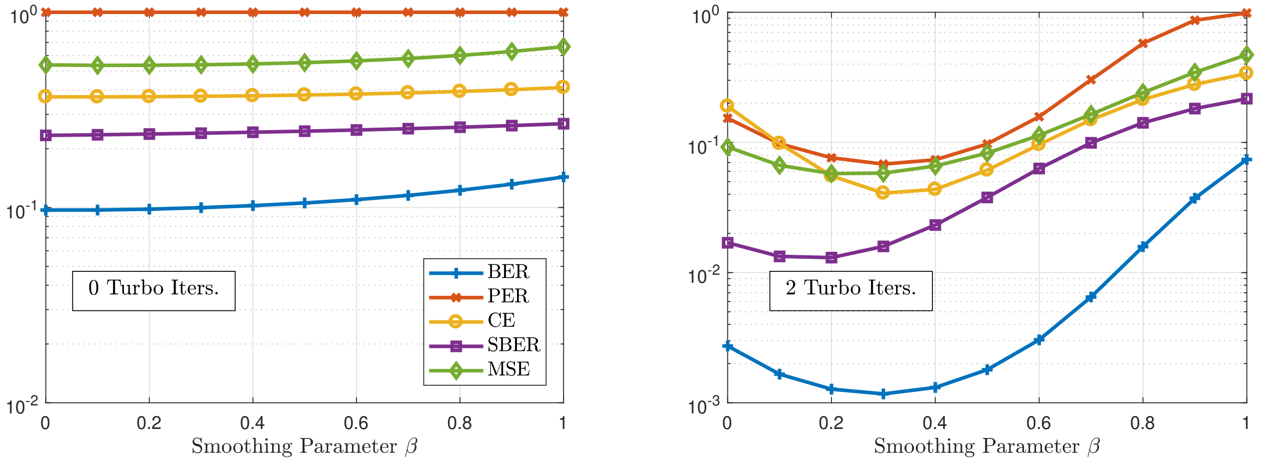

Figure 3 – Comparison of alternative loss metrics with BER and PER as a function of β for a 2-layer equalizer.

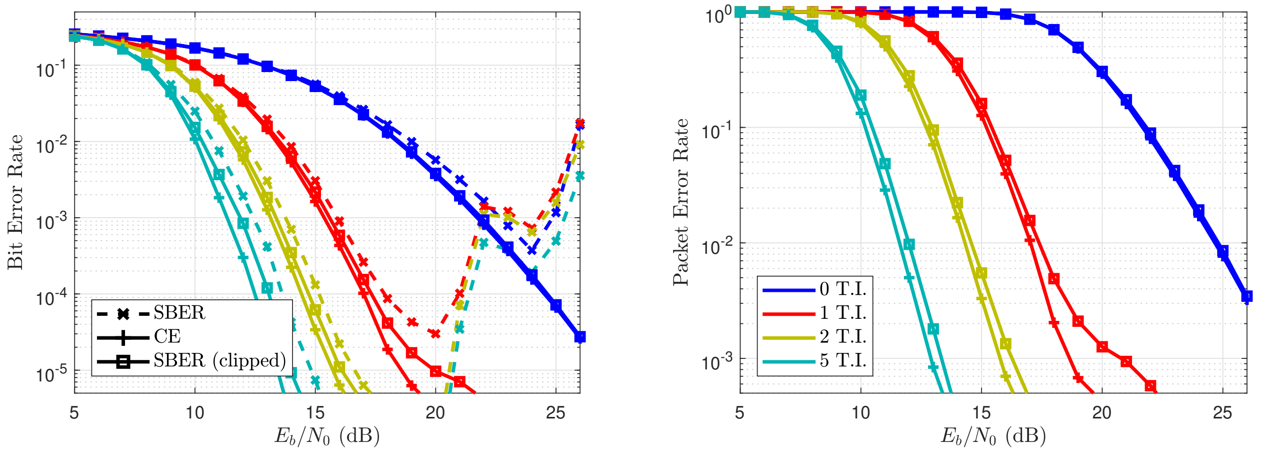

Figure 4 – Comparison of a 2-layer DL-SEP with trained parameters from different loss metrics.

used for smoothing. In this setting, for β varying from 0 to 1, BER, PER and the three loss functions above are measured.

In Figure 3, these loss functions are compared in the Proakis C channel for $E_b/N_0 = 13$ dB, when a 2-layer (2 self-iterations) DL-SEP receiver is used with a recursive systematic convolutional (RSC) channel code of polynomials $[7, 5]_8$ and rate $1/2$ with 8-PSK modulation. It is seen that the minimum of SBER manages to accurately follow the optimum BER when the system is not turbo-iterated (i.e. $I_A = 0$, left plot), but not necessarily when there is feedback from the decoder (right plot). On the other hand, the minimum of CE provides overall a closer estimate to the optimum PER, whether the system is iterated or not. MSE does not provide a sufficiently accurate representation of optimal error rates.

Hence, we have carried out the training framework for both CE and SBER loss functions, in order to evaluate their impact on the actual receiver performance. The training of SBER appeared to be numerically more challenging than CE; the ADAM optimizer tends to seek values of β outside the interval $[0, 1]$, and trained values for low I_A can be negative. This results in severe degradation of BER performance at high SNR, and to overcome this limitation, we have manually clipped learned parameters to $[0, 1]$. The impact of this approach is illustrated in Figure 4-(a), where the raw use of trained parameters from SBER causes unacceptable BER degradations above 20 dB, and interestingly the use of clipping slightly improves the receiver decoding threshold.

In Figure 4-(b), we compare the PER behaviour of a 2-layer DL-SEP when parameters are trained with CE and SBER (clipped), and CE is shown to outperform SBER overall, except in the absence of prior information, i.e. 0 T.I. (turbo iterations) where they nearly yield the same PER.

3.3 Interpretation of the Learned Parameters

Next, the values of the learned parameters (in our case the damping coefficients which control the receiver convergence behaviour) are discussed to illustrate that such hybrid learning techniques provide explainable results. In order to correctly interpret results, we recall that damping factors close to one means that the feedback from the previous self-iteration (equivalently, layer) is privileged. If damping factors of all self-iterations

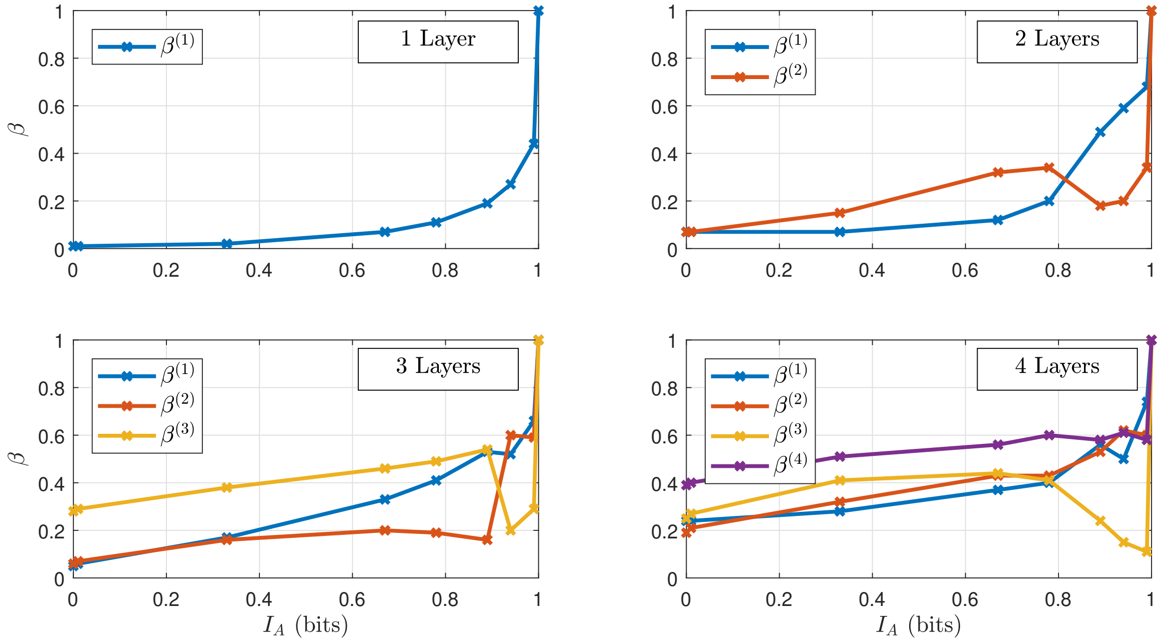

Figure 5 – Learned values of the damping parameters as a function of the mutual information I_A of prior LLRs provided by the decoder, for 1 to 4 inner "self-iteration" layers.

are high, the feedback from the decoder is weighted more. If damping factors are close to zero, the current feedback from the demapper takes more importance.

Figure 5 shows the values of the learned damping parameter as a function of the prior information $I_A \in \{0, 0.33, 0.67, 0.78, 0.89, 0.94, 0.99, 1\}$ for different number of layers, by using the CE. The plotted values of β correspond to those used for simulations with CE in Figure 4. When there is no or very weak prior information $I_A \approx 0$, confidence is given to the demapper feedback ($\beta \approx 0$), but in case of multiple layers, the damping factors of the deeper ones are higher (i.e. their feedback has weaker impact), thus controlling residual error propagation. When the prior information from the decoder is perfect, there is no need to equalize, hence $\beta \to 1$, otherwise a compromise is made across the layers. In particular, oppositely to the cases with low I_A, when I_A gets closer to 1 increased damping is used at the first layers in order thus privileging decoder feedback.

3.4 The Impact of SNR-adaptative Learned Parameters

In our previous works, as described in Subsection 3.1, the training is performed over a vast range of signal-to-noise ratio (SNR) interval to provide overall adequate learned parameters. Thus, in the context of DL-SEP, learned parameters are only dependent on the prior information I_A (or equivalently, the prior covariance $v^{p(\tau)}$). In this section, we explore the effects adapting parameters to the operating point conditions by making learned hyper-parameters also SNR-dependent. Hence the learning procedure is changed to use a *fixed* noise variance, in order to have a set of learned parameters $\beta^{(s)}(I_A, \sigma_w^2)$ tabulated as a function of I_A and σ_w^2.

In Figure 6 we compare the PER performance obtained with CE loss function in the conventional approach with parameters obtained through learning with SNR-dependent CE. The obtained improvements with the SNR-dependent parameters is less than 0.025 dB and thus we can conclude that the originally proposed framework which solely uses the prior information I_A for parameter adaptation is robust. Indeed, as the feedback from the decoder naturally incorporates some information about the state of the SNR, this learning framework appears to be attractive for optimizing SISO detectors.

4 Conclusion

The use of deep unfolding on iterative physical layer algorithms enables to overcome their modelling or complexity limitations by exploiting the powerful optimization capabilities of stochastic training strategies. In this paper, we aim to get a better grasp of unfolded turbo detection techniques that are optimized with hybrid-AI by further analysing the response of the proposed receiver to different learning parameters.

Our complementary investigations let us think that the promising soft BER loss functions, which provided reduced packet error rates in decoding applications, is not well adapted for turbo detection, while cross entropy on extrinsic soft bits seems to match better to the actual PER behaviour. Moreover, choosing the right loss function is fundamental also for reducing the complexity of the receiver. The cross entropy loss function matches

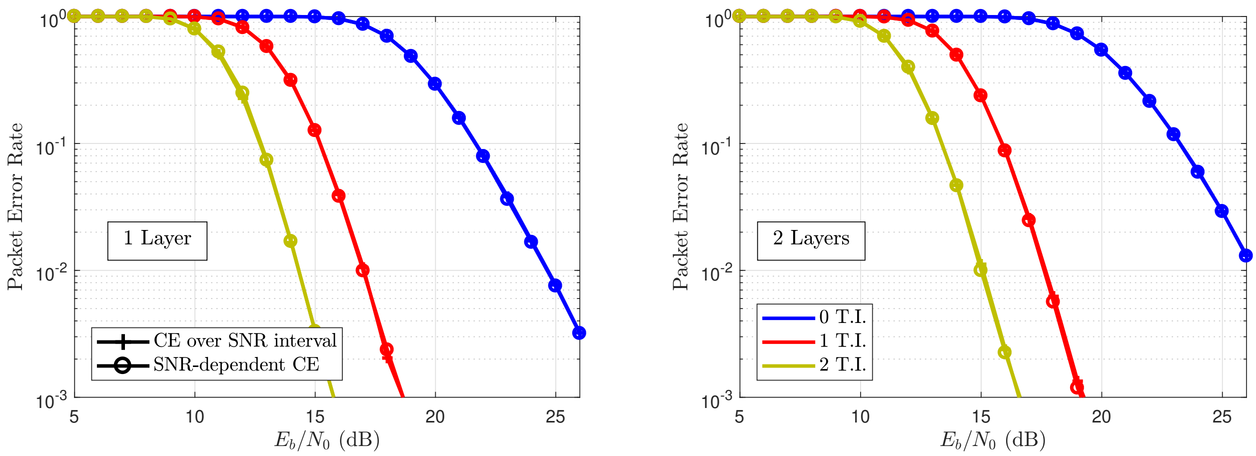

Figure 6 – Comparison of SNR-adaptative CE-based learned parameters with SNR-invariant CE-based learned parameters.

well with extrinsic information behaviour at the output of the demapper, and hence it allows to dynamically select optimal damping parameters only as a function of the average mutual information from the decoder, while adding SNR information is not necessary and brings no significant improvement.
Finally, by using deep unfolding, the neural network implemented at the receiver inherits the structure and fundamental behaviour of the underlying EP-based Bayesian inference algorithm, thus considerably simplifying the interpretation of the impact of the learned parameters on the receiver behaviour.

5 References

[1] O. Simeone, "A very brief introduction to machine learning with applications to communications systems," *IEEE Transactions on Cognitive Communications*, vol. 4, pp. 648 – 664, Dec. 2018.

[2] A. Balatsoukas-Stimming and C. Studer, "Deep unfolding for communications systems: a survey and some new directions," *a*, vol. abs/1906.05774, June 2019.

[3] W. Guo, "Explainable arficial intelligence (xai) for 6g: Improving trust between human and machine," *a*, vol. abs/1911.04542, Sept. 2019.

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

[5] C. Douillard, M. Jézéquel, C. Berrou, A. Picart, *et al.*, "Iterative correction of intersymbol interference: Turbo-equalization," *Eur. Trans. Telecomm.*, vol. 6, pp. 507–511, Sept. 1995.

[6] M. Tüchler and A. C. Singer, "Turbo equalization: An overview," *IEEE Trans. on Inf. Theory*, vol. 57, pp. 920–952, Feb. 2011.

[7] H. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proceedings of the IEEE*, vol. 95, pp. 1295–1322, June 2007.

[8] M. Senst *et al.*, "How the framework of expectation propagation yields an iterative IC-LMMSE MIMO receiver," in *IEEE 54th GLOBECOM*, Dec. 2011.

[9] S. Şahin, C. Poulliat, A. M. Cipriano, and M.-L. Boucheret, "Doubly iterative turbo equalization: Optimization through deep unfolding," in *IEEE 30th PIMRC*, pp. 1–6, Sept. 2019.

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

[11] S. Şahin, A. M. Cipriano, C. Poulliat, and M.-L. Boucheret, "A framework for iterative frequency domain EP-based receiver design," *IEEE Transactions on Communications*, vol. 66, pp. 6478–6493, Dec. 2018.

[12] E. Nachmani, Y. Be'ery, *et al.*, "Learning to decode linear codes using deep learning," in *54th Annual Allerton Conf.*, pp. 341–346, Sept. 2016.

[13] M. Borgerding *et al.*, "AMP-inspired deep networks for sparse linear inverse problems," *IEEE Transactions on Signal Processing*, vol. 65, pp. 4293–4308, Aug. 2017.

[14] S. Ten Brink, "Designing iterative decoding schemes with the extrinsic information transfer chart," in *AEU Int. JEC*, vol. 54, Jan. 2000.

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

[16] M. Lian, F. Carpi, C. Häger, and H. D. Pfister, "Learned belieft-propagation decoding with simple scaling and snr adaptation," *a*, vol. abs/1901.08621, Jan. 2019.