

Récepteurs Itératifs par Propagation de l’Espérance : Optimisation par Dépliage Profond

Serdar ŞAHİN^{1,2}, Antonio Maria CIPRIANO¹, Charly POUILLIAT², Marie-Laure BOUCHERET²

¹THALES, 4 avenue des Louvresses, 92230, Gennevilliers, France

²IRIT - Toulouse INP, CNRS, 2 Rue Charles Camichel, 31000, Toulouse, France

{serdar.sahin, antonio.cipriano}@thalesgroup.com
{charly.poulliat, marie-laure.boucheret}@enseeiht.fr

Résumé – Ce papier étudie la conception de récepteurs itératifs à faible complexité à travers l’utilisation du principe de la propagation de l’espérance, un algorithme d’inférence bayésienne approximée. Une stratégie de dépliage profond est utilisée pour interpréter ce récepteur comme un réseau multi-couches permettant ainsi d’améliorer les performances par une optimisation des paramètres de cet algorithme au travers d’une phase d’apprentissage. Pour cette dernière, une nouvelle fonction de coût est proposée pour l’apprentissage avec des turbo-récepteurs.

Abstract – This paper investigates the design of low complexity iterative receivers through the use of expectation propagation, an approximate Bayesian inference algorithm. Moreover, a deep unfolding strategy is used to interpret this receiver as a multi-layer network and to improve its performance by optimizing its parameters through learning. To this end, a novel cost function is proposed for learning suited to turbo-receivers.

1 Introduction

La gestion et la mitigation de l’interférence est un problème important pour les systèmes de communications. D’un côté, dû à la pénurie de ressources fréquentielles et dû aux contraintes temporelles strictes, ces systèmes doivent tolérer de l’interférence intra-système. D’autre part, lorsqu’il s’agit de transmissions large bande, les environnements de propagation difficiles introduisent de l’interférence entre-symboles (ISI).

La détection et le décodage optimal au sens du taux d’erreur paquet correspond au critère du maximum a posteriori (MAP) conjoint pour la détection et le décodage, qui a cependant une complexité calculatoire trop élevée. La technique de modulation codée à bits entrelacés (BICM) permet de mettre en place des turbo-récepteurs [1], moins complexes, qui peuvent se rapprocher des performances optimales asymptotiquement.

Il existe un grand nombre de turbo-récepteurs dans la littérature, et la plupart sont dérivés avec des méthodes d’inférence bayésienne approximées tel que la propagation de croyance (BP) ou BP avec l’approximation gaussienne (GABP) [2, 3]. Plus récemment, des méthodes d’inférence plus avancées telles que la propagation de l’espérance (EP) [4] ont été considérées pour la conception de récepteurs en couche physique [5, 6, 7]. EP permet entre-autre de concevoir un détecteur à entrées et sorties souples (SISO) itératif constitué d’une étape de filtrage optimisée selon le critère d’erreur quadratique moyenne minimum (MMSE) et d’une étape de démodulation MAP par symbole. Il existe aussi une famille d’algorithmes de passage de messages approximés (AMP), dérivés à partir de EP, avec une utilisation répandue pour la détection parcimonieuse [8, 9].

Par ailleurs, l’apport possible de l’apprentissage profond aux problématiques de détection liées à la couche physique est un sujet d’actualité. Dans ce contexte, le dépliage "profond" permet d’obtenir des structures de réseaux multi-couches à partir de certains algorithmes itératifs [10]. Par exemple, une version dépliée de BP a été utilisée pour améliorer la performance du décodage canal pour des matrices de parité denses [11], en rajoutant du lissage sur les messages de BP.

En vue de ces développements, cet article étudie la conception de turbo-récepteur avec l’EP dans un contexte de transmissions à BICM avec une forme d’onde mono-porteuse circulaire. Par la suite, nous analysons les liens qui existent entre les récepteur EP et des récepteurs issue de l’apprentissage profond et nous proposons une fonction de coût dédié aux turbo-récepteurs, en exploitant l’information a priori. Pour finir, nous appliquons cette fonction de coût pour optimiser les paramètres du récepteur à base de l’EP à travers l’apprentissage profond.

2 Modèle du système

Dans cet article, nous considérons des schémas de type BICM pour des transmissions mono-porteuses par bloc et circulaires sur un canal sélectif en fréquence (i.e. des schémas permettant une égalisation fréquentielle en réception). K_b bits d’information \mathbf{b} sont codés par un code de canal, puis entrelacés¹ pour donner le mot de code binaire \mathbf{d} de taille K_d . Le modulateur

1. Les techniques élaborées dans cet article peuvent s’étendre à des transmissions à entrées et/ou sorties multiples (MIMO), voire à des transmissions non-orthogonales multi-utilisateurs, avec respectivement l’usage d’une modulation codée à bits entrelacés espace-temps (ST-BICM) et d’un récepteur avec suppression d’interférence entre-utilisateurs [12].

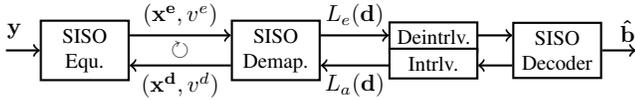


FIGURE 1 – La structure du récepteur à deux-boucles avec EP.

d'amplitude en quadrature sans-mémoire φ associe le mot de Q -bits $\mathbf{d}_k \triangleq [d_{Qk}, \dots, d_{Q(k+1)-1}]$ au symbole $x_k \in \mathcal{X}$, issu de la constellation \mathcal{X} , $|\mathcal{X}| = M$ et $Q = \log_2 M$. On obtient alors le bloc de données $\mathbf{x} \in \mathcal{X}^K$ tel que $K_d = QK$. On utilisera aussi $\varphi_q^{-1}(x_k)$ ou $d_{k,q}$ pour se référer à d_{kQ+q} .

Sous l'hypothèse d'une synchronisation parfaite au récepteur, les observations du signal bande de base sont données par $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, où \mathbf{w} est un bruit blanc gaussien de variance σ_w^2 . Dans le domaine fréquentiel, les observations s'écrivent

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

où $\underline{\mathbf{x}} = \mathcal{F}_K \mathbf{x}$, $\underline{\mathbf{w}} = \mathcal{F}_K \mathbf{w}$ et $\underline{\mathbf{H}} = \mathcal{F}_K \mathbf{H} \mathcal{F}_K^H$. La matrice du canal \mathbf{H} est donc une matrice circulante, ayant pour première ligne la réponse impulsionnelle $\mathbf{h} = [h_0, \dots, h_{L-1}, \mathbf{0}_{K-L-1}^T]$, où $L < K$ est l'étalement du canal². \mathcal{F}_K est la matrice de transformée de Fourier discrète (DFT) unitaire de taille K , et $\mathcal{F}_K \mathcal{F}_K^H = \mathbf{I}_K$. Avec les propriétés de la DFT, on a $\underline{\mathbf{H}} = \text{Diag}(\underline{\mathbf{h}})$ où $\underline{h}_k = \sum_{l=0}^{L-1} h_l \exp(-2j\pi kl/K)$.

Nous considérerons un récepteur itératif où le décodeur et le détecteur échangent de l'information extrinsèque avec des log-rapports de vraisemblances (LLRs) associés aux bits codés \mathbf{d} . On notera, du point de vue du détecteur, les LLRs a priori $L_a(d_p)$ et extrinsèques $L_e(d_p)$ et la distribution de probabilité discrète (PMF) a priori de x_k est $\mathcal{P}_k(\alpha)$ pour $\alpha \in \mathcal{X}$. On note enfin $\tau = 0, \dots, \mathcal{T}$ l'indice courant d'une turbo-itération.

3 Égaliseur EP fréquentiel pour schémas BICM

Propagation de l'espérance : La densité de probabilité a posteriori (APP) des bits d'information, sachant les observations, et une connaissance parfaite du canal est $p(\mathbf{b}, \mathbf{d}, \mathbf{x} | \mathbf{y})$, et le critère MAP conjoint cherche le maximum de sa marginale sur \mathbf{b} . L'approche BICM simplifie l'estimation de cette densité de probabilité (PDF) en permettant de factoriser l'expression précédente, et d'y appliquer un algorithme d'inférence approximée, sous réserve d'un entrelaceur suffisamment long et d'un mapping sans mémoire. On considère ici une approche par passage de messages basée sur EP, qui étend l'approche BP en affectant une famille de PDF à chaque noeud de variable (VN). Par la suite, les messages arrivant vers ce VN sont contraints à appartenir à cette famille [4], en utilisant une projection de type Kullback-Leibler. Cette dernière est équivalente à une *méthode des moments (moment matching)* pour la famille exponentielle.

2. Ce modèle s'applique à diverses formes d'onde mono-porteuses utilisant un préfixe cyclique (CP), une séquence de zéros (ZP) ou un mot unique (UW), avec un filtrage de mise en forme temporelle ou spectrale. Cela comprend aussi l'usage d'une mono-porteuse à accès multiple par répartition en fréquence (SC-FDMA) ou une signalisation plus rapide que Nyquist (FTN).

Algorithme 1 Récepteur SEP à deux boucles (DL-SEP)

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

- 1: Initialise le décodeur $L_a^{(0)}(\mathbf{d}_k) = 0, \forall k$.
- 2: **pour** $\tau = 0$ to \mathcal{T} **faire**
- 3: Initialise l'égaliseur $x_k^{e(\tau,0)} = 0, \forall k$ and $v^{e(\tau,0)} = +\infty$.
- 4: **pour** $s = 0$ to \mathcal{S} **faire**
- 5: Màj la distribution APP de x_k , avec $\forall k, \forall \alpha \in \mathcal{X}$, $\mathcal{D}_k^{(\tau,s)}(\alpha) \propto \exp(-|x_k^{e(\tau,s)} - \alpha|^2 / v^{e(\tau,s)}) \mathcal{P}_k^{(\tau)}(\alpha)$.
- 6: Calculer la moyenne et la variance APP de $x_k, \forall k$
 $\mu_k^{d(\tau,s)} \triangleq \mathbb{E}_{\mathcal{D}}[x_k] = \sum_{\alpha \in \mathcal{X}} \alpha \mathcal{D}_k^{(\tau,s)}(\alpha)$,
 $\gamma^{d(\tau,s)} \triangleq K^{-1} \sum_k \text{Var}_{\mathcal{D}}[x_k]$.
- 7: Le retour vers l'égaliseur est obtenu avec la division de la PDF $\mathcal{CN}(\mu_{p,k}^d, \gamma_p^d)$, par la PDF $\mathcal{CN}(x_{p,k}^e, v_p^e)$
 $v^{*(\tau,s+1)} \triangleq 1 / (1/\gamma^{d(\tau,s)} - 1/v^{e(\tau,s)})$,
 $x_k^{*(\tau,s+1)} \triangleq v^{*(\tau,s+1)} \left(\frac{\mu_k^{d(\tau,s)}}{\gamma^{d(\tau,s)}} - \frac{x_k^{e(\tau,s)}}{v^{e(\tau,s)}} \right)$.
- 8: On évite les éventuels mauvaises point-fixes de EP (extrema locaux), avec du lissage ($0 \leq \beta \leq 1$) [4],
 $x_k^{d(\tau,s)} = (1 - \beta)x_k^{*(\tau,s)} + \beta x_k^{d(\tau,s-1)}$,
 $v^{d(\tau,s)} = (1 - \beta)v^{*(\tau,s)} + \beta v^{d(\tau,s-1)}$.
- 9: Le filtrage MMSE s'effectue avec, $k = 0, \dots, K - 1$,
 $\xi^{(\tau,s)} = K^{-1} \sum_k |h_k|^2 / (\sigma_w^2 + v^{d(\tau,s)} |h_k|^2)$,
 $f_k^{(\tau,s)} = h_k / [\xi^{(\tau,s)} (\sigma_w^2 + v^{d(\tau,s)} |h_k|^2)]$,
 $\underline{x}_k^{e(\tau,s)} = \underline{x}_k^{d(\tau,s)} + f_k^{(\tau,s)*} (y_k - h_k \underline{x}_k^{d(\tau,s)})$,
 $v^{e(\tau,s)} = 1/\xi^{(\tau,s)} - v^{d(\tau,s)}$.
- 10: **fin pour**
- 11: Màj PMF APP $\mathcal{D}_k^{(\tau,S+1)}(\alpha)$ et LLRs extrinsèques sont
 $L_e^{(\tau)}(d_{k,q}) = \ln \frac{\sum_{\alpha \in \mathcal{X}_q^0} \mathcal{D}_k^{(\tau,S+1)}(\alpha)}{\sum_{\alpha \in \mathcal{X}_q^1} \mathcal{D}_k^{(\tau,S+1)}(\alpha)} - L_a^{(\tau)}(d_{k,q})$,
avec $\mathcal{X}_q^b = \{\alpha \in \mathcal{X} : \varphi_q^{-1}(x) = b\}$, $b \in \mathbb{F}_2$, puis les fournir au décodeur pour obtenir les LLRs a priori $L_a^{(\tau+1)}(\mathbf{d}_k), \forall k$.
- 12: **fin pour**

EP avec des signaux temporellement blancs : EP offre une grande liberté pour la résolution d'un problème d'inférence. Dans l'objectif d'avoir le meilleur compromis performance - complexité, il faut choisir avec attention la famille des PDF des VNs et l'ordonnancement des messages sur le graphe associé au détecteur. Travailler avec des PDFs gaussiennes amène à des solutions mettant en oeuvre des étapes de filtrage de type MMSE et de l'annulation d'interférence. D'autre part, la structure de la matrice de covariance choisie pour la modélisation impacte la complexité du détecteur. Par exemple, on peut considérer une structure de matrice diagonale (EP-diagonale), qui néglige toutes les corrélations, ou une structure de matrice proportionnelle à l'identité (EP-scalaire) [13]. EP-diagonale a été utilisé pour la turbo-égalisation dans le domaine temporel [14], où le récepteur a une structure de banc de filtres MMSE auto-itérée, avec une complexité calculatoire quadratique en K .

D'autre part, EP-scalaire (SEP) a été utilisé pour les turbo-récepteurs, en considérant les signaux temporellement blancs,

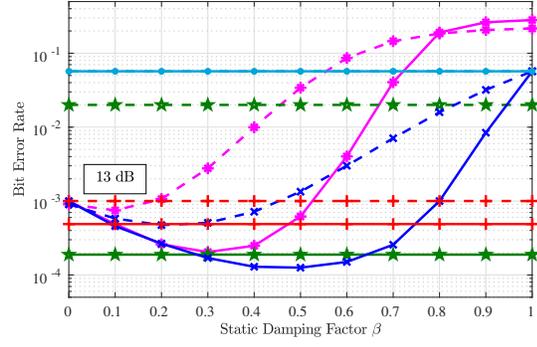
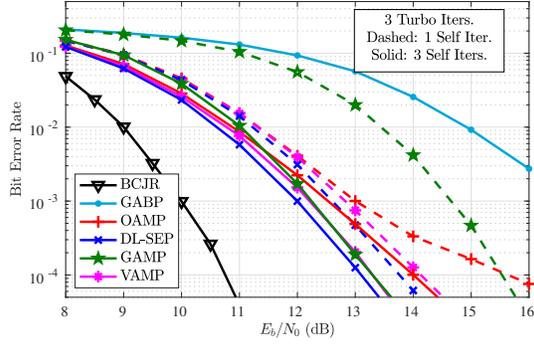


FIGURE 2 – Comparaison de BER dans le canal Proakis C avec du 8-PSK codé par RSC $[1, 5/7]_8$ coded 8-PSK.

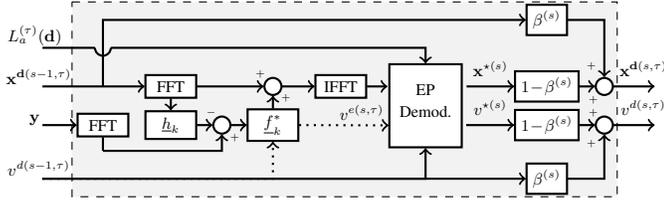


FIGURE 3 – “Learned-DL-SEP” : $s^{\text{ème}}$ couche d’égalisation EP.

pour divers problèmes de détection dans le domaine fréquentiel [7], avec des détecteurs à complexité quasi-linéaire, en exploitant la transformée de Fourier rapide (FFT). Les résultats expérimentaux indiquaient que SEP atteint asymptotiquement les mêmes performances de détection que EP-diagonale et cette observation est en ligne avec une analyse formelle qui montre que, sous conditions d’indépendance asymptotique, ces deux approches donnaient lieu à des points fixes identiques [15]. Cet récepteur SEP à deux boucles de [7] (DL-SEP), est illustré sur la figure 1, et il est décrit dans l’algorithme 1.

Liens avec AMP : L’algorithme AMP généralisé (GAMP) approxime les estimations extrinsèques de EP par des a posteriori, et essaye de compenser les corrélations introduites avec un terme additif dit de “réaction”. L’AMP orthogonal (OAMP) [8] et l’AMP vectorielle (VAMP) [9] sont des algorithmes qui possèdent une forte ressemblance à SEP, sur le problème de détection considéré, mais les heuristiques d’implémentation (lissage, calcul de variance) créent des différences de comportement. Ces algorithmes, connus habituellement pour leur faible complexité, sont comparés au DL-SEP dans [7], et le taux d’erreur binaire (BER) de ces récepteurs sont comparés sur la partie gauche de la fig. 2, pour des transmissions 8-PSK, avec une taille de bloc $K = 256$ avec un code convolutif, avec les hyperparamètres (tel que le lissage) étant optimisés. La partie droite de la figure 2 illustre l’impact des hyperparamètres sur ces algorithmes ; on observe que le récepteur à EP-scalaire peut atteindre des BER plus faibles que les alternatives AMP.

4 EP avec apprentissage profond

Déplieement profond pour la turbo-détection : Le déplieement profond permet de générer une structure multi-couches à partir d’un algorithme itératif, parfois en y ajoutant des hyper-

paramètres supplémentaires à apprendre. Cela a été utilisé avec BP et du lissage pour améliorer les performances de décodage canal [11]. D’autre part, dans [16], l’algorithme VAMP, sans lissage, est déplié, en rendant tous les paramètres des filtres et des éléments non-linéaires “libres” pour apprentissage. Après l’optimisation de ces paramètres par apprentissage, on constate que les paramètres appris possèdent les valeurs des paramètres indiqués par VAMP. Ainsi, les algorithmes SEP/VAMP/OAMP purs (sans lissage, i.e. $\beta = 0$), sont déjà optimaux par rapport à leurs structures, et ne nécessitent pas d’apprentissage.

Cependant, comme la fig. 2 l’indique, le lissage améliore considérablement les performances, et le BER suit une évolution régulière avec un minimum local pour β . Ainsi nous proposons de déplier DL-SEP, en considérant chaque auto-itération comme une couche d’égalisation, fig. 3, suivie d’une couche de demapping pour constituer un réseau EP “profond”, illustré sur la fig 4, avec les paramètres $\theta = [\beta^{(1)}, \dots, \beta^{(S)}]$ à apprendre.

Apprentissage pour les turbo-récepteurs : Pour l’apprentissage sur le réseau profond associé à EP, nous proposons une nouvelle fonction de coût \mathcal{L} , capable de suivre les dynamiques de la turbo-détection. L’entropie-croisée des bits-souples extrinsèques du DL-SEP avec les bits transmis, est au coeur de \mathcal{L} qui dépend aussi des LLRs a priori provenant du décodeur

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

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

où I_A est l’information mutuelle moyenne entre $L_a^{(\tau)}(\mathbf{d})$ et \mathbf{d} , $L_e(\mathbf{d}, I_A)[n]$ est la sortie extrinsèque du réseau, obtenue lorsqu’on fournit les a priori $L_a(d_{k,q}, I_A)[n]$ avec $n = 1, \dots, N$ étant l’indice de l’échantillon. Nous générons les a priori suivant $\mathcal{N}((1 - 2d_{k,q})\mu_a, 2\mu_a)$, avec $\mu_a = J^{-1}(I_A)$ et $J(\mu) \triangleq 1 - \mathbb{E}_{L \sim \mathcal{N}(\mu, 2\mu)}[\log_2(1 + e^{-L})]$. Ce choix est directement inspiré de l’approche de synthèse des fonctions EXIT [17].

\mathcal{L} permet alors d’apprendre les valeurs optimales de θ , en fonction de I_A , et il existe un lien bijectif entre I_A , et la covariance a priori $v^{a(\tau)} \triangleq K^{-1} \sum_k \text{Var}_{\mathcal{P}}[x_k]$. Donc, les valeurs apprises sont tabulées en fonction de $v^a \in [0, \sigma_x^2]$, et le récepteur calibre la valeur de $\theta^{(\tau)}$, en interpolant cette table selon la valeur mesurée de $v^{a(\tau)}$ dans la turbo-itération τ courante.

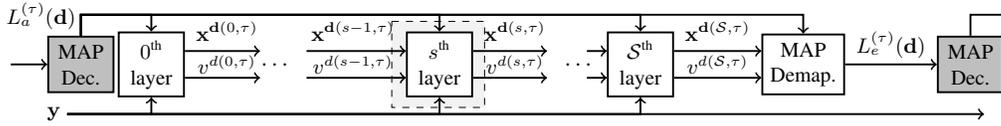


FIGURE 4 – “Learned-DL-SEP” : Réseau EP profond déplié, à la $\tau^{\text{ème}}$ turbo itération.

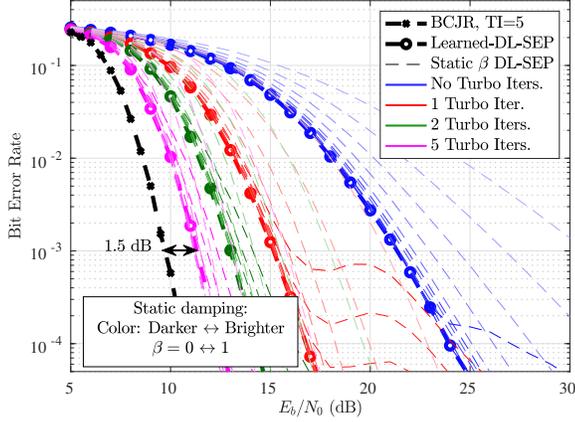


FIGURE 5 – Comparaison dans Proakis C avec 8-PSK codé par RSC $[1, 5/7]_8$ du DL-SEP et “Learned-DL-SEP” avec $\mathcal{S} = 3$.

On procède à l’apprentissage avec l’optimiseur ADAM [18], avec un taux d’apprentissage initiale de 0.025 et des mini-lots de 200 échantillons de :

- σ_w^2 , avec $\text{SNR}_c = 20 \log_{10} \sigma_x / \sigma_w$ de loi uniforme,
- des mots de code aléatoires \mathbf{d} , parmi les 2^{K_d} possibilités,
- réalisations du vecteur de bruit \mathbf{w} et du canal \mathbf{H} ,
- réalisations de $L_a(\mathbf{d}, I_a)[n]$, pour $n = 1, \dots, N$.

Cette approche permet d’optimiser DL-SEP (ou autre turbo-récepteur déplié), en le considérant comme un réseau profond.

5 Résultats numériques

On propose d’évaluer le DL-SEP déplié, afin de vérifier que l’apprentissage profond permet d’atteindre les performances optimales. Pour illustration, on considère un canal Proakis-C, $\mathbf{h} = [0.23, 0.46, 0.69, 0.46, 0.23]$, avec $\text{SNR}_c \in [5, 20]$ dB et nous avons trouvé que le nombre d’échantillons a priori $N = 25$ était suffisant, avec 150 itérations d’apprentissage, et une précision de 0.05 sur β . On apprend les poids β pour $I_A \in \{0, 0.33, 0.67, 0.78, 0.89, 0.94, 0.99, 1\}$.

Les performances du “Learned-DL-SEP” affichées sur la fig. 5, et ceux de DL-SEP avec lissage statique (à travers tous les itérations), avec β qui varie de 0 à 1 avec des pas de 0.1. DL-SEP avec un β faible a un bon seuil de détection, mais il est sensible à la propagation d’erreurs à fort SNR, et un β élevé ralentit la convergence et augmente le seuil de détection. “Learned-DL-SEP” parvient à s’adapter dynamiquement à la situation, avec l’apprentissage le permettant d’évaluer les valeurs optimales de $\beta^{(s,\tau)}$ en fonction de $v^{a(\tau)}$. On constate que “Learned-DL-SEP” opère sur l’enveloppe convexe de l’ensemble des performances de DL-SEP, en adaptant la valeur du β selon la qualité

du retour du décodeur. Pour finir, avec 3 couches, ce récepteur est à 1.5 dB du BCJR pour $\text{BER} = 10^{-3}$ et $\mathcal{T} = 5$ pour une complexité nettement moindre.

6 Conclusion

Ce papier étudie l’optimisation des turbo-récepteurs à faible-complexité à base de EP à travers des méthodes de l’apprentissage profond. Avec une nouvelle fonction de coût dédiée à la prise en compte de l’information a priori au cours de la turbo-détection, nous parvenons à optimiser les paramètres de ces algorithmes pour un scénario cible, avec un coût d’apprentissage relativement faible.

Références

- [1] C. Douillard *et al.*, “Iterative correction of intersymbol interference : Turbo-equalization,” *ETT.*, vol. 6, no. 5, pp. 507–511, Sep. 1995.
- [2] 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.
- [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. Minka *et al.*, “Divergence measures and message passing,” Microsoft Research, Tech. Rep., 2005.
- [5] M. Senst *et al.*, “How the framework of expectation propagation yields an iterative IC-LMMSE MIMO receiver,” in *GLOBECOM’11*, Dec. 2011.
- [6] X. Meng, Y. Wu *et al.*, “Low complexity receiver for uplink SCMA system via expectation propagation,” in *IEEE WCNC*, Mar. 2017, pp. 1–5.
- [7] S. Şahin *et al.*, “A framework for iterative frequency domain EP-based receiver design,” *IEEE TCOM*, vol. 66, no. 12, pp. 6478–6493, Dec. 2018.
- [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,” in *IEEE Proc. in ISIT*, Jun. 2017, pp. 1588–1592.
- [10] J. L. J. R. Hershey and F. Wenginger, “Deep unfolding : Model-based inspiration of novel deep architectures,” Mitsubishi Electric Research Labs, Tech. Rep. TR2014-117, 2014, version 12.0.1.
- [11] E. Nachmani, Y. Be’ery *et al.*, “Learning to decode linear codes using deep learning,” in *54th Annual Allerton Conf.*, Sep. 2016, pp. 341–346.
- [12] S. Şahin, C. Poulliat *et al.*, “Spectrally efficient iterative MU-MIMO receiver for SC-FDMA based on EP,” in *IEEE PIMRC’18*, Sep. 2018.
- [13] M. Opper and O. Winther, “Expectation consistent approximate inference,” *J. of Mach. Learning Res.*, vol. 6, no. Dec., pp. 2177–2204, 2005.
- [14] I. Santos *et al.*, “Turbo EP-based equalization : A filter-type implementation,” *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 4259–4270, Sep. 2018.
- [15] B. Çakmak *et al.*, “Expectation propagation for approximate inference : Free probability framework,” in *IEEE ISIT*, Jun. 2018, pp. 1276–1280.
- [16] M. Borgerding *et al.*, “AMP-inspired deep networks for sparse linear inverse problems,” *IEEE TSP*, vol. 65, no. 16, pp. 4293–4308, Aug. 2017.
- [17] S. Ten Brink, “Designing iterative decoding schemes with the extrinsic information transfer chart,” in *AEU Int. JEC*, vol. 54, no. 6, Jan. 2000.
- [18] D. P. Kingma and J. Ba, “Adam : a method for stochastic optimization,” in *Proc. Int. Conf. Learn. Representations*, May 2015.