

Le Smart Kalman filter : une approche basée sur l'apprentissage profond pour l'estimation du canal

Antoine SIEBERT^{1,2} Guillaume FERRÉ¹ Bertrand LE GAL¹ Aurélien FOURNY²

¹Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400, Talence, France

²THALES, France

Résumé – Dans les communications sans fil, le signal reçu peut être fortement altéré par l'environnement et ainsi être affecté d'interférence entre symboles (IES). Afin de correctement estimer les symboles reçus, il faut minimiser cette IES. Pour cela, le canal de propagation doit être estimé dans le but d'être compensé. Le filtre de Kalman (KF pour Kalman filter en anglais) est un algorithme d'estimation incontournable pour les systèmes linéaires de par son optimalité en terme d'erreur quadratique moyenne minimale (EQMM). Cependant, les implémentations d'un KF dans des systèmes réels posent souvent des difficultés car certains paramètres nécessitent d'être manuellement fixés. Dans cet article, nous présentons le Smart Kalman filter (SKF), une architecture hybride qui allie un KF avec les capacités de l'apprentissage profond pour estimer les paramètres optimaux du KF au cours du temps. Nous démontrons dans ce papier l'efficacité du SKF qui offre un gain entre 3 dB et 5 dB par rapport à un algorithme des moindres carrés (LS pour Least Squares en anglais) dans un contexte de canaux variant brutalement dans le temps.

Abstract – In digital wireless communications, the received signal can be strongly altered by the environment and may contain inter-symbol interference (ISI). To remove or reduce the ISI, the impulse response of the propagation channel can be estimated. The Kalman filter (KF) is an inescapable estimation algorithm in linear systems because of its optimality in term of minimum mean square error (MMSE) under certain assumptions. However, in real conditions, implementations of KF are often difficult because of the necessity of hand-tuning parameters. In this paper, we present the Smart Kalman filter (SKF), a hybrid architecture, that combines a KF and the power of neural networks to extract relevant features from data to benefit from an adaptive KF that is automatically well-tuned over time. We demonstrate in this paper that our SKF is 3 dB to 5 dB better than Least Square (LS) algorithm in a time-varying channel estimation context with abrupt Doppler frequency variations.

1 Introduction

Lors d'une communication numérique sans fil, l'environnement joue un rôle majeur. En effet, en raison d'effets tels que la réflexion, la diffraction ou la diffusion, le récepteur reçoit à chaque instant une observation composée de différents signaux provenant de la même source. Ce phénomène porte le nom de multitrajét. Le multitrajét étant la cause de l'IES, il doit être minimisé ou supprimé dans le but de pouvoir estimer correctement chaque symbole. Ce processus s'appelle l'égalisation. Afin d'égaliser le signal reçu, le canal de propagation doit être estimé. Ce canal de propagation est entièrement caractérisé par sa réponse impulsionnelle h (CIR pour channel impulse response en anglais). Afin d'estimer h , des séquences de symboles pilotes sont généralement insérées dans les trames à transmettre. Différents algorithmes peuvent être implémentés pour estimer la CIR, comme par exemple l'algorithme LS [9], l'algorithme des moindres carrés moyens (LMS pour Least Mean Squares en anglais) ou l'algorithme des moindres carrés récursifs (RLS pour Recursive Least Squares en anglais). Bien que l'estimation de la CIR par un KF ait déjà été réalisée [1], cet algorithme n'est pas très utilisé dans les systèmes réels car certains paramètres nécessitent d'être réglés manuellement. Pour éviter cette contrainte, l'Interacting Multiple Model (IMM) a été développé [2]. L'idée est de combiner plusieurs KF afin de couvrir d'avantage de situations, en échange d'un coût calculatoire plus important. Cependant, la difficulté du paramétrage persiste.

Depuis quelques années, différents travaux ont mis en avant

l'apprentissage profond pour la couche physique, et en particulier les réseaux de neurones récurrents ont prouvé leur capacité à extraire les informations utiles des données, et à comprendre des schémas complexes. Des travaux utilisant les réseaux de neurones pour la couche physique ont été réalisés sur de la reconnaissance de modulations, sur du codage/décodage [8], de l'égalisation [5], ou encore de l'estimation de la CIR [4].

À partir des travaux existants de la littérature, nous proposons ici un algorithme hybride à base de KF et de réseaux de neurones. Les réseaux serviront à estimer les paramètres du KF au cours du temps afin de rendre viable son utilisation au sein de systèmes réels. L'architecture proposée, nommée Smart Kalman filter (SKF), est décrite dans la section 2. Dans la section 3, les différentes méthodes d'estimation des paramètres du KF sont détaillées. Puis les performances de l'architecture proposée sont mesurées dans un contexte de canaux de Rayleigh avec des trajets multiples et de brusques variations de fréquence Doppler normalisée. Ces variations traduisent notre contexte applicatif militaire, et peuvent par exemple représenter un porteur en mouvement dans un environnement évoluant rapidement ou changeant de milieu. Les performances obtenues sont ensuite comparées avec des approches usuelles dans la section 4. Enfin une conclusion sur l'architecture est proposée dans la section 5 avec une perspective des futurs travaux.

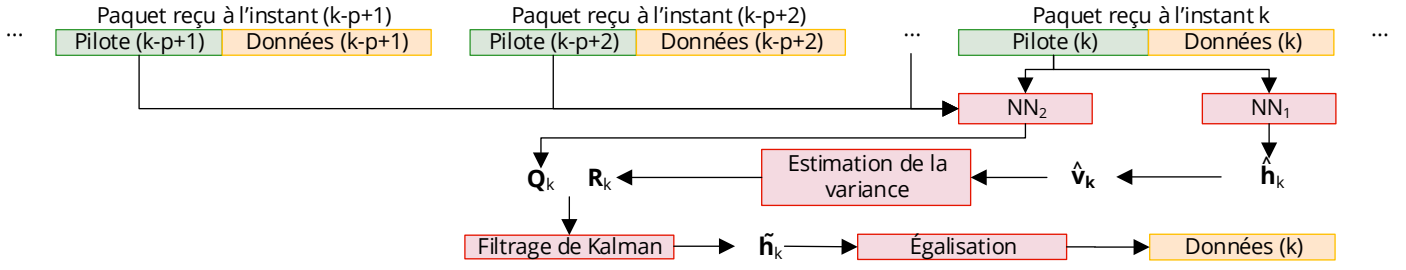


FIGURE 1 : Principe de l'architecture proposée : le Smart Kalman filter

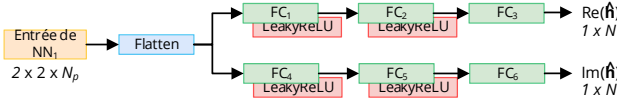


FIGURE 2 : Architecture du NN₁ qui estime \mathbf{h}

2 Présentation de l'architecture proposée : le Smart Kalman filter

Dans un contexte discret, le KF est connu pour ses capacités à suivre les paramètres d'un système linéaire \mathbf{h} dont l'évolution est donnée par l'équation

$$\mathbf{h}_k = \mathbf{F}_k \mathbf{h}_{k-1} + \mathbf{w}_k \quad (1)$$

avec \mathbf{F}_k la matrice d'évolution, et \mathbf{w}_k le bruit de processus à l'instant k , un bruit Gaussien centré et de matrice de covariance \mathbf{Q}_k . Dans notre contexte d'estimation du canal, aucune information n'est donnée a priori sur l'évolution de \mathbf{h} et donc la matrice \mathbf{F}_k est prise égale à l'identité. De plus, le processus est observé à chaque instant par une mesure \mathbf{z}_k donnée par

$$\mathbf{z}_k = (\mathbf{h}_k * \mathbf{s}) + \mathbf{v}_k \quad (2)$$

où \mathbf{h}_k est la CIR à l'instant k , \mathbf{s} la séquence de symboles pilotes émise, et \mathbf{v}_k un bruit additif. Cependant, afin d'être utilisé dans un système réel, le KF demande à chaque instant k deux paramètres : \mathbf{R}_k qui est la matrice de covariance du bruit de mesure sur l'observation \mathbf{z}_k , et \mathbf{Q}_k la matrice de covariance du bruit de processus. La matrice \mathbf{Q}_k n'a pas de sens physique et nécessite donc un paramétrage manuel, ou la création de méthodes permettant de l'adapter à chaque instant. Elle correspond au degré de variation du processus suivi entre deux instants. Dans ce papier, nous proposons une méthode basée sur l'apprentissage profond afin d'estimer les matrices \mathbf{Q}_k et \mathbf{R}_k à chaque instant k afin de bénéficier d'un KF adaptatif.

L'objectif de notre architecture est d'estimer la CIR à l'aide d'un KF combiné avec deux réseaux de neurones nommés NN₁ et NN₂ dans la suite du papier. Pour cela, les matrices \mathbf{R}_k et \mathbf{Q}_k sont estimées à chaque instant k . Le premier réseau de neurones NN₁ est utilisé pour estimer la CIR $\hat{\mathbf{h}}_k$ à partir de la séquence pilote reçue à l'instant k . $\hat{\mathbf{h}}_k$ est ensuite utilisée pour estimer la variance du bruit additif afin de construire \mathbf{R}_k . Le second réseau de neurones NN₂ utilise les p dernières séquences de symboles pilotes reçues pour prédire \mathbf{Q}_k .

La fig.1 illustre ce principe. Cela permet d'obtenir un KF dont les paramètres \mathbf{Q}_k et \mathbf{R}_k varient au cours du temps. L'architecture permet donc de bénéficier d'un KF correctement

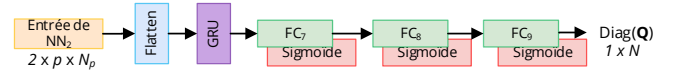


FIGURE 3 : Architecture du NN₂ qui estime la matrice \mathbf{Q}

paramétré dans le but d'obtenir une seconde estimation $\hat{\mathbf{h}}$ de la CIR. L'égalisation peut ensuite s'effectuer.

3 Méthodes d'estimation des matrices

Dans cette section, les méthodes utilisées pour estimer \mathbf{Q}_k et \mathbf{R}_k sont présentées, ainsi que les architectures des deux réseaux de neurones utilisés.

3.1 Estimation de \mathbf{R}_k

L'estimation de \mathbf{R}_k à chaque instant k est obtenue à l'aide de NN₁ qui utilise la séquence de symboles pilotes de longueur N_p reçue à l'instant k . Le réseau NN₁ est un réseau de neurones Multi Layer Perceptron (MLP), dont l'architecture est illustrée en fig.2. Le réseau prend en entrée la partie réelle et imaginaire de la séquence pilote émise et reçue puis produit une estimation de la partie réelle et imaginaire de la CIR de longueur N . Le réseau est composé d'une première couche qui vectorise l'entrée. Ensuite, il contient deux blocs de trois couches entièrement connectées, chacune activée par une fonction LeakyReLU afin d'accélérer l'entraînement du réseau de neurones. À l'instant k , la CIR $\hat{\mathbf{h}}_k$ est obtenue, et l'estimation du bruit de mesure

$$\hat{v}(i) = (\mathbf{s} * \hat{\mathbf{h}}_k)(i) - z_k(i) \quad (3)$$

est calculée pour $i \in \{0, \dots, N_p - 1\}$. La variance du bruit estimé $\sigma_{\hat{v}}^2$ peut être obtenue en utilisant l'estimateur non biaisé de la variance empirique. La matrice \mathbf{R}_k est ensuite définie par $\mathbf{R}_k = \sigma_{\hat{v}}^2 \mathbf{I}$ et peut être estimée à chaque instant.

3.2 Estimation de \mathbf{Q}_k

En considérant l'indépendance des coefficients de la CIR, la matrice de covariance \mathbf{Q}_k peut s'écrire

$$\mathbf{Q}_k = \text{Diag}(\sigma_0^2, \dots, \sigma_{N-1}^2) \quad (4)$$

Avec pour tout $i \in \{0, \dots, N - 1\}$, σ_i^2 représente la variance du $i^{\text{ème}}$ coefficient. Afin d'estimer \mathbf{Q}_k au cours du temps, un second réseau de neurone est utilisé. Son architecture est

présentée en fig.3. Le réseau consiste en une couche Gated Recurrent Unit (GRU) [6], suivie de trois couches entièrement connectées activées par une fonction sigmoïde. Cette architecture est basée sur les travaux de [3]. Contrairement au réseau présenté dans [3], nous avons remplacé la couche Long-Short-Term-Memory (LSTM) par une couche GRU afin de réduire la complexité calculatoire. Afin de prendre en compte l'évolution du canal, l'entrée du réseau est un vecteur composé des p dernières séquences pilotes reçues. La matrice \mathbf{Q}_k est obtenue en sortie du NN_2 .

4 Résultats numériques et discussion

Afin d'étudier les performances de l'architecture proposée, des simulations ont été réalisées. Dans cette section, les paramètres des simulations sont introduits. Les entraînements de NN_1 et NN_2 sont détaillés. Puis les performances du SKF sont comparées avec un IMM ainsi qu'un LS dans un contexte de canaux de Rayleigh avec du multitrajet.

4.1 Paramètres des simulations

Dans les simulations suivantes, des paquets de 256 symboles sont envoyés. Chaque paquet contient une séquence de symboles pilotes de taille $N_p = 5$, qui représente moins de 2% du paquet, et 251 symboles de données. La modulation QPSK est utilisée. La CIR est constante au sein de chaque paquet et varie entre les paquets. À la réception, une synchronisation parfaite en temps et en fréquence est supposée. Un égaliseur linéaire EQMM de longueur 15 est utilisé. Concernant le nombre de séquences pilotes p considéré pour estimer \mathbf{Q}_k au cours du temps, la valeur $p = 2$ a été fixée. De multiples expérimentations ont été réalisées avec $p \geq 3$ mais aucune amélioration des performances n'a été observée. L'IMM utilisé dans la dernière partie est composé de 4 KF avec des paramètres \mathbf{R}_k et \mathbf{Q}_k tels que $(\mathbf{R}_k, \mathbf{Q}_k) \in \{(10^{-1}, 10^{-1}), (10^{-3}, 10^{-1}), (10^{-1}, 10^{-3}), (10^{-3}, 10^{-3})\}$. Ces paramètres permettent de définir un IMM équilibré à bas et haut SNR, pour des variations lentes et rapides de la CIR. Le vecteur μ_0 de l'IMM possède ses 4 coefficients égaux à 0.25. Le vecteur μ est le vecteur de probabilité qui représente le poids des différents vecteurs d'état de chaque KF dans le vecteur d'état en sortie de l'IMM. Quant à la matrice de probabilité de transition $\mathbf{\Pi} \in \mathcal{M}_{4,4}(\mathbb{R})$, ses coefficients diagonaux sont fixés à 0.94 et ceux extra diagonaux à 0.02.

4.2 Résultats de l'estimation de \mathbf{R}_k

La matrice \mathbf{R}_k s'obtient à l'aide de NN_1 . La base de données d'entraînement de NN_1 est composée de 5×10^4 vecteurs de taille $1 \times N$ dont les L premiers coefficients sont non nuls, L uniformément tiré entre 1 et 5. Les coefficients non nuls de ces vecteurs suivent une loi normale complexe centrée réduite qui représente la CIR. L'intérêt est d'obtenir un réseau de neurones capable d'estimer la CIR d'un canal quelconque, de profondeur quelconque entre 1 et 5. Durant l'entraînement, la taille du batch est 128, l'optimiseur Adam a été utilisé et la fonction perte est l'EQM.

Concernant l'architecture de NN_1 , les tailles des couches ont été choisies empiriquement. Elles sont un compromis entre performance et complexité calculatoire. Les tailles de FC_1 et

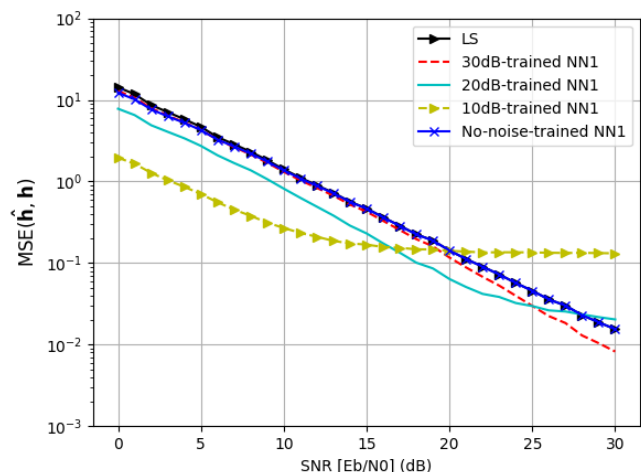


FIGURE 4 : $\text{MSE}(\hat{\mathbf{h}}, \mathbf{h})$ vs. SNR pour des entraînements de NN_1 à différents niveaux de SNR

FC_4 sont 128, FC_2 et FC_5 sont de taille 32 et enfin FC_3 et FC_6 sont de taille $N = 5$. Plusieurs entraînements ont été réalisés avec différents SNR, les résultats sont présentés en fig.4. La figure montre que les réseaux de neurones entraînés sur un SNR spécifique propose les meilleures performances au voisinage de leur SNR d'entraînement, à l'exception du réseau entraîné sans bruit additif qui semble équivalent à un LS. C'est pour cette raison que les réseaux entraînés à 10 dB, 20dB et 30 dB sont ceux retenus afin d'évaluer les performances de l'architecture globale.

4.3 Résultats de l'estimation de \mathbf{Q}_k

Les tailles des couches de NN_2 ont été également choisies de manière empirique. La taille de FC_7 est 32, FC_8 est de taille 16 et FC_9 est de taille 5. La couche GRU est composée d'une cellule GRU avec vecteur caché de taille 128.

Le NN_2 chargé d'estimer \mathbf{Q}_k a été entraîné sur des canaux variant au cours du temps, et en particulier sur des canaux de Rayleigh corrélés à spectre de Jakes qui sont les plus utilisés pour simuler une absence de trajet direct entre émetteur et récepteur. Les canaux ont été générés suivant la méthode des sinusoides, basée sur le modèle de Clarke [7]. Ce qui caractérise principalement ces canaux est la fréquence Doppler maximale normalisée \bar{f}_d . Plus elle est élevée, plus les canaux varient rapidement au cours du temps.

L'entraînement de NN_2 a été réalisé sur une base de données de taille 2×10^6 . Chaque canal de la base contient $L = 5$ trajets, et \bar{f}_d varie de manière indépendante et uniforme tous les 100 paquets dans $\{10^{-3}, 10^{-2}, 10^{-1}\}$. Le canal varie régulièrement de manière brusque afin de représenter notre contexte applicatif militaire. La taille du batch est de 200, un optimiseur Adam et la fonction perte EQM ont été utilisés. Différents entraînements ont été réalisés à plusieurs niveaux de SNR i.e. 10 dB, 20 dB et 30 dB. Une nouvelle fois, les trois réseaux ont été conservés afin d'étudier les performances de l'architecture globale.

La comparaison des performances est réalisée entre le SKF, le LS et l'IMM. Les taux d'erreurs binaires (BER pour bit error rate en anglais) sont présentés dans la fig.5. Par exemple, la notation "SKF NN_1 10dB - NN_2 20dB" désigne une architecture où NN_1 a été entraîné à 10 dB et NN_2 a été entraîné à 20

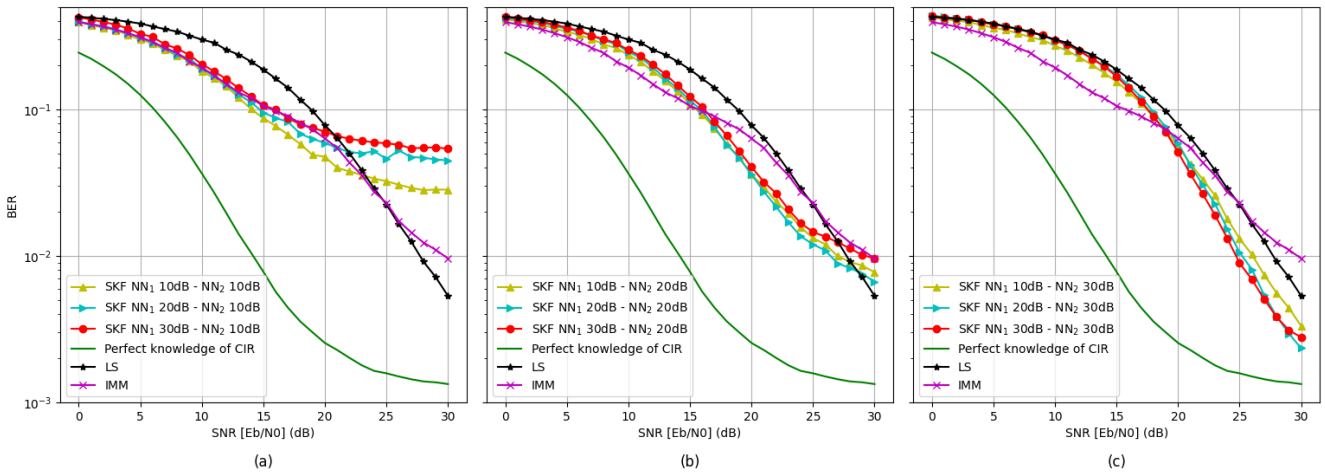


FIGURE 5 : Taux d'erreurs binaires obtenus en utilisant différents algorithmes d'estimation de la CIR, suivi d'une égalisation par MMSE. La fig.5 (a) est obtenue avec un NN_2 entraîné à 10 dB, la fig.5 (b) est obtenue avec un NN_2 entraîné à 20 dB, et la fig.5 (c) est obtenue avec un NN_2 entraîné à 30 dB.

dB. La fig.5 montre que les performances obtenues dépendent principalement du niveau de SNR d'entraînement du NN_2 . En particulier, La fig.5 (a) montre que le SKF offre à bas SNR un gain de 5 dB sur le LS, et des performances similaires à l'IMM. En revanche, la fig.5 (c) montre qu'à haut SNR, le SKF performe mieux que le LS de 3 dB et de 5 dB par rapport à l'IMM.

5 Conclusion

Dans ce papier, une architecture hybride basée sur un filtre de Kalman aidé de deux réseaux de neurones a été proposée pour l'estimation de la CIR. Les performances ont été mesurées dans un contexte de canaux de Rayleigh avec du multitrajet et des variations brusques de la fréquence Doppler normalisée. Les performances (BER) ont été comparées à celles obtenues avec une estimation du canal réalisée par un LS et par un IMM. L'étude montre la viabilité des réseaux de neurones à estimer les paramètres du KF. En effet, le SKF propose un gain entre 3 dB et 5 dB par rapport au LS, et 5 dB de gain par rapport à l'IMM à haut SNR. De plus, un autre avantage de cette architecture est le fait que les paramètres du KF, notamment la matrice \mathbf{Q} est automatiquement prédite à chaque instant. Pour la suite de l'étude, il pourrait être intéressant d'adapter cette architecture aux canaux de Rayleigh à évanouissement rapide, c'est-à-dire en considérant que la CIR varie au sein du paquet. Afin d'évaluer la viabilité et la pertinence de l'approche en vue de son implémentation dans un système embarqué, il pourrait être opportun d'optimiser les réseaux de neurones proposés.

6 Bibliographie

Références

[1] Aqiel ALMAMORI et Seshadri MOHAN : Estimation of channel state information for massive mimo based on received data using kalman filter. pages 665–669, 2018.

[2] H.A.P. BLOM et Y. BAR-SHALOM : The interacting multiple model algorithm for systems with markovian switching coefficients. *IEEE Transactions on Automatic Control*, 33(8):780–783, 1988.

[3] Changhao CHEN, Chris Xiaoxuan LU, Bing WANG, Niki TRIGONI et Andrew MARKHAM : Dynanet : Neural kalman dynamical model for motion estimation and prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 32(12):5479–5491, 2021.

[4] Chia-Hsin CHENG, Yung-Fa HUANG, Hsing-Chung CHEN et Tsung-Yu YAO : Neural network-based estimation for ofdm channels. In *2015 IEEE 29th International Conference on Advanced Information Networking and Applications*, pages 600–604, 2015.

[5] Xing CHENG, Dejun LIU, Chen WANG, Song YAN et Zhengyu ZHU : Deep learning-based channel estimation and equalization scheme for fbmc/oqam systems. *IEEE Wireless Communications Letters*, 8(3):881–884, 2019.

[6] Kyunghyun CHO, Bart van MERRIENBOER, Dzmitry BAHDANAU et Yoshua BENGIO : On the properties of neural machine translation : Encoder-decoder approaches, 2014.

[7] R. H. CLARKE : A statistical theory of mobile-radio reception. *The Bell System Technical Journal*, 47(6):957–1000, 1968.

[8] I. ORTUNO, M. ORTUNO et J.A. DELGADO : Error correcting neural networks for channels with gaussian noise. In *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, volume 4, pages 295–300 vol.4, 1992.

[9] Manisha B. SUTAR et Vikram S. PATIL : Ls and mmse estimation with different fading channels for ofdm system. 1:740–745, 2017.