A novel LLR scaling factor selection using LDL Bayes classifier

Seunghun Yu, Kwonyeol Park, Jongmin Cho,
Kyusuk Mo, Shichang Noh and Min-Ho Shin
CP S/W Development Team
Samsung Electronics.,
Hwasung-si, Korea Republic
{sh0703.yu, kwon10.park, jongmin5.cho, kyusuk1127.mo, sichang.no, minho.shin}@samsung.com

***Abstract*—In this paper, we suggest a novel learning algorithm that determines a scaling factor to adjust the log likelihood ratio (LLR) extracted from the symbol detector. The LLR scaling factor to overcome the constraints of HW can affect the performance depending on the determined value. It has the ambiguity that selects the value of LLR scaling factor since determining the value would be changed rapidly under the small difference of environment. On the other hand, among machine learning types, label distribution learning (LDL) is known to be beneficial for learning ambiguous data because it distinguishes the differences regarding each label as the distribution. After the proposed learning algorithm using LDL Bayes classifier learns the ambiguity of each scaling factor, the learned Bayes classifier determines the LLR scaling factor offering higher performance. As a result of the simulation, the proposed learning algorithm provides improved performance compared to the existing fixed LLR scaling factor.**

***Index Terms*—Symbol detection, LLR, scaling factor, LDL, Bayes classifier**

I. Introduction

In the 5G new radio (NR) networks are being widely displayed worldwide, both academia and industry have started to move beyond 5G and explore 6G networks. The 6G network will be designed as a more complex and sophisticated system beyond the three main concepts in 5G, i.e., 1) Enhanced Mobile Broadband (eMBB), 2) Ultra-Reliable and Low latency communication (URLLC) and 3) Massive Machine Type Communication (mMTC) [1]. In order to achieve higher performance and much robustness, challenging attempts are being continued in both academia and industry to introduce various machine learning technologies into wireless communication systems. Moreover, most wireless chipset manufacturers are installing neural processing units (NPUs) and high-performance graphics processing units (GPUs) on chipsets to commercialize machine learning that requires high complexity and computation [2].

Aforementioned the above, wireless communication adopting deep learning has been introduced after reducing the computation effort. Nowadays, research on applying machine learning to signal processing has been also actively conducted in the field of wireless communication. As multi-input multi-output (MIMO) is one of the important elements in current cellular systems, [3] devises deep machine learning-based MIMO detection. In addition, the proposed detection shows significant accuracy under ill and varying channels. Various fancy applications of machine learning are also discussed to adopt the physical layer in [4]. Their application of convolutional neural networks (CNN) for modulation classification achieves satisfying accuracy compared to the present scheme. In addition, [5] results in better performance than the previous interference whitening (IW) scheme thereby utilizing reinforcement learning based IW mode.

However, there are still many domains in wireless communication where machine learning has not been applied into, since only recently has collaboration research between wireless communications and machine learning been established [6]. One of those domains is to apply into currently implemented hardware systems. For instance, quantizing LLR values is more useful since the variation in these values extracted from the symbol detector is so huge that it increases complexity. When LLR is quantized, an additional scaling factor is considered to maintain its property without transformation and ensure high performance from decoder's perspective [7]. By the way, selection of scaling factors for reducing variation of LLR is highly ambiguous because many parameters affect this scaling factor. Machine learning can also be applied to currently designed communication systems.

Therefore, we propose the learning algorithm to apply a LDL-based Bayesian classifier for selection of LLR scaling parameters in symbol detection. LDL is a type of machine learning designed to learn the ambiguity inherent in labels by learning the probability distribution of each label [8]. Also, this classifier is learned using data from the actual terminal in this paper. In the section of simulation results, the proposed learning algorithm results in improved performance in terms of block error rate (BLER) than the result of fixed scaling factor. In other words, the simulation is conducted based on the mobile system currently in use and obtains meaningful results.

II. System Model

We consider the User (UE) has M antennas and receives the transmitted signal from M antennas of the base station (BS). There is $M \times 1$ a data stream vector $\boldsymbol{x} = [x_0, x_1, \ldots, x_{M-1}]^T$ transmitted through each transmitted antenna as a spatial multiplexing scheme. Thus, we express $M \times 1$ received signal vector $\boldsymbol{y} = [y_0, y_1, ..., y_{M-1}]^T$ as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n} \\ &= \sum_{i=0}^{M-1} \boldsymbol{h}_i x_i + \boldsymbol{n} \end{aligned} \tag{1}$$

where $\boldsymbol{H} = [\boldsymbol{h}_0, \boldsymbol{h}_1, ..., \boldsymbol{h}_{M-1}]$ is $M \times M$ channel coefficient matrix and $\boldsymbol{n} \sim \mathcal{CN}(0, \sigma_n^2)$ is the additive white Gaussian noise (AWGN) with covariance $E\{\boldsymbol{n}\boldsymbol{n}^H\} = \sigma_n^2 \boldsymbol{I}$. Suppose that the i-th layer's transmit symbol x_i is from some constellation $\mathbb{C}_i$ which is one of the modulations among QPSK, 16QAM, 64QAM and 256 QAM with 2^{m_i} points. $m_i = 2, 4, 6, 8$ is mapped in to QPSK, 16QAM, 64QAM and 256QAM respectively and symbol x_i is normalized such that $E\{|x_i|^2\} = 1$. Each modulated symbol x_i carries a $m_i \times 1$ bit vector and $b_{i,l}$ is defined as the l-th bit of x_i. Assuming no prior information and receiver to generate the LLR for bit with Log-MAP (maximum a posterior), $b_{i,l}$ is written as

$$\begin{aligned} \Lambda(b_{0,l}) &= \log \frac{P(b_{0,l} = 0|\boldsymbol{y})}{P(b_{0,l} = 1|\boldsymbol{y})} \\ &= \log \frac{P(\boldsymbol{y}|b_{0,l} = 0)P(b_{0,l} = 0)}{P(\boldsymbol{y}|b_{1,l} = 0)P(b_{1,l} = 0)} \\ &= \log \frac{\sum\limits_{x_0:b_{0,l}=0} \sum\limits_{x_1} \cdots \sum\limits_{x_{M-1}} P(\boldsymbol{y}|x_0, x_1, ...x_{M-1})}{\sum\limits_{x_0:b_{1,l}=0} \sum\limits_{x_1} \cdots \sum\limits_{x_{M-1}} P(\boldsymbol{y}|x_0, x_1, ...x_{M-1})} \\ &= \log \frac{\sum\limits_{x_0:b_{0,l}=0} \sum\limits_{x_1} \cdots \sum\limits_{x_{M-1}} e^{-\frac{||\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}||^2}{\sigma^2}}}{\sum\limits_{x_0:b_{1,l}=0} \sum\limits_{x_1} \cdots \sum\limits_{x_{M-1}} e^{-\frac{||\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}||^2}{\sigma^2}}} \end{aligned} \tag{2}$$

where $P(\boldsymbol{y}|\boldsymbol{x}) \sim \exp\left(-\frac{||\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}||^2}{\sigma^2}\right)$. To obtain a simple solution from the above equation, we use dimension reduced max log-MAP which is a method of approximation in the same way as [9]. Consequently, LLR is represented as following

$$\Lambda(b_{0,l}) = \frac{1}{\sigma^2}\left\{\min_{b_{1,l}=1} ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2 - \min_{b_{0,l}=1} ||\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}||^2\right\} \tag{3}$$

Meanwhile, because LLR has a huge variation in use as above mentioned, LLR value for hardware implementation is quantized. In order to maintain LLR's properties and high performance, a scaling parameter is multiplied before quantizing the values as the following

$$\Lambda(\cdot)' = \Lambda(\cdot)2^q \tag{4}$$

where q is scaling factor. Especially, only a certain extent value which hardware can control is extracted from equation (4). In other words, the scaling factor q is important to include the characteristics of the original LLR value. According to [10], when $\Lambda(\cdot)'$ is quantized on $2n+1$ bits, quantification function Ψ is defined with the interval of the quantization A as

$$\Psi(\Lambda(\cdot)') = sat\left(\lfloor \Lambda(\cdot)' \cdot \frac{2^{2n} - 1}{A} + 0.5 \rfloor, 2^{2n} - 1\right) \tag{5}$$

where $sat(a, b) = a$ if a is into $[-b, b]$ and $sat(a, b) = sign(a) \cdot b$ otherwise.

A. Label Distribution Learning

LDL has two types of learning such as single label learning (SLL) and multi label learning (MLL). SLL is that one label is determined by a training data instance and MLL is that various labels can be determined through a training data instance. We let instance be a characteristic of the channel and map each range of LLR scaling factor q value to a label. Especially, the instance is a vector consisting of SINR (Signal-to-interference-plus-noise ratio), modulation order, new transmission and layers. Total number of labels is $\boldsymbol{c}$ and each label can be expressed as probability distribution according to specific instance $\boldsymbol{a}_i$. Here, each label is mapped into the number of configurable scaling factors q. Therefore, the distribution of scaling factor q is equal to the label distribution through the specific instance $\boldsymbol{a_i}$. Assuming that $d_{\boldsymbol{a}_i}^{q_j}$ is the probability of j-th scaling factor q_j corresponding to instance $\boldsymbol{a_i}$, data set D_i including all of the probability for scaling factors q corresponding to $\boldsymbol{a_i}$ can be expressed as $D_i = \{d_{\boldsymbol{a}_i}^{q_1}, d_{\boldsymbol{a}_i}^{q_2}, ..., d_{\boldsymbol{a}_i}^{q_c}\}$ and depicted in Fig.1. In addition, since D_i is the data set of distribution, the following equation is satisfied.

$$\sum_{D_i} d_{\boldsymbol{a}_i}^{q_j} = 1 \tag{6}$$

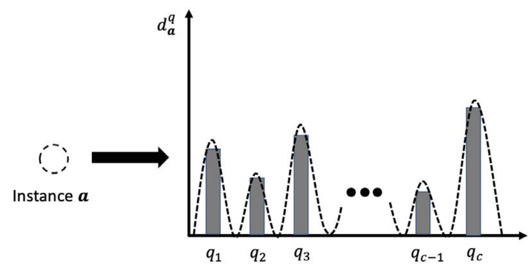

Fig. 1. Distribution probability of LLR scaling factor given instance

III. Proposed selection algorithm

It is necessary to define the probability regarding each scaling factor q_j. In the sake of this, the estimator should be proposed to compare which scaling factor q_j bring the high quality signal in LLR form. We define the high quality signal LLR form before proposing the estimator. Assuming that the LLR value is quantized by $2n+1$ units as equation (5) and there are total L bits from symbol detector, the set of the number of quantized LLR is expressed as $\mathcal{L} = \{\Delta_1, \Delta_2, ..., \Delta_{2n+1}\}$. Also, Δ_k is notated as following

$$\Delta_k = \sum_{l=1}^{L} \delta_{\Psi(\Lambda(b_{0,l})'),k} \tag{7}$$

$$\delta_{\Psi(\Lambda(b_{0,l})'),k} = \begin{cases} 1 & \text{if } \Psi(\Lambda(b_{0,l})') = k \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where $\delta_{\Psi(\Lambda(b_{0,l})'),k}$ is kronecker delta function. If the l-th bit is quantized as k, Δ_k would be only counted. LLR is the output from the symbol detector and the high quality signal is expected to be clustered close to both ends such as Δ_1 and Δ_{2n+1}. In other words, assuming set Δ_{n+1} to value 0 and distribute the rest of LLR symmetrically from Δ_{n+1}. The more LLR is distributed at the ends such as Δ_1 and Δ_{2n+1}, we expect that it is the better signal quality. The good quality signal's LLR distribution (Q) is represented thereby putting the half of total number of LLRs at both ends respectively and loss weighting the rest. The following probability $P(\Delta_k, Q)$ indicates the probability of distribution in the corresponding quantized value Δ_k regarding LLS distribution Q.

$$P(\Delta_k, Q) = \begin{cases} \frac{L-1}{2L} & \text{if } k = 1 \text{ or } 2n+1, \\ \tau_{k-1} P(\Delta_1, Q) & \text{if } 2 \leq k < n+1, \\ \tau_{2n+1-k} P(\Delta_1, Q) & \text{otherwise.} \end{cases} \tag{9}$$

Here $\tau_i = 2^{-i}$ is loss weighting constant to represent the distribution of good signal LLR. Next, the distribution Q_j when the scaling factor q_j is selected is expressed as follows as defined above notation (6).

$$P(\Delta_k, q_j) = \frac{\Delta_k}{\sum_{i=1}^{2n+1} \Delta_i} \text{ , } \Delta_k \in \mathcal{L} \tag{10}$$

We need to denote the estimator for calculation of the difference between the LLR of high quality signal Q and the LLR derived from scaling factor q_j. Thus, we introduce the estimator $E(\cdot|\cdot)$ based on Kullback–Leibler divergence (KLD) which is applied to measure the differences between two distributions.

$$E(Q_j|\boldsymbol{a}_i) = \sum_{k=1}^{2n+1} f(k|Q_j, \boldsymbol{a}_i) \tag{11}$$

where $f(k|Q_j, \boldsymbol{a}_i)$ is not the same with KLD in that the calculation is different at $k = 1$ and $k = 2n+1$ as the below.

$$f(k|Q_j, \boldsymbol{a}_i) = \begin{cases} P(\Delta_k, q_j) \log \frac{P(\Delta_k, q_j)}{P(\Delta_k, Q)} & \text{if } P(\Delta_k, q_j) \\ & > P(\Delta_k, Q) \\ & , k = 1 \text{ or } 2n+1 \\ P(\Delta_k, Q) \log \frac{P(\Delta_k, Q)}{P(\Delta_k, q_j)} & \text{else} \end{cases} \tag{12}$$

Therefore, $E(\cdot, \cdot)$ measures values using LLR derived when a specific scaling factor is selected under instance $\boldsymbol{a}_i$. In order to satisfy the distribution characteristics from (6), measured value should be expressed as the probability. Applying softmax activation function, the probability of measured value is written as

$$p(q_j, \boldsymbol{a}_i) = d_{\boldsymbol{a}_i}^{q_j} = \frac{\exp\left(E(q_j, \boldsymbol{a}_i)\right)}{\sum_{j=1}^{c} (\exp(E(q_j, \boldsymbol{a}_i)))} \tag{13}$$

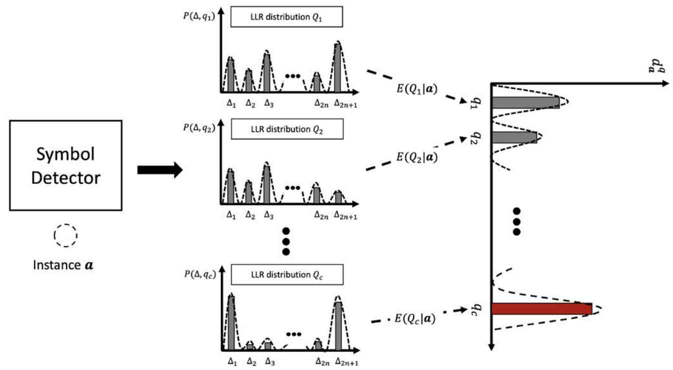

Fig. 2. Process of extracting label distribution

A. *Proposed Selection Algorithm Of LDL Based-Bayes Classifier*

Applying the notations which are introduced as the previous, we suggest the learning algorithm in this subsection. First of all, collect various data sets that consist of an instance and the distribution corresponding to each scaling factor. When specific instance $\boldsymbol{a}_i$ is given, distribution of LLR is calculated through equation (10) regarding scaling factor q_j respectively. Then, the probability of each scaling factor is determined from distribution of LLR adopting equation (12) and (13). According to the above, Fig.2 diagrammatically can show the process of creating a data set by representing each scaling factor as a probability with the LLR extracted from the symbol detector. One data set D_i is determined for instance $\boldsymbol{a}_i$, so assuming there are as many instances as $\mathcal{I}$, $\mathcal{I}$ data sets would be collected. Next step, we let naive Bayes classifier learn label distribution through data sets acquired from the first step. Finally, LDL based-Bayes classifier chooses the scaling factor which has the highest probability compared to the others when new instance $\boldsymbol{a}$ is given. All of the steps regarding the proposed selection algorithm of LDL based-Bayes classifier are summarized as $\mathbf{Algorithm}1$

Algorithm 1 Proposed Selection Algorithm of LDL Based-Bayes Classifier

Input: Data set $D_i = \{(\boldsymbol{a}_i, d_{\boldsymbol{a}_i}^{q_1}), (\boldsymbol{a}_i, d_{\boldsymbol{a}_i}^{q_2}), ...(\boldsymbol{a}_i, d_{\boldsymbol{a}_i}^{q_c})\}$
    naive Bayes classifier $\mathcal{NB}$
    Number of learning iteration $\mathcal{I}$
Process:
    for $i = 1, 2, \cdots, \mathcal{I}$:
        fitting naive Bayes classifier$\mathcal{NB}$
    end
Output: $q_{j^*} = \text{argmax}_{d_{\boldsymbol{a}}^{q_j}} \mathcal{NB}(D)$

IV. Simulation Results

In this section, simulation results are shown when the scaling factor is selected adopting LDL based-Bayes classifier proposed in the previous and when the scaling factor value is fixed. The data sets to be used in the simulation are collected based on the test results utilizing Samsung's communication processor chip (S5E8535) and Anritsu's MT8000A which is RF test solution for 5G base stations. Due to the characteristics of LDL, data should be required regarding each label. Thus, tests are conducted when selecting each scaling factor under the same environment. Detailed test parameters are shown in TABLE I.

TABLE I
SYSTEM PARAMETERS FOR COLLECTING DATA SETS

System parameter	
Test equipment	MT8000A (Anritsu corp.)
Test terminal	Exynos 1330* (S5E8535)
Radio access network	NR (stand alone)
Band	n78
Band width	100MHz
Layers	1
MCS (Modulation order)	22 (256QAM), 11 (64QAM), 5 (16QAM), 2 (QPSK)
HARQ	Not supported
Transmit power control	-0.5 unit [dBm]
LLR scaling factor (q)	0, 1, 2, 3
Distribution set of LLR	9 ($\mathcal{L} = \{\Delta_1, ..., \Delta_9\}$, $n = 4$)

Parameters based on 3gpp release 15
Exynos 1330* (S5E8535): Samsung Electronics' SoC chipset

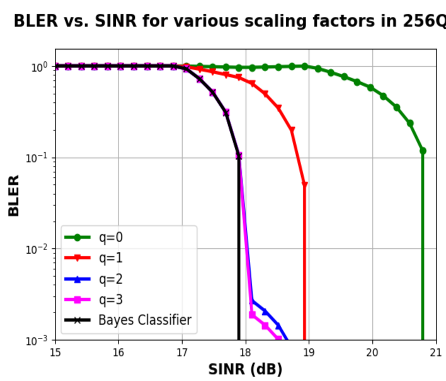

Fig. 3. BLER vs. SINR for various scaling factors in 256QAM

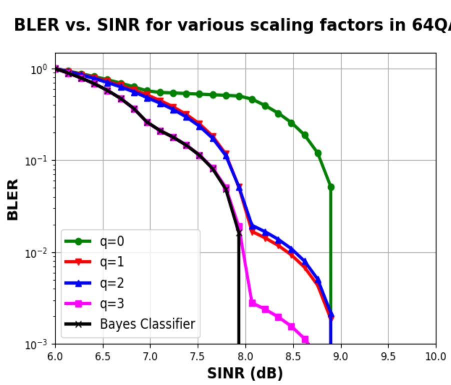

Fig. 4. BLER vs. SINR for various scaling factors in 64QAM

When simulation is proceeding, SINR, layers, the number of retransmission and modulation order are the elements of the instance vector from data sets. As in the system parameter table, the key elements to affect learning are SINR and modulation order since the number of layers is 1 and HARQ is not supported. The following from Fig.3 to Fig.6 show the simulation results of BLER performance when fixing each scaling factor $q = \{0, 1, 2, 3\}$ and LDL based-Bayes classifier selects the scaling factor. As $q = 0$ is 20.8 dB level, $q = 1$ is 18.9 dB, $q = 2, 3$ is 18 dB level, and LDL based-Bayes classifier is 17.9 dB level in BLER 10^{-2} of Fig.3. respectively, LDL based-Bayes classifier get approximated 0.1 dB performance gain compared to best fixed scaling factor $q = 2, 3$. Although LDL based-Bayes classifier also selects one of scaling factors among 0, 1, 2 and 3, the gain difference was seen due to selecting a scaling factor stochastically. Also, Fig.4 shows that LDL based-Bayes classifier gets 0.1dB improved performance compared to $q = 3$ at the BLER $10^{-2.5}$. LDL based-Bayes classifier shows the same BLER performance as $q = 2, 3$ and 4 except for BLER from $10^{-1.5}$ to $10^{-2.5}$ in Fig.5. In the case of 16QAM, the scaling factor $q = 0$ bring much lower performance compared to other scaling factors $q = 1, 2, 3$. Therefore, LDL based-Bayes classifier seems to be similar to the BLER pattern with scaling factors $q = 1, 2, 3$. For the BLER 10^{-2} in Fig.6, the scaling factor $q = 3$ and LDL based-Bayes classifier indicate SINR -3.1 dB level, whereas scaling factor $q = 0, 1, 2$ shows SINR -0.9dB level equally. In the case of QPSK, the scaling factor $q = 3$ is

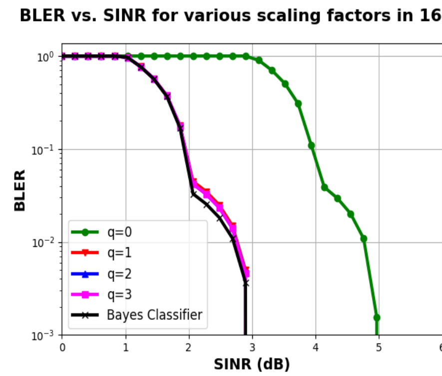

Fig. 5. BLER vs. SINR for various scaling factors in 16QAM

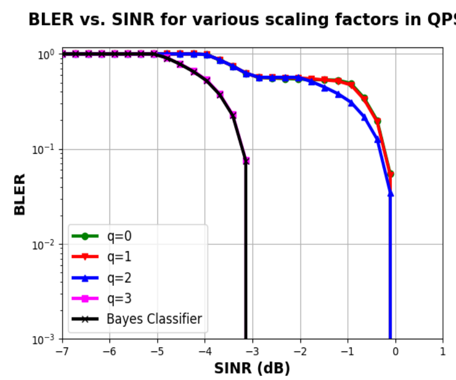

Fig. 6. BLER vs. SINR for various scaling factors in QPSK

much better compared to other scaling factors. Therefore, LDL based-Bayes classifier seems to show the same results as selecting only $q = 3$. As a result, through the above simulation, adopting the proposed LDL based-Bayse classifier improves performance compared to the existing fixed scaling factor. Also, if the significantly better scaling factor exists compared to other scaling factors in terms of the performance, the proposed scheme has the same effect as selecting this scaling factor.

V. Conclusion

In this paper, it is devised to apply machine learning in terms of actual operation of wireless terminals. In particular, when extracting the LLR from the block of symbol detector in the wireless terminal, we focus on selecting the scaling factor applied to compensate for the hardware limitation. Since there was ambiguity in selecting this scaling factor, we propose the algorithm to which machine learning technique called LDL is applied. In addition, data sets are collected through an actual device adopted by Exynos 1330 (S5E8535) and MT8000A which is Anritsu's protocol call box. Simulation results are shown by learning proposed learning algorithms and using those extracted data sets. In the section of simulation results, the proposed algorithm, LDL based-Bayes classifier, accomplishes the better performance compared to other fixed scaling factors. As a future work, we are going to apply and model to other wireless communication blocks that have ambiguities in the selection of parameters.

References

[1] E. Dahlman and S. Parkvall, "5G NR: The Next Generation Wireless Access Technology," , Academic Press, Aug. 2018.

[2] C. Jiang et al., "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[3] N. Samuel, T. Diskin, and A. Wiesel, "Deep MIMO detection," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sapporo, Japan, 2017, pp. 1–5.

[4] T. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[5] K. Park, P. *et al.*, "The Reinforcement Learning based Interference Whitening Scheme for 5G," in *Proc. IEEE 93rd Veh. Technol. Conf. (VTC Spring)*, Helsinki, Finland, 2021, pp. 1–5.

[6] C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Communications Surveys & Tutorials*, vol.21, issue. 3, pp. 2224-2287, Mar. 2019.

[7] J. Sidon, Y. Alon and S. Serfaty, "Wireless communication unit and method for receiving a wireless signal," WIPO (PCT). Patent No. 2007092744A2, Feb., 2007.

[8] X. Geng, "Label Distribution Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol.28, issue. 7, pp. 1734-1748, Mar. 2016.

[9] K. Hwang, S. Yoon and Y. Kwon, "Log-likelihood ration (LLR) generating apparatus and method in Multiple Input Multiple Output (MIMO) antenna communication system," U.S. Patent No. 7746951B2, June, 2010.

[10] D. Declercq, M. Fossorier and E. Biglieri (Editor), "Channel Coding: Theory, Algorithms, and Applications: Academic Press Library in Mobile and Wireless Communications," , Academic Press, Aug. 2013.