

Machine Learning in 6G Wireless Communications

Tomoaki OHTSUKI^{†a)}, *Fellow*

SUMMARY Mobile communication systems are not only the core of the Information and Communication Technology (ICT) infrastructure but also that of our social infrastructure. The 5th generation mobile communication system (5G) has already started and is in use. 5G is expected for various use cases in industry and society. Thus, many companies and research institutes are now trying to improve the performance of 5G, that is, 5G Enhancement and the next generation of mobile communication systems (Beyond 5G (6G)). 6G is expected to meet various highly demanding requirements even compared with 5G, such as extremely high data rate, extremely large coverage, extremely low latency, extremely low energy, extremely high reliability, extreme massive connectivity, and so on. Artificial intelligence (AI) and machine learning (ML), AI/ML, will have more important roles than ever in 6G wireless communications with the above extreme high requirements for a diversity of applications, including new combinations of the requirements for new use cases. We can say that AI/ML will be essential for 6G wireless communications. This paper introduces some ML techniques and applications in 6G wireless communications, mainly focusing on the physical layer.

key words: artificial intelligence (AI), machine learning (ML), deep learning (DL), neural network (NN), deep neural network (DNN), 6G, deep transfer learning (DTL)

1. Introduction

Digital Transformation (DX), which transforms society, economy, and industry using digital technology represented by rapidly developing AI (Artificial Intelligence), is attracting much attention. Information and Communication Technology (ICT) infrastructure plays an important role in DX. It is no exaggeration to say that mobile communication systems, represented by the 5th generation mobile communication system (5G), are the core of the ICT infrastructure. 5G is expected for various use cases in industry and society. 5G has three functional requirements: enhanced Mobile Broadband (eMBB), Ultra-Reliable and Low Latency Communications (URLLC), and Massive Machine Type Communications (mMTC). In 5G (New Radio (NR) Release 15), which is currently in service, best-effort services that emphasize downlink speed are mainly realized as a result of standardization in 3GPP, focusing on eMBB and some URLLC among them [1]. In the future, it is expected that services that take advantage of large data uploads and services that guarantee communication quality, particularly for industrial applications, will be required. Therefore, 5G Enhancement

is expected to improve the performance of the uplink and realize communication quality assurance.

Mobile communication systems are evolving every 10 years, and by 2030, when the next generation of mobile communication systems (Beyond 5G (6G)) is expected to be in use, various social issues and use cases are expected to be addressed. As shown above, 6G will be required to support the data traffic that is expected to continuously increase as mobile communication services become more sophisticated and diverse. Also, 6G will be required to meet the extremely high-performance requirements that will support the resolution of social issues and new use cases in the 2030s. The Ministry of Internal Affairs and Communications (MIC) has presented the following three social images for the 2030s when 6G is expected to be used in “Beyond 5G Promotion Strategy — Roadmap towards 6G —” [2]: “Inclusive society,” “Sustainable society,” and “Dependable society.”

The year 2030 is also the target year for achieving the Sustainable Development Goals (SDGs) adopted at the United Nations Summit in 2015. 6G is expected to support the realization of these goals as a social infrastructure. According to the “Beyond 5G Promotion Strategy,” in addition to further upgrading of the characteristic functions of 5G, such as eMBB, URLLC, and mMTC, 6G must be equipped with four new functions: “ultra low power consumption,” “autonomy,” “scalability,” and “ultra security and resiliency.” In addition to the above functions, [1] also lists lower cost (lower cost per bit) and sensing as requirements.

To meet those high requirements, various techniques need to be developed and used in 6G. Several companies and research institutes have issued white papers about B5G and 6G [1]–[8]. In those white papers, we can see many common requirements as shown below such as in [1].

- Extreme high data rate/capacity
 - Peak data rate > 100 Gbps exploiting new spectrum bands
 - > 100× capacity
 - Extreme-high uplink capacity
- Extreme low latency
 - E2E very low latency < 1 ms
 - Always low latency
- Extreme coverage extension
 - Gbps coverage everywhere
 - New coverage areas, e.g., sky (10000 m), sea

Manuscript received July 8, 2022.

Manuscript revised July 20, 2022.

Manuscript publicized August 10, 2022.

[†]The author is with the Faculty of Science and Technology, Keio University, Yokohama-shi, 223-8522 Japan.

a) E-mail: ohtsuki@ics.keio.ac.jp

DOI: 10.1587/transcom.2022CEI0002

(200 NM), space, etc.

- Extreme high reliability
 - Guaranteed QoS for wide range of use cases (upto 99.99999% reliability)
 - Secure, private, safe, resilient, . . .
- Extreme low energy & cost
 - Affordable mmW/THz NW & devices
 - Devices free from battery charging
- Extreme massive connectivity
 - Massive connected devices (10 M/km²)
 - Sensing capabilities & high-precision positioning (< 1 cm)

To meet these high requirements, many key techniques are mentioned in the white papers such as in [5].

- AI/ML-driven air interface design and optimization
- Expansion into new spectrum bands and new cognitive spectrum sharing methods
- The integration of localization and sensing capabilities into system definition
- The achievement of extreme performance requirements on latency and reliability
- New network architecture paradigms involving sub-networks and RAN-core convergence
- New security and privacy schemes

As mentioned above, AI/ML will be essential for 6G wireless communications with the extremely high requirements for a diversity of applications, including new combinations of the requirements for new use cases. Note that AI is a simulation of human intelligence or experience by machines, while ML is an application of AI with the ability to automatically learn and improve from experience without being explicitly programmed. AI is a much broader concept than ML. A white paper on ML in 6G wireless communication networks has been issued by University Oulu [8]. In the white paper, several applications of ML in each layer are presented, physical layer, medium access control (MAC) layer, application layer, and also for the security of wireless networks. As the application of ML at the physical layer, the following areas are shown: channel coding, synchronization, positioning, channel estimation, beamforming, and physical layer optimization with ML. Also, in the application of ML at the MAC layer, the following use cases are shown: Federated Learning (FL) for orientation and mobility prediction in wireless virtual reality networks, predictive resource allocation in machine-type communications, predictive power management, and asymmetric traffic accommodation.

In [9] more applications of AI/ML in each layer are presented, physical layer, network layer, and application layer. As the application of ML at the physical layer, the following areas are shown: Channel tracking/equalization/decoding, pathloss prediction/estimation, intelligent beamforming, modulation mode selection, anti-jamming, channel access

control, spectrum sensing/management/allocation, physical-layer security, and so on. Also, in the application of ML at the network layer, the following areas are shown: Caching, traffic classification, anomaly detection, throughput optimization, latency minimization, attack detection, intelligent routing, traffic prediction/control, access control, source encoding/decoding, and so on.

As shown above, there are ongoing standardization efforts to exploit AI in cellular systems. The third generation partnership project (3GPP) has standardized a network data analytics function (NWDAF) for data collection and analytics in automated cellular networks [10]. In addition to 3GPP, the O-RAN Alliance is targeting to realize an intelligent radio access network (RAN). Networks are now required to support a wide variety of applications and becoming increasingly complex. In such a case, it may become difficult to optimize operations and networks manually as in the past. It will become essential to realize more autonomous and automated operations utilizing AI and ML. To realize such a vision, the O-RAN Alliance is studying RAN configurations (architectures) that can optimize network design and operations while utilizing AI/ML, as well as open interfaces are also being considered. The function called “RIC (RAN Intelligent Controller)” specified by the O-RAN Alliance is positioned at the center of the realization of this intelligent RAN.

In this paper, we introduce some ML techniques and their applications in 6G wireless communications, mainly focusing on the physical layer. We first introduce end-to-end learning of communication systems through neural network-based autoencoders [11]. We then introduce some ML techniques for massive multiple-input multiple-output (MIMO). We introduce a neural network-based belief propagation (BP) algorithm for massive MIMO signal detection [12], [13]. This algorithm is based on the idea of deep unfolding that unfolds the iterations of an inference algorithm into a layer-wise structure like a neural network [14]. We also introduce a signal detection based on the BP algorithm with a deep learning (DL)-based denoising technique based on the deep image prior (DIP) [15]. In massive MIMO systems, as the name mentions, the number of antennas is large so the number of channels that needs to be estimated is also large. Due to the time-varying characteristics of the channel, the length of pilot signals is limited so that that of orthogonal pilot signals is finite. Thus, the same pilot signals are reused in neighboring cells, which deteriorates the channel estimation performance. This defect is referred to as pilot contamination. We introduce two neural network-based schemes to reduce the effects of pilot contamination [16]. We also consider channel state information (CSI) feedback, where the amount of feedback information is the issue in massive MIMO. We introduce the neural network-based CSI feedback scheme, where we explain the idea of deep transfer learning (DTL) and present the DTL-based CSI feedback scheme [17].

In B5G and 6G, it is also essential to utilize new spectrum bands such as mmWave bands and tera-hertz frequency

bands to achieve an extremely high data rate. However, the systems using high-frequency bands, such as the mmWave systems, suffer from severe pathloss. Thus, it is essential to use beamforming with large antenna array gains in mmWave communications. Therefore, it needs to use a lot of antennas in mmWave communications. Since the power consumption and cost of radio-frequency (RF) chains are both high, the current mmWave systems employ not full digital beamforming, in which each antenna is attached to an RF chain, but hybrid beamforming in general. Hybrid beamforming is effective but requires a large overhead in a beam search/selection phase. We introduce our proposed DL-based analog beam selection scheme with low overhead [18]. Finally, we conclude this paper and discuss the future direction of ML in 6G wireless communications.

Note that some wireless scenarios mentioned in this paper are also studied in 5G. However, as mentioned above, there are new extremely high requirements for new use cases in 6G. Also, there are new combinations of the requirements for new use cases in 6G. Those are the differences between 5G and 6G even for the same wireless scenarios.

2. AI-Based Wireless Communications

2.1 End-to-End Learning of Communication Systems Through Neural Networked-Based Autoencoders

In general, a communication system consists of a transmitter, a channel, and a receiver, where a transmitter and a receiver are split into multiple signal processing blocks. Conventional signal processing tries to optimize each block separately or sometimes jointly. The block-based optimization provides a good performance in general. However, block-based optimization does not always provide the best possible end-to-end performance. Joint optimization can provide better performance than block-based optimization in general. However, joint optimization of multiple signal blocks is often computationally prohibitive. A learned end-to-end optimization through deep learning can provide superior performance.

The general communication systems mentioned above can be seen as a kind of autoencoder [19]. Thus, an autoencoder has been used to model the communication system and optimize the communication system in an end-to-end manner as shown in Fig. 1 [11]. In general, an autoencoder is used to find a lower-dimensional representation of its input at an intermediate layer like compression, while it still can reconstruct the input signal as the output of the decoder. On the other hand, the autoencoder modeling the communication system tries to learn representations of the transmitted signals \mathbf{x} of the messages \mathbf{s} that are robust to the channel impairments, such as noise, fading, distortion, and so on, so that the transmitted message can be recovered with a small probability of error. This autoencoder is sometimes referred to as the ‘‘channel autoencoder.’’ [11] presents a comparison of the block error rate (BLER) performance between Hamming (7,4) coded binary phase-shift keying (BPSK) with

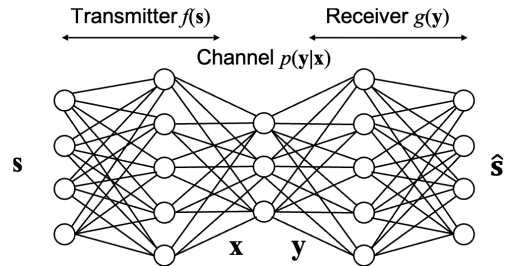


Fig. 1 Modeling the communications system by autoencoder, where \mathbf{s} is the transmitted information, \mathbf{x} is the transmitted signal, \mathbf{y} is the received signal, $\hat{\mathbf{s}}$ is the estimated transmitted information, $f(\cdot)$ is the transmitter function, $g(\cdot)$ is the receiver function, and $p(\mathbf{y}|\mathbf{x})$ is the conditional probability density function of the channel.

maximum likelihood (ML) decoding and that of the trained autoencoder (7,4). The autoencoder is shown to achieve the same BLER performance as that of the Hamming (7,4) code with ML decoding. Thus, it can be said that the autoencoder has learned the encoder and decoder function without any prior knowledge.

One of the drawbacks of the end-to-end learning of communication systems through autoencoders is that the gradient of the instantaneous channel transfer function must be known, which is not practical. Moreover, the channel usually comprises some processes of the transmitter and the receiver, such as quantization, which are non-differentiable, and thus the gradient-based training through backpropagation cannot be used.

To overcome these drawbacks, [20] proposes a learning algorithm that enables the training of communication systems with an unknown channel model or with non-differentiable components. The proposed algorithm iterates between the training of the receiver using the true gradient of the loss, and that of the transmitter using an approximation of the loss function gradient. The proposed algorithm is shown to achieve the performance identical to that of the algorithm with training with a channel model using backpropagation on additive white Gaussian noise (AWGN) and Rayleigh block-fading (RBF) channels.

2.2 Deep Learning for Massive MIMO

2.2.1 Massive MIMO Detection

Massive MIMO that uses a massive number of antenna elements on the transmitter side is one of the key technologies of 5G and 6G. Massive MIMO can achieve high spectral efficiency and accommodate a large number of users. However, in a massive MIMO system, it is difficult to detect signals from a large number of users. Also, the complexity of signal detection becomes high. As simple signal detection techniques, linear detection methods such as zero-forcing (ZF) and minimum mean squared error (MMSE) are known. However, those need the inverse matrix calculation, which results in a large computational complexity for massive MIMO systems. The signal detection technique

based on the belief propagation (BP) algorithm, referred to as BP detection, is one of the promising techniques [21]–[25]. The BP algorithm calculates a marginal probability of unobserved variables by message passing in a factor graph. In BP detection, symbol replicas are generated from the propagated messages at each iteration, and the log-likelihood ratio (LLR) of each symbol is updated as a message by removing the interfering component of the received signals. BP detection can achieve near-optimal detection performance with lower complexity [21]–[25]. Moreover, BP detection does not require the matrix-inversion calculation, which is attractive for massive MIMO systems. However, there are some issues in BP detection. In BP detection errors occur in the propagated messages due to residual interference and noise in the received signals after interference removal. Also, due to multiple loops included in the MIMO channel, a message with error propagates throughout the factor graph, which results in the degradation of the convergence performance and the detection performance. A damping factor is introduced to control the message updates to improve the BP detection performance. The damping factors are used to average two successive messages by expanding the BP iteration to the neural network so that the detection performance and convergence performance can be improved. In conventional works, the damping factors are tuned heuristically in general. However, a heuristic-based selection often results in suboptimal performance.

Neural networks have the ability to learn the fundamental information of the model. Deep unfolding is a technique to unfold the iterations of an inference algorithm into a layer-wise structure like neural networks [14]. Deep unfolding capitalizes the well-known signal processing model and the ability of DL. It can solve problems for which precise modeling is not available. It can also approximate computationally complex operations by a deep neural network (DNN). In deep unfolding, model parameters are de-coupled across layers that can be trained easily and discriminatively using gradient-based methods. There are similarities between the message passing factor graph and DNN as shown in Table 1 [12]. Thus, DNN is employed to improve the convergence performance of BP, which is referred to as DNN-based damped BP (DNN-dBP) [12]. DNN-dBP trains the damping factors by unfolding the BP iteration to the neural network. By using the trained damping factors, it is possible to improve the convergence performance of BP. In [13], we derived the damping factors that are robust to the channel mismatches between training and testing using DNN-dBP. Fig. 2 shows the structure of the proposed DNN-dBP with node selection. In this method, observation nodes to be updated in one iteration are selected so that the spatial correlation becomes low. Thus, the channel correlation among the selected nodes in BP detection is lowered and the convergence performance of BP is improved. Therefore, the dumping factors derived based on this method are robust to the channel mismatches between training and testing.

In the context of image processing, deep image prior (DIP) has been reported as a method to remove noise with-

Table 1 Similarities between BP FG and DNN.

BP FG	DNN
Nodes	Neurons
Transmitted signals	Input data
Received signals	Output data
l -th iteration	l -th hidden layer
Belief messages	Hidden signals
Message updating rules	Mapping function between layers
Correction factor	Parameters

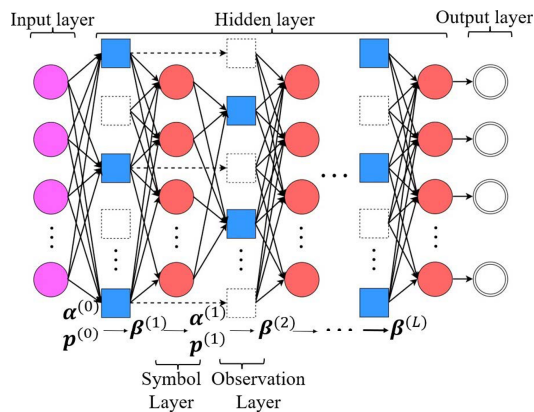


Fig. 2 A structure of DNN-dBP with node selection [13].

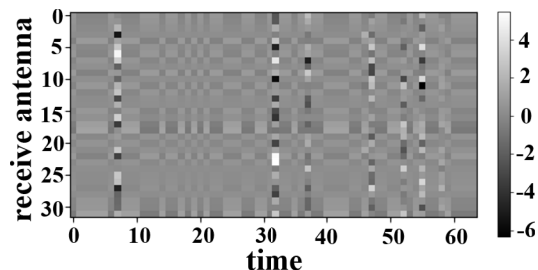


Fig. 3 A heatmap of the received signal.

out the need for teacher data [15]. DIP learns a single input image and optimizes the parameters of the convolutional neural network (CNN) by the gradient descent method to obtain a reconstructed image in general. DIP can be said to exploit the difference in the learning speed of neural networks for images. It has been shown in [15] that DIP learns faster for natural images than for random images such as noise. DIP can use this difference in the learning speed to output a clean image, i.e., an image with reduced noise, by stopping learning before learning noise. In [26], a heatmap of the received signal is generated as shown in Fig. 3 using the “receive antenna index” and “time index” as dimensions. In each heatmap, each receive antenna receives the same transmitted symbols after interference removal, so the correlation is high in the domain of the receive antenna. Suppose the inter-frame channel is assumed to be constant. In that case, each receive antenna receives a symbol pattern of transmitted symbols at each time, resulting in high correlations in the time domain. Based on these correlations, DIP can reduce residual interference and noise. In [26], we introduced a

massive MIMO BP detection using DIP with a DNN-trained scaling factor. In BP detection, we create the heatmap of the received signals after interference removal at each iteration so that it correlates. By applying DIP to the heatmap of the received signals, it is possible to reduce residual interference and noise. After applying DIP, the variance of the interference and noise components changes. To bring the variance closer to its true value, we scale it. Because it is difficult to calculate the value of the variance after applying DIP theoretically, we train the scaling factors offline using DNN-BP. By scaling the variance, it is possible to improve the reliability of the message. Figure 4 shows the BER performance versus SNR in dB in the correlated channel (the correlation factor $\rho = 0.3$) where the modulation scheme is QPSK, 16×16 MIMO, and the number of BP iterations is 7.

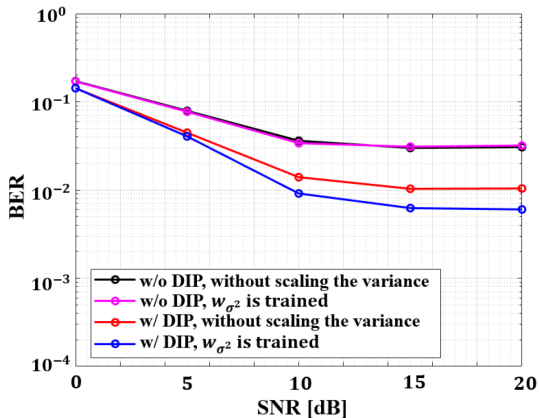


Fig. 4 BER performance vs SNR (dB) per receive antenna in the correlated channel ($\rho = 0.3$) where the modulation scheme is QPSK, 16×16 MIMO, and the number of BP iterations is 7.

It can be seen that the BER performance is improved by applying DIP. It can be also seen that the BER performance of the proposed method with the trained scaling factor is better than that without the trained one.

2.2.2 Pilot Contamination

In massive MIMO, the number of channels that needs to be estimated is large. Since the number of orthogonal pilot signals is limited when we limit the length of those, the same pilot signals need to be reused in neighboring cells. The degradation of the channel estimation performance by reusing the same pilot signals is referred to as pilot contamination. In [27] a covariance-aided channel estimation is proposed, in which the MMSE channel estimation is derived. This scheme can remove the pilot contamination completely when the covariance matrices satisfy a certain non-overlapping condition. However, this assumption is not so practical.

Recently, DL is expected to improve the channel estimation performance in massive MIMO. In [28], DL is integrated into direction-of-arrival (DoA) estimation and channel estimation in massive MIMO systems. In [16] we propose two methods of DL-aided channel estimation to reduce the effects of pilot contamination. One uses a neural network consisting of fully connected layers, while the other uses a CNN. Figure 5 shows the frameworks of the proposed methods where the upper and lower parts show the structure in the neural network-based estimation using the fully connected layers and the CNN-based estimation using the convolutional layers, respectively. Neural networks, particularly CNN, can extract features of spatial information from the contaminated signals. It is shown that the former method

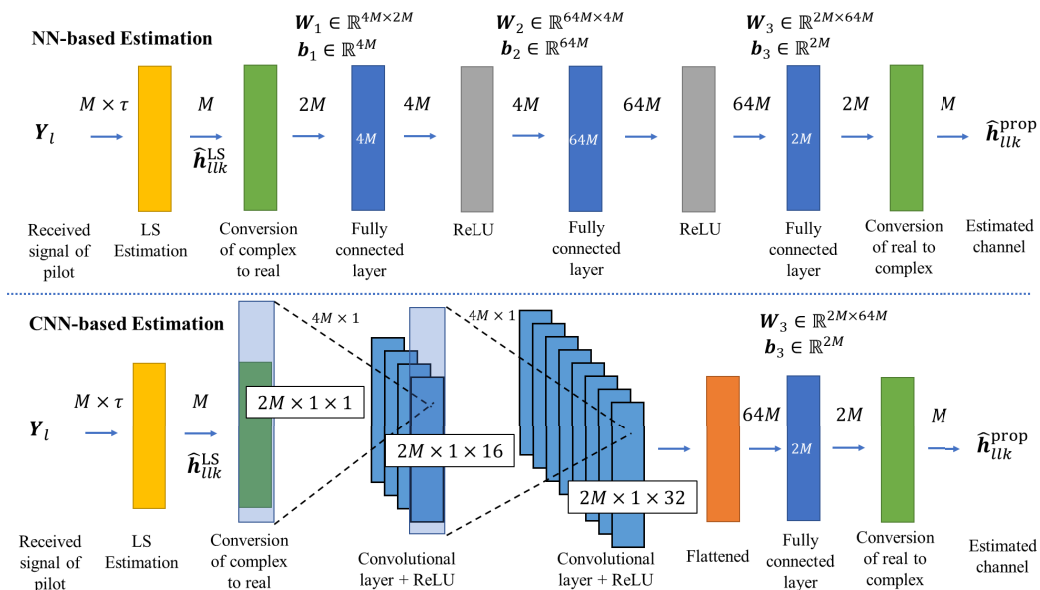


Fig. 5 A framework for the proposed methods [16]. The upper part and the lower part show the structure in the neural network-based estimation using the fully connected layers and the CNN-based estimation using the convolutional layers, respectively.

is better in terms of the training speed, while the latter one can estimate the channel more accurately.

2.3 Deep Transfer Learning

Transfer learning (TL) is a machine learning method where a model trained for a task is used as a starting point for a model on a different related task. TL is a popular technique in DL such as for computer vision and natural language processing where a large amount of computation and time resources are required to train a model from scratch. In TL, a domain and a task are defined. A domain D is defined as a pair $D = \{\mathcal{X}, P(X)\}$, which consists of a feature space \mathcal{X} and a marginal distribution $P(X)$ over the feature space, where $X = \{x_1, \dots, x_n\} \in \mathcal{X}$. A task is defined as a pair $T = \{\mathcal{Y}, f(\cdot)\}$. \mathcal{Y} is the label space, and given $y_i \in \mathcal{Y}$, $f(\cdot)$ is a function that predicts y_i corresponding to x_i . Using the definitions of a domain and a task, TL can be defined as follows [29]:

Transfer learning : Given a source domain D_S , a source task T_S , a target domain D_T , and a target task T_T , the aim of TL is to improve the learning of the target prediction function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S , where $D_S \neq D_T$ or $T_S \neq T_T$.

Deep transfer learning: DTL is a method that combines deep learning with TL. Given that the TL task is defined by $\langle D_S, T_S, D_T, T_T \rangle$, which is a DTL task when the target prediction function $f_T(\cdot)$ for T_T is a non-linear function approximated by DNN.

DTL has been applied to wireless communications as well, such as CSI feedback, beamforming, signal detection, physical layer security, and so on. In [17], DTL is used to generate the CSI feedback deep learning model for each target channel model whereas the Clustered Delay Line (CDL) channel model [30] is used to simulate the wireless environments. Specifically, the DNN is trained as the source model by using a large number of CDL-A samples as source data. The source model is then fine-tuned with a small number of CDL-B, CDL-C, CDL-D, and CDL-E samples, i.e., target data, respectively. Based on this procedure, a target model for each target channel can be obtained with a small number of samples and a short training time. Figure 6 shows the system model of the CSI feedback scheme based on DTL [17]. Figure 7 shows the NMSE performance of the DTL scheme [17] in FDD massive MIMO systems where the target channel is CDL-A. The frequencies of the uplink and downlink channels are set to 2.0 GHz and 2.1 GHz, respectively. The numbers of antennas of UE and BS are 2 and 32, respectively. The number of subcarriers is set to 72 with a spacing of 15 kHz, and the number of OFDM symbols to 14. The estimated CSI of UE and feedback CSI of BS are assumed to be error-free. The compression ratio is set to 1/8. The number of source data samples used to train the source model is set to 50,000, and that of target data samples used for fine-tuning is varied as 200, 500, 1000, 2000, and 4000. The red dotted line with the label ‘‘CDL-A (NLOS)’’ represents the NMSE performance of the source model trained using CDL-A as the

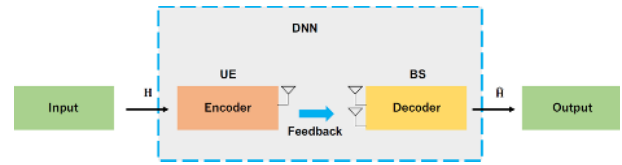


Fig. 6 The system model of the CSI feedback scheme based on DTL [17].

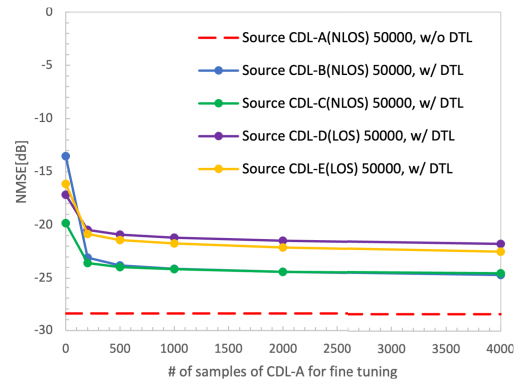


Fig. 7 The NMSE performance versus the number of target data samples where the target channel is CDL-A.

source data. We can see that there is a performance degradation, but the DTL scheme achieves good NMSE performance with the small number of target data samples. We can also see that in the DTL scheme, different source models provide different NMSE performances. In this environment where the target channel is CDL-A (NLOS), the DTL scheme provides the best NMSE performance when the source model is CDL-B (NLOS) and CDL-C (NLOS). As mentioned before, the source model selection is important for the DTL scheme. Some discussion about the source model selection criteria in the DTL scheme can be found in [31].

2.4 mmWave Communications

In wireless communications, there have been continuous and tremendous efforts to increase capacity by expanding spectrum and improving spectral efficiency and spatial reuse. It is very important to utilize new spectrum bands such as mmWave bands and tera-hertz frequency bands. A significant amount of research has been ongoing to improve and realize mmWave systems. However, mmWave systems suffer from severe pathloss. Thus, it is essential to use beamforming with large antenna array gains in mmWave communications. In mmWave communications, the power consumption and cost of RF chains are both high. Hybrid beamforming is a promising technique to balance tradeoffs between cost and performance. Since mmWave communications need to use a large number of antennas, the channel estimation is also the challenging task. Against the challenge, a switched beamforming scheme has been proposed [32] in which the best beams to steer are found within the codebook. Among beam selection schemes, an exhaustive search scheme achieves the best performance but requires a

large overhead particularly when a large number of beams are employed [33]. A hierarchical beam search proposed in [34] can reduce the beam training overhead by two-stage beam training. In the hierarchical beam search scheme, BS and UE, equipped with multiple-tier codebooks, sweep wider beams first and iteratively thin the search space for the best narrow beam. The hierarchical beam search scheme can provide a good trade-off among the performance, the time, and the large overhead. In [35] a beam selection scheme using DL is proposed to reduce the overhead. The DL model estimates the qualities (received power) of all the beams from a few beam measurements. The authors introduce the DL-based image reconstruction approach to the beam selection where the received power matrix is transformed into a power map by assigning the received power to the corresponding color. However, since the beams used for measurements are selected randomly, the performance of the scheme can be largely affected by the beam searching area [35].

In [18], we proposed a DL-based low overhead analog beam selection scheme in which two different-width beams are steered, wide beams for pilot signals and narrow beams for data signals. To change the beam widths without losing beamforming gain, a balance beam is implemented in our proposed scheme, which concentrates a radiation pattern over the target area. Based on the wide-beam measurements, the proposed super-resolution-inspired DL predicts the beam qualities (received powers) of narrow beams where the spatial correlation in the beam qualities is utilized with a CNN to improve the estimation accuracy. Moreover, the proposed scheme predicts beam qualities to reduce the frequency of beam training. The proposed scheme transmits the pilot signal only every other channel coherence time to reduce the training overhead. The current received powers with narrow beams are predicted based on the past pilot signals. Thus, the training time can be reduced by half. To capture spatiotemporal correlations, the proposed model is designed with a convolutional long short-term memory (LSTM) network. Figure 8 shows an idea of the proposed super-resolution-inspired DL scheme. Here, the received power matrix obtained by 4×4 DFT beams is transformed into a power map by assigning the received power to the corresponding color. The low-resolution beam domain image is input to the super-resolution-inspired DL network to output the high-resolution beam domain image corresponding

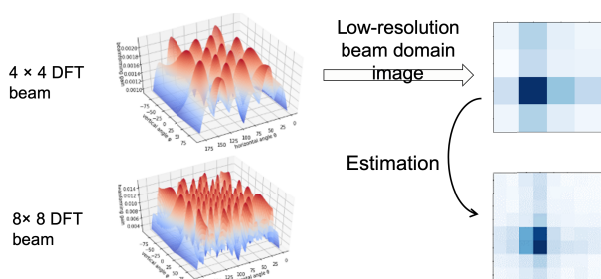


Fig. 8 Estimation of received power of narrow beam from that of wide beam based on super resolution [18].

to the power map obtained by such as 8×8 DFT beams. It is shown in [18] that the proposed beam selection achieves a performance comparable to that of the exhaustive search scheme. Note that the number of beam measurements per coherence time is 8 for the proposed scheme and 64 for the exhaustive search scheme.

3. Conclusions

In this paper, I presented an overview of some ML techniques and applications in 6G wireless communications, mainly focusing on the physical layer. One of the challenges in applying ML to real systems is the dynamic environments. The environments of wireless communications dynamically change. ML makes inferences and predictions using data. Therefore, if the statistical properties of the data change over time, the performance of the system using ML may degrade. To use ML in wireless communications, DTL that I introduced its applications in wireless communications is one of the promising solutions. Another solution is the meta-learning that learns how to learn [17]. Another challenge is that ML, particularly DL-based solutions, usually require a large amount of training data and computational resources. To apply DL-based solutions, we need to carefully consider those requirements.

A common problem with AI is that the parameters obtained as a result of training are difficult to interpret. That is, it is difficult to interpret why the characteristics obtained by AI are the way they are. This is called the interpretability problem. However, to use AI in a real system, it is necessary to be able to understand and explain why such characteristics are obtained. Explainable AI (XAI), which is an ML model whose results and processes leading to them are interpretable by humans, has been actively studied in recent years. A typical technique to realize XAI is LIME [36], which is a local approximation approach to represent AI's decision logic for specific input data in an interpretable form. XAI is also an important technology for 6G.

References

- [1] NTT DOCOMO, "White Paper: 5G Evolution and 6G (Version 4.0)," Jan. 2022. https://www.docomo.ne.jp/english/binary/pdf/corporate/technology/whitepaper_6g/DOCOMO_6G_White_PaperEN_v4.0.pdf
- [2] The Ministry of Internal Affairs and Communications (MIC), "Beyond 5G Promotion Strategy—Roadmap towards 6G—," June 2020. https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/presentation/pdf/Beyond_5G_Promotion_Strategy-Roadmap_towards_6G-.pdf
- [3] National Institute of Information and Communications Technology (NICT), "Beyond 5G/6G White Paper (English Version 2.0)," June 2022. https://beyond5g.nict.go.jp/images/download/NICT_B5G6G_WhitePaperEN_v2_0.pdf
- [4] KDDI Corporation, KDDI Research, Inc., "Beyond 5G/6G White Paper (Version 2.0.1)," Oct. 2021. https://www.kddi-research.jp/sites/default/files/kddi_whitepaper_en/pdf/KDDI_B5G6G_WhitePaperEN_2.0.1.pdf
- [5] NOKIA Bell Labs, "Communications in the 6G Era," Sept. 2020. https://d1p0gxncu0lvz.cloudfront.net/documents/Asset_20200909

- 220306.pdf
- [6] H. Viswanathan and P.E. Mogensen, "Communications in the 6G Era," *IEEE Access*, vol.8, pp.57063–57074, 2020.
 - [7] Ericsson: "6G – Connecting a cyber-physical world," Feb. 2022. <https://www.ericsson.com/4927de/assets/local/reports-papers/white-papers/6g--connecting-a-cyber-physical-world.pdf>
 - [8] 6G Flagship, University of Oulu, "6G Visions on paper," <https://www.6gflagship.com/white-papers/>
 - [9] J. Du, C. Jiang, J. Wang, Y. Ren, and M. Debbah, "Machine learning for 6G wireless networks: Carrying forward enhanced bandwidth, massive access, and ultrareliable/low-latency service," *IEEE Veh. Technol. Mag.*, vol.15, no.4, pp.122–134, Dec. 2020.
 - [10] 3GPP TR 23.791, "Study of Enablers for Network Automation for 5G," June 2019.
 - [11] 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.
 - [12] X. Tan, W. Xu, K. Sun, Y. Xu, Y. Be'ery, X. You, and C. Zhang, "Improving massive MIMO message passing detectors with deep neural network," *IEEE Trans. Veh. Technol.*, vol.69, no.2, pp.1267–1280, Feb. 2020.
 - [13] J. Tachibana and T. Ohtsuki, "Learning and analysis of damping factor in massive MIMO detection using BP algorithm with node selection," *IEEE Access*, vol.8, pp.96859–96866, 2020.
 - [14] J.R. Hershey, J.L. Roux, and F. Wenginger, "Deep unfolding: Model-based inspiration of novel deep architectures," [arXiv:1409.2574v4](https://arxiv.org/abs/1409.2574v4), 2014.
 - [15] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp.9446–9454, June 2018.
 - [16] H. Hirose, T. Ohtsuki, and G. Gui, "Deep learning-based channel estimation for massive MIMO systems with pilot contamination," *IEEE Open J. Veh. Technol.*, vol.2, pp.67–77, 2021.
 - [17] J. Zeng, J. Sun, G. Gui, B. Adebisi, T. Ohtsuki, H. Gacanin, and H. Sari, "Downlink CSI feedback algorithm with deep transfer learning for FDD massive MIMO system," *IEEE Trans. Cogn. Commun. Netw.*, vol.7, no.4, pp.1253–1265, Dec. 2021.
 - [18] H. Echigo, Y. Cao, M. Bouazizi, and T. Ohtsuki, "A deep learning-based low overhead beam selection in mmWave communications," *IEEE Trans. Veh. Technol.*, vol.70, no.1, pp.682–691, Jan. 2021.
 - [19] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, USA, 2016.
 - [20] F.A. Aoudia and J. Hoydis, "Model-free training of end-to-end communication systems," *IEEE J. Sel. Areas Commun.*, vol.37, no.11, pp.2503–2516, Nov. 2019.
 - [21] W. Fukuda, T. Abiko, T. Nishimura, T. Ohgane, Y. Ogawa, Y. Ohwatari, and Y. Kishiyama, "Low-complexity detection based on belief propagation in a massive MIMO system," *Proc. VTC 2013-Spring, Dresden, Germany*, pp.1–5, June 2013.
 - [22] J. Yang, C. Zhang, X. Liang, S. Xu, and X. You, "Improved symbol based belief propagation detection for large-scale MIMO," *Proc. IEEE Workshop Signal Process. Syst. (SiPS)*, pp.1–6, Oct. 2015.
 - [23] F. Long, T. Lv, R. Cao, and H. Gao, "Single edge based belief propagation algorithms for MIMO detection," *Proc. IEEE Sarnoff Symposium, Princeton, USA*, pp.1–5, May 2011.
 - [24] Y. Gao, H. Niu, and T. Kaiser, "Massive MIMO detection based on belief propagation in spatially correlated channels," *Proc. IEEE 11th Int. ITG Conf. Syst. Commun. Coding*, pp.1–6, Feb. 2017.
 - [25] T. Takahashi, S. Ibi, T. Ohgane, and S. Sampei, "On normalized belief of Gaussian BP in correlated large MIMO channels," *2016 International Symposium on Information Theory and Its Applications (ISITA)*, pp.458–462, Nov. 2016.
 - [26] J. Tachibana and T. Ohtsuki, "Massive MIMO detection with BP algorithm using DIP for removing residual interference and noise and DNN for learning scaling factor," *IEICE Technical Report, RCS2020-253*, March 2021 (in Japanese).
 - [27] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol.31, no.2, pp.264–273, Feb. 2013.
 - [28] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol.67, no.9, pp.8549–8560, Sept. 2018.
 - [29] S.J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol.22, no.10, pp.1345–1359, Oct. 2010.
 - [30] 3GPP TR 38.901: "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project; Technical Specification Group Radio Access Network, 2017.
 - [31] M. Inoue and T. Ohtsuki, "Evaluation of source data selection for DTL based CSI feedback method in FDD massive MIMO systems," *IEICE Technical Report, RCS2022-62*, June 2022 (in Japanese).
 - [32] P. Wang, Y. Li, L. Song, and B. Vucetic, "Multi-gigabit millimeter wave wireless communications for 5G: From fixed access to cellular networks," *IEEE Commun. Mag.*, vol.53, no.1, pp.168–178, Jan. 2015.
 - [33] L. Wei, Q. Li, and G. Wu, "Exhaustive, iterative and hybrid initial access techniques in mmwave communications," *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp.1–6, 2017.
 - [34] M. Giordani, M. Mezzavilla, C.N. Barati, S. Rangan, and M. Zorzi, "Comparative analysis of initial access techniques in 5G mmWave cellular networks," *2016 Annual Conference on Information Science and Systems (CISS)*, pp.268–273, 2016.
 - [35] C.-H. Lin, W.-C. Kao, S.-Q. Zhan, and T.-S. Lee, "Bsnet: A deep learning-based beam selection method for mmWave communications," *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pp.1–6, 2019.
 - [36] M.T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you?" Explaining the predictions of any classifier," *KDD'16: Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1135–1144, Aug. 2016.



Tomoaki Ohtsuki received the B.E., M.E., and Ph.D. degrees in Electrical Engineering from Keio University, Yokohama, Japan in 1990, 1992, and 1994, respectively. From 1994 to 1995 he was a Post Doctoral Fellow and a Visiting Researcher in Electrical Engineering at Keio University. From 1993 to 1995 he was a Special Researcher of Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists. From 1995 to 2005 he was with the Science University of Tokyo. In 2005 he joined

Keio University. He is now a Professor at Keio University. From 1998 to 1999 he was with the department of electrical engineering and computer sciences, University of California, Berkeley. He is engaged in research on wireless communications, optical communications, signal processing, and information theory. Dr. Ohtsuki is a recipient of the 1997 Inoue Research Award for Young Scientist, the 1997 Hiroshi Ando Memorial Young Engineering Award, Ericsson Young Scientist Award 2000, 2002 Funai Information and Science Award for Young Scientist, IEEE the 1st Asia-Pacific Young Researcher Award 2001, the 5th International Communication Foundation (ICF) Research Award, 2011 IEEE SPCE Outstanding Service Award, the 27th TELECOM System Technology Award, ETRI Journal's 2012 Best Reviewer Award, 9th International Conference on Communications and Networking in China 2014 (CHINACOM '14) Best Paper Award, 2020 Yagami Award, and The 26th Asia-Pacific Conference on Communications (APCC2021) Best Paper Award. He has published more than 240 journal papers and 470 international conference papers. He served as a Chair of IEEE Communications Society, Signal Processing for Communications and Electronics Technical Committee. He served as a technical editor of the IEEE Wireless Communications Magazine and an editor of Elsevier Physical Communications. He is now serving as an Area Editor of the IEEE Transactions on Vehicular Technology and an editor of the IEEE Communications Surveys and Tutorials. He is also serving as the IEEE Communications Society, Asia Pacific Board Director. He has served as general-co chair, symposium co-chair, and TPC co-chair of many conferences, including IEEE GLOBECOM 2008, SPC, IEEE ICC 2011, CTS, IEEE GLOBECOM 2012, SPC, IEEE ICC 2020, SPC, IEEE APWCS, IEEE SPAWC, and IEEE VTC. He gave tutorials and keynote speeches at many international conferences including IEEE VTC, IEEE PIMRC, IEEE WCNC, and so on. He was Vice President and President of the Communications Society of the IEICE. He is a senior member and a distinguished lecturer of the IEEE, a fellow of the IEICE, and a member of the Engineering Academy of Japan.