Multi-User Hybrid Beamforming System Based on Deep Neural Network in Millimeter-Wave Communication

LAKJU SUNG and DONG-HO CHO, (Senior Member, IEEE)
Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, South Korea
Corresponding author: Dong-Ho Cho (dhcho@kaist.ac.kr)

This work was supported by the BK21 PLUS and Institute for Information and Communications Technology Promotion (IITP) Grant funded by the Korean Government (MSIT), Development of Simultaneous Wireless Information and Power Transfer System Using Same Resource Based on Multiple Antennas under Grant 2017-0-00708.

ABSTRACT
Millimeter-wave (mmWave) communication with a large bandwidth can result in a significantly improved data rate in wireless communications. To overcome high path-loss in the mmWave frequency band, beamforming technology is necessary. Especially, there has been widespread interest in development of hybrid beamforming (HB) technologies, in view of reducing cost and power consumption in massive multiple input multiple output (MIMO) systems. Some of existing researches on HB algorithms assumed perfect channel state information (CSI) and the others used beam training process in case of assuming imperfect CSI. When beam training process is used, enough beam training has to be conducted to achieve sufficient system performance in massive MIMO systems, which results in significant training overhead. Thus, it is necessary to reduce beam training complexity. Compared to state-of-the-art technology, we propose a multi-user HB system using codebooks based on a deep neural network (DNN) in this paper. In our proposed scheme, beam codewords for the base station (BS) and all users can be inferred using limited beam training in cases when the channel state information (CSI) is unknown. In order to apply the proposed scheme to situations where the CSI is unknown, reference radio frequency (RF) beamformers were introduced. Also, the proposed DNN structure is designed considering introduced reference RF beamformers. By using the proposed DNN with reference RF beamformers, the proposed system can inferred optimal beam codewords with limited beam training. Results obtained from simulations indicate that the proposed scheme can achieve almost the same performance as a conventional scheme with less beam training complexity. We also show that the performances achieved by the proposed scheme are gradually increased as training epoch is increased, eventually converging to a steady-state value.

INDEX TERMS
Hybrid beamforming, deep neural network, limited feedback beamforming, mmWave communication.

I. INTRODUCTION
Millimeter-wave (mmWave) communication is a technology that shows considerable promise in that it allows a significantly increased data rate using a large bandwidth in wireless communications [1]. Since path-loss is high in mmWave frequency band, it is important to solve this problem. Thus, a large antenna array based beamforming technology is necessary to overcome the high path loss in mmWave frequency band [2]. In cellular communications, fully-digital beamforming systems have been used to increase data rate [3], [4]. To use fully-digital beamforming technology, radio frequency (RF) chains have to be connected to all antennas. In massive multiple input multiple output (MIMO) systems, a large number of RF chains are needed due to large antenna array, which causes significant cost and power consumption. Thus, a hybrid analog and digital beamforming system is commonly used in mmWave communication to overcome problems due to high cost and power consumption, which are caused by the use of many RF chains in a large antenna array. Also, using small number of RF chains and low cost phase shifters, hybrid beamforming (HB) system can
achieve almost the same system performance as fully-digital beamforming system. Therefore, research on HB systems in mmWave communication based on massive MIMO is very important.

A great number of HB algorithms have been proposed for mmWave communication [5]–[8]. An HB algorithm based on orthogonal matching pursuit (OMP) was proposed in a single-user MIMO environment by exploiting sparsity of mmWave channel [5]. Also, HB algorithms were designed in a multi-user multiple input single output (MISO) system [6]. The HB algorithms proposed in [5] and [6] assumed that perfect channel state information (CSI) could be obtained but it is almost impossible to achieve perfect CSI in a massive MIMO system. Thus, a codebook-based HB system using imperfect CSI was developed [7]. In [7], analog beamformers of BS and users were found by using codebook based beam training. By using codebook based beam training, effective channel information can be acquired. Then, baseband beamformer of BS are designed based on effective channel information to minimize inter-user interference. Furthermore, there have been some attempts to design HB algorithms based on codebook using a dual polarized array antenna. By using dual polarized array antenna, data rate can be increased in case of imperfect CSI [8]. Due to exploiting imperfect CSI in the HB algorithms proposed in [7] and [8], beam training process was introduced. In order to achieve better performance, the number of beam training has to be increased according to the number of antennas at BS and users, which causes significant training overhead in massive MIMO systems. Therefore, it is necessary to reduce beam training complexity in HB systems.

Aside from encouraging developments of HB systems, the use of deep neural networks (DNNs) has resulted in significant performance improvements in several areas, especially in image processing [9]. Recently, DNNs have been also applied to communication systems actively to increase system performance. Thus, a number of studies on communication systems using DNN has been increasing steadily [10]–[14]. Sparse code multiple access (SCMA) assisted by deep learning was proposed in [10]. Then, autoencoder assisted DNN structure was used to generate codewords and determine decoding strategy. In addition, power control scheme using convolutional neural network (CNN) was proposed [11]. By using CNN, spectral and energy efficiency were increased with much lower signaling overhead. There were also researches on beamforming systems using learning approach [12]–[14]. A coordinated beamforming system based on DNN for users with high mobility was described in multi-point to point environment [12]. In [12], several access points (APs) decide their analog beamformers based on received signal at one receive antenna. An analog beam selection method was proposed in [13] based on not DNN structure but a support vector machine (SVM). In addition, hybrid precoder was designed based on autoencoder structure in point to point environment [14]. In [13], however, perfect CSI has to be input value of the proposed learning model. Also, angles of arrival and departure are input values of the proposed scheme in [14]. Thus, another pre-processes to know CSI are necessary to train learning model. Therefore, HB algorithms proposed in [13] and [14] are hard to be used directly in practical system.

In this paper, we propose a DNN-based HB system for a multi-user environment in mmWave communication, which can achieve a sub-optimal rate performance with low beam training complexity when the CSI is unknown. Then, the main contributions of this paper can be summarized as follows.

- We propose a DNN-based HB system considering a multi-user environment. Unlike conventional algorithms, the proposed DNN based HB system can work with imperfect CSI. Because of using imperfect CSI in the proposed DNN based HB system, we introduce reference RF beamformer which receives pilot signals transmitted from users into beam training procedure. By using reference RF beamformer, DNN structure is changed compared to conventional schemes.
- By using the proposed DNN for HB trained by supervised learning, beam codewords of the BS and users can be inferred simultaneously based only on signals received at the reference RF beamformer using limited beam training. This allows HB system to be designed with significantly lower beam training complexity than conventional scheme.
- We use simulations to evaluate the performance of the proposed system. After, having shown that the proposed system can be used to achieve almost the same performance as a conventional HB system using a low beam training complexity, we also show that the performances of the proposed scheme are increased gradually when the training epoch is increased, converging eventually to a steady-state value.

The rest of this paper is organized as follows. In Section II, system model and channel model are described. The proposed multi-user DNN based hybrid beamforming system is presented in section III. Here, beam training procedure and layer architecture of the proposed scheme are introduced. Simulation results of the proposed scheme are given in section IV. Finally, conclusion of this paper is made in section V.

**Notations:** $\mathbb{C}$ and $\mathbb{R}$ denote the field of complex numbers and real, respectively. Bold capital letter and small letter mean matrix and vector, respectively. Also, $(\cdot)^s$, $(\cdot)^T$ and $(\cdot)^H$ are complex conjugate, transpose and Hermitian transpose, respectively. Finally, $I$ is identity matrix and $N(m, C)$ represents a complex Gaussian random vector with mean $m$ and covariance $C$.

## II. SYSTEM MODEL

In the proposed DNN based HB system, it is assumed that the BS has $N_t$ antennas and each user has $N_r$ antennas. A hybrid beamformer at the BS is composed of a baseband beamformer, $F_{BB} = [f_{BB,1} \cdots f_{BB,K}] \in \mathbb{C}^{N_{RF} \times N_t}$, and a RF beamformer, $F_{RF} = [f_{RF,1} \cdots f_{RF,K}] \in \mathbb{C}^{N_t \times N_{RF}}$, where $N_{RF}$ is the number of RF chains at BS, $N_t$ is the number of data

VOLUME 8, 2020 91617
The antenna radiation pattern vectors of the $p$th antenna at the BS and the $q$th antenna at user $k$, respectively, and $\Phi^j$ is uniformly distributed over $[0, 2\pi]$. We consider a vertical half-wavelength dipole antenna radiation pattern [17].

\[ \alpha_k^j = \mathbf{R}_p^{(\theta_{r,k}^j, \phi_{r,k}^j)} \mathbf{R}_q^{(\theta_{l,k}^j, \phi_{l,k}^j)} e^{j\Phi^j}. \]  

(2)

In the present context, $\mathbf{R}_p^{(\theta_{r,k}^j, \phi_{r,k}^j)}$ and $\mathbf{R}_q^{(\theta_{l,k}^j, \phi_{l,k}^j)}$ are the antenna radiation pattern vectors of the $p$th antenna at the BS and the $q$th antenna at user $k$, respectively, and $\Phi^j$ is uniformly distributed over $[0, 2\pi]$. We consider a vertical half-wavelength dipole antenna radiation pattern [17].

The received signal at the $k^{th}$ user before receive beamforming is denoted $\hat{y}_k$, and can be represented as

\[ \hat{y}_k = \mathbf{H}_k \mathbf{F}_{RF} \mathbf{f}_{BB,k} x_k + \sum_{n=1,n\neq k}^{K} \mathbf{H}_k \mathbf{F}_{RF} \mathbf{f}_{BB,n} x_n + \mathbf{n}_k, \]  

(3)

where $x_k$ is the transmitted signal for the $k^{th}$ user and $\mathbf{n}_k \in \mathbb{C}^{N_t \times 1}$ is an additive white Gaussian noise (AWGN) vector with $\mathbf{n}_k \sim N(0, \sigma^2 \mathbf{I})$. The signal received at the $k^{th}$ user after receive beamforming is denoted as $y_k$, and is expressed as $y_k = \mathbf{w}_{RF,k}^\ast \hat{y}_k$.

\section*{B. CODEBOOK BASED BEAMFORMER MODEL AND PROBLEM FORMULATION}

To overcome the high path-loss in a mmWave system, it is necessary to use several beams to acquire CSI. However, it is very hard to obtain perfect CSI in a massive MIMO system due to the large amount of channel feedback information. A codebook based beam training technique is therefore used to obtain sub-optimal beam sets at the BS and the users. We use a discrete fourier transform (DFT) codebook and there are two reasons why DFT codebook is employed. First reason is that mmWave channel has sparsity, so DFT codebook is well suitable for mmWave channel environments [7]. Second reason is that phase shifters are only used to design RF beamformer considering practical environment. Then, DFT beam can be implemented simply by using phase shifters. Then, the DFT codebook at the BS can be defined as

\[ \mathbf{C}' = \begin{bmatrix} \mathbf{c}'_{1,1}, \cdots, \mathbf{c}'_{1,j}, \cdots, \mathbf{c}'_{ele,N_{ele}^{R}}, \mathbf{c}'_{azi,N_{azi}^{R}} \end{bmatrix}, \]  

(4)

where $\mathbf{c}'_{i,j} \in \mathbb{C}^{N_t \times 1}$ is the DFT weight vector for the $i^{th}$ elevation and $j^{th}$ azimuth direction at the BS, and $N_{ele}^{R}$ and $N_{azi}^{R}$ are the number of elevation and azimuth directions of the DFT beams at the BS, respectively. The number of codewords in $\mathbf{C}'$ can be expressed as $|\mathbf{C}'| = N_{ele}^{R} N_{azi}^{R}$. Like the BS codebook, the DFT codebook at the user can be written as $\mathbf{C}'$ with $|\mathbf{C}'| = N_{ele}^{U} N_{azi}^{U}$, where $N_{ele}^{U}$ and $N_{azi}^{U}$ are the number
of elevation and azimuth directions of the DFT beams at the user, respectively. The BB beamformer is then assumed to be ideal zero-forcing (ZF) beamformer to eliminate inter-user interference.

In the proposed HB system, we need to find \( F_{BB} \), \( F_{RF} \) and \( w_{RF} \) to achieve the maximum sum rate. The rate of the \( k \)th user, \( R_k \), can be written as

\[
R_k = \log_2(1 + \frac{P_k}{K} \left| w_{RF,k} H_k F_{BB,k} \right|^2 + \sigma^2),
\]

where \( P_t \) is the total transmit power and \( K \) is the number of users. Then, the sum rate of the proposed system, \( R \), can be represented as \( R = \sum_{k=1}^{N_k} R_k \). Here, finding optimal solutions of \( w_{RF,k} \), \( F_{RF} \) and \( F_{BB,k} \) maximizing sum rate is very hard when CSI is unknown. Therefore, we find sub-optimal solutions of \( w_{RF,k} \) and \( F_{RF} \) maximizing \( \left| w_{RF,k} H_k F_{BB,k} \right|^2 \). After then, \( F_{BB} \) is solved using ZF beamformer corresponding to \( w_{RF,k} \) and \( F_{RF} \).

III. PROPOSED DNN BASED HB SYSTEM

We now consider the details of our DNN based HB system using supervised learning. Using the proposed system, codewords of the BS and all users are found based on codebooks with imperfect CSI. We first explain the DNN-based HB system and analyze the beam training complexity of the proposed system.

A. OPERATION OF PROPOSED DNN BASED HB SYSTEM

Due to high path-loss in mmWave communication, BS could not receive pilot signals transmitted by one antenna of MS. Also, without the use of devices such as switches and attenuators, the BS cannot know all the signals coming into each antenna due to the hybrid beamforming architecture. Thus, channel matrix elements between BS and MS antennas which will be the input matrix of the DNN cannot be obtained accurately in practical system. In the proposed system, therefore, BS and MSs transmit and receive pilot signals using their own beam codewords, respectively.

To predict the analog beamformers of the BS and all users, beam training in the proposed system is proceeded as follows. The \( k \)th user transmits signals sequentially using all the codewords contained in the codebooks, and the same signals using all the codewords in the codebooks are transmitted \( N_p \) times. Then, the BS receives the signals using \( N_p \) different specific RF beamformers composed of codewords contained in the BS codebook. These \( N_p \) specific RF beamformers are called reference RF beamformers in this paper. The \( p \)th reference RF beamformer receiving the transmitted signals from the \( k \)th user, \( F_{ref}^p \), \( \in \mathbb{C}^{N_{RF} \times 1} \), can be denoted as

\[
F_{ref}^p = [e_{ref,1}^p, \ldots, e_{ref,N_{RF}}^p] \in \mathbb{C}^r,
\]

where \( F_{ref}^q \cap F_{ref}^q = \emptyset \) in the case when \( p \neq q \) and \( 1 \leq p, q \leq N_p \). It is assumed that the channels between the BS and all users are reciprocal for simplicity [18]. Here, all codewords contained in the one reference RF beamformer are DFT beams having different azimuth and elevation look direction each other to achieve independent AoA information.

The reference RF beamformers configuration in BS codebook is shown in Fig. 2 when \( N_{azi} = N_{ele} = 16, N_{RF} = 8 \), and \( N_p = 4 \). Fig. 2 shows four reference RF beamformers used in this paper. When \( N_p = 2 \), \( F_{ref}^1 \) and \( F_{ref}^2 \) are only used. Also, \( F_{ref}^3 \) is only used in case of \( N_p = 1 \). The configuration of the reference RF beamformers, however, can be changed depending on the system environments.

FIGURE 2. Reference RF beamformers configuration in BS codebook.

The received signal vector from the \( k \)th user when using a transmit beamformer of \( c_{i,j}^r \) and a receive beamformer of \( F_{ref}^p \), \( r_{k,c_{i,j}^r}^p \in \mathbb{C}^{N_{RF} \times 1} \), can be represented as

\[
r_{k,c_{i,j}^r}^p = F_{ref}^p H_k c_{i,j}^r s_k + n_k.
\]

The input vectors for the proposed DNN is then \( r_k = \left[ (r_1^T)^T, \ldots, (r_{N_p}^T)^T \right]^T \), where \( r_k^p \) is defined as

\[
r_k^p = \left[ (r_{k,c_{i,1}^r}^p)^T, \ldots, (r_{k,c_{i,N_{RF} \times azi}}^p)^T \right]^T.
\]

The proposed DNN-based HB system predicts the beam codewords of the BS and all users using only \( r_k \) which is the received signal based on reference RF beamformers at the BS. Thus, only limited beam training is used to obtain \( r_k \).

B. LAYER ARCHITECTURE AND TRAINING OF PROPOSED DNN FOR HB

To predict the analog beamformers of the BS and all users, the proposed DNN is trained by supervised learning, which requires labeled data in this paper. The labeled data is called as training data. To obtain training data, conventional codebook based HB algorithm is used [7]. In [7], sub-optimal beam codewords for the \( k \)th user, \( \tilde{F}_{RF,k}^p \) and \( \tilde{w}_{RF,k}^p \), maximizing \( \left| \tilde{F}_{RF,k}^p H_k^H \tilde{w}_{RF,k}^p \right|^2 \) were found. Then, \( r_k \) is mapped to the sub-optimal beam codeword set.
composed of $\tilde{f}_{RF,k}$ and $\tilde{w}_{RF,k}$. The solved set of $\tilde{f}_{RF,k}$ and $\tilde{w}_{RF,k}$ is indexed to a single integer $0 \leq e_k \leq |C'| |C'| - 1$, and one-hot encoded $\tilde{z}^e$ is written as $\tilde{z}^e_k$. Here, $r_k$ is obtained during the process of finding $\tilde{f}_{RF,k}$ and $\tilde{w}_{RF,k}$ based on conventional codebook based HB algorithm. It means that additional beam training is not needed to obtain $r_k$. In view of practical environment, the BS can obtain training data and train the proposed DNN while the BS supports users based on conventional codebook based HB algorithm. Test data is also generated in the same way as for training data. The proposed DNN based HB system, however, predicts beam codewords of the BS and all users, so the test data was not included in the training data. Therefore, the training data and test data are achieved from different channel matrix, so training data and test data have to be independent.

In the proposed DNN based HB system, the analog beamformer for the $k$th user, $\tilde{f}_{RF,k}$ and $\tilde{w}_{RF,k}$, are inferred based on only $r_k$. The proposed DNN layer architecture for HB is depicted in Fig. 3, in which the proposed DNN is composed of $L$ hidden layers. All hidden layers except the final one are composed of three sub-layers, namely the fully connected (FC) layer, the batch normalization layer, and the rectifier linear unit (ReLU) layer. The final hidden layer is composed of an FC layer and a softmax layer.

The input vector of the proposed DNN, $x_{in}$ is composed of $Re\{r_k\}$ and $Im\{r_k\}$, which are vectors composed of real and imaginary components of the elements in $r_k$, respectively. All the elements in $x_{in}$ are normalized by the maximum element in $x_{in}$ using input normalization. The output vector of the input normalization is $x_1$. When the input vector of the FC layer in the $m$th hidden layer is $x_m$, the output vector of the FC layer in the $m$th hidden layer, $y_m$, is written as $y_m = W_m x_m + b_m$, where $W_m$ and $b_m$ are the weight and bias of the FC layer in the $m$th hidden layer, respectively. Here, $W_m$ and $b_m$ are parameters learned through training of the proposed DNN. Then, $y_m$ is fed into the batch normalization layer, which can accelerate the training of the DNN and prevent overfitting [19]. The output vector of the batch normalization layer in the $m$th hidden layer, $y_m^{BN}$ is defined as $y_m^{BN} = \gamma y_m - \text{Mean}(y_m) + \beta$ where $\epsilon = 10^{-6}$ is a constant to avoid divergence, and $\gamma$ and $\beta$ are scaling and shift factors, respectively. Here, $\gamma$ and $\beta$ can also be learned via the training of the DNN. Finally, $y_m^{BN}$ is fed into the ReLU layer, which is the activation function [20]. To prevent gradient vanishing, ReLU is used in this paper. Using the activation function in ReLU, the output of the $m$th hidden layer becomes $\max(y_m^{BN}, 0)$. In the final hidden layer, the output vector of the FC layer, $y_L$, is fed into the softmax layer, which is the classifier. When the output vector of the softmax layer is $y_{pr}$, all the elements in $y_{pr}$ are normalized such that $y_{pr,e} = \frac{\exp(y_{pr,e})}{\sum \exp(y_{pr,e})}$, where $y_{pr,e}$ and $y_{L,e}$ are the $e$th elements of $y_{pr}$ and $y_L$, respectively.

Having obtained the labeled training data set, we now need to train the proposed DNN for HB, in order to minimize the loss function. Here, a cross-entropy function, $L(\Omega)$, is considered as the loss function and can be expressed as

$$L(\Omega) = - \sum_e \frac{c^e}{c^e} \log(y_{pr,e}), \quad (9)$$

where $\Omega$ is the set of all the parameters in the proposed DNN, and $c^e$ is the $e$th element of $z^e$. In order to minimize the loss function, we use an adaptive moment estimation, in the form of an off-the-shelf stochastic gradient descent algorithm [21].

When training of the proposed DNN is over, $y_{pr}$ can be obtained via the proposed DNN based on $r_k$ obtained at the BS. The proposed DNN finds the index of the element with maximum value among all the elements of $y_{pr}$. Using this index, the analog beamformers of the BS and the $k$th user can be identified directly.

C. BEAM TRAINING COMPLEXITY OF DNN-BASED HB SYSTEM

It is assumed that the proposed DNN for HB is already trained in a practical service environment [12]–[14]. Also, the ZF beamformer is used to find $F_{BB}$, and we therefore focus on the complexity of finding the analog beamformers of the BS and users. Then, we consider not the training complexity of DNN but the beam training complexity. To utilize the proposed
DNN for beam training, DNN has to be trained once. After the proposed DNN for beam training has been trained once, there is no need to train the proposed DNN again later. In practical environments, the already trained proposed DNN have to be used. Therefore, we have focused on the beam training complexities of the proposed DNN based HB system.

In a conventional HB, the size of the BS codebook is related to the complexity of the beam training process. In conventional HB, the required beam training complexity can be written as $O(K \frac{N_{RF}}{c})$. In the proposed system, however, all the codewords in the BS codebook are not used but only the reference RF beamformers are required for beam training. Therefore, the required beam training complexity for finding analog beamformers is $O(KN_p |C'|)$. Also, we compare two conventional beam training methods which are sequential downlink-downlink (SDD) method proposed in [22] and beam training scheme proposed in IEEE 802.11 ad [23].

In Fig. 4, beam training complexities according to size of BS codebook are shown in case of $K = N_{RF} = 8$. As the size of BS codebook is increased to improve system performance, the beam training complexities of the all conventional schemes is increased [7], [22]–[24]. The beam training complexities of the proposed scheme, however, are constant regardless of the size of BS codebook. Thus, the proposed method has a significant advantage in massive MIMO systems in view of complexity reduction for beam training.

![FIGURE 4. Beam training complexities according to size of BS codebook.](image)

IV. SIMULATION RESULTS AND DISCUSSIONS

In this section, the performances of the proposed DNN-based HB were compared with those of a conventional algorithms [7], [22], [23]. The performances of the proposed DNN-based HB were also compared with those of fully-digital beamforming with perfect CSI, using the unconstrained block diagonalization algorithm described in [25] and random beam selection choosing the beam codewords of the BS and all users randomly. For simplicity, it was assumed that $L_k = 2$ and a uniform planar array of $16 \times 16$ and $4 \times 4$ are used as an antenna array of the BS and each user, respectively. It was also assumed that $N_{t,d} = 16$, $N_{t,e} = 4$, $N_{r,d} = 4$, and $N_{r,e} = 4$. All parameters for training of DNN were determined experimentally based on fine-tuning. The number of total layer, $L$, was set to 6, in which each FC layer has 4,096 hidden nodes, and the learning rate of the adaptive moment estimation was set to 0.001. In the training of the DNN for HB, a total of 8,000 batches were used, and the size of each batch was set to 1,000. Also, 1,600,000 data were used as test data.

In Fig. 5, the sum rates of four schemes were shown according to their signal-to-noise ratio (SNR) when $K = 8$. The performances of the DNN-based HB were increased with $N_p$. It means that the performances of the DNN based HB are increased when the the amount of feedback information in the DNN for HB is increased. When $N_p$ was 4, the performance of the DNN-based HB was almost the same as that of the conventional HB, but the beam training complexity was much lower. We also noted that the DNN based HB outperforms random beam selection in the case when $N_p = 1$, meaning that beam codewords inferred by the proposed scheme are better than beam codewords which are selected randomly, even if $N_p = 1$. In addition, the proposed DNN based HB had better performance than 802.11 ad method even if $N_p = 2$. Also, the proposed DNN based HB outperformed conventional SDD method even if $N_p = 1$. Then, as shown in Fig. 4, beam training complexity of conventional SDD method was almost the same as that of DNN based HB in case of $N_p = 2$. Also, beam training complexity of 802.11 ad was higher than that of SDD method and the proposed DNN based HB. Therefore, the proposed DNN based HB achieved better performance compared to SDD method and 802.11 ad with lower beam training complexity.

When $N_p = 4$, sum rates according to the number of RF chain were shown in Fig. 6. As the number of RF chain is increased, the sum rates of the conventional and proposed schemes are also increased. In addition, the proposed scheme achieved the almost same performance as the conventional scheme regardless of the number of RF chain.

![FIGURE 5. Sum rates according to SNR.](image)
It was also investigated that the performances of the proposed DNN based HB compared to conventional HB for different number of BS antennas and user antennas. Then, it was assumed that size of codebooks and number of antennas are the same. In Fig. 7 and 8, sum rate according to number of BS and user antennas are shown, respectively. For simplicity, the number of user antennas was fixed to 16 in Fig. 7 and the number of BS antennas was fixed to 64 in Fig. 8. Even if the number of BS and user antennas are increased, the performance of the proposed DNN based HB achieved almost the same performance as that of conventional HB when \( N_p = 4 \). Thus, it was investigated that the proposed DNN based HB worked well in massive MIMO system.

In case of 30 dB SNR, the performance variations of the DNN based HB were shown in Fig. 9, in which the training epoch is the number of iterations for training all the data samples. As the training epoch was increased, the performances of the DNN based HB were also increased. We noted that the sum rate of the DNN based HB for all \( N_p \) converges gradually to a steady-state value according to the increase of the training epoch. Regardless of SNR, the performances of the proposed DNN converge to the steady-state values.

In Table 1, the running times of conventional HB and DNN based HB were described. For running conventional HB and DNN based HB, NVIDIA RTX 2080 GPU and 1.12.0 tensorflow were used in linux environments. Then, it was assumed to find beam codewords for one user to obtain running time based on completely trained DNN for HB. The running time of the proposed DNN based HB was about 64.7% lower than that of conventional HB for \( N_p = 4 \). Also, the running time of the proposed DNN based HB was decreased as \( N_p \) was decreased because of the number of nodes being reduced. It can be possible to further reduce running time of the

| Scheme               | Running time |
|----------------------|--------------|
| Conventional HB      | 5.6 ms       |
| Proposed DNN based HB (\( N_p = 4 \)) | 3.4 ms |
| Proposed DNN based HB (\( N_p = 2 \)) | 2.5 ms |
| Proposed DNN based HB (\( N_p = 1 \)) | 1.85 ms |
proposed DNN based HB depending on the hardware and software capability for the DNN. Thus, the proposed DNN based HB has strength in massive MIMO system in view of running time.

V. CONCLUSION
In this paper, we have proposed a DNN based HB systems for multi-user environments. By using only the received signals based on reference RF beamformers, the sub-optimal beam codewords of the BS and users have been inferred simultaneously reducing beam training complexity. Using a deep learning technique, we showed via simulation that the proposed scheme can achieve almost the same performance as a conventional scheme with a much lower beam training complexity. The performance of the proposed scheme was increased as the number of RF chains was increased. We also showed that the performances of the proposed scheme converge gradually to a steady-state value when the training epoch is increased.

REFERENCES
[1] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!,” IEEE Access, vol. 1, pp. 335–349, May 2013.
[2] S. Han, C.-L. I, Z. Xu, and C. Rowell, “Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G,” IEEE Commun. Mag., vol. 53, no. 1, pp. 186–194, Jan. 2015.
[3] F. Rusek, D. Persson, B. Kiong Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” IEEE Signal Process. Mag., vol. 30, no. 1, pp. 40–60, Jan. 2013.
[4] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 742–758, Oct. 2014.
[5] Q. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., “Spatially sparse precoding in millimeter wave MIMO systems,” IEEE Trans. Wireless Commun., vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
[6] F. Sohrabi and W. Yu, “Hybrid digital and analog beamforming design for large-scale antenna arrays,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 501–513, Apr. 2016.
[7] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., “Limited feedback hybrid precoding for multi-user millimeter wave systems,” IEEE Trans. Wireless Commun., vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
[8] L. Sung, D. Park, and D.-H. Cho, “Limited feedback hybrid beamforming based on dual polarized array antenna,” IEEE Commun. Lett., vol. 22, no. 7, pp. 1486–1489, Jul. 2018.
[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in Proc. NIPS, Stateline, NV, USA, 2012, pp. 1097–1105.
[10] M. Kim, N.-I. Kim, W. Lee, and D.-H. Cho, “Deep learning-sided SCMA,” IEEE Commun. Lett., vol. 22, no. 4, pp. 720–723, Apr. 2018.
[11] W. Lee, M. Kim, and D.-H. Cho, “Deep power control: Transmit power control scheme based on convolutional neural network,” IEEE Commun. Lett., vol. 22, no. 6, pp. 1276–1279, Jun. 2018.
[12] A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, “Deep learning coordinated beamforming for highly-mobile millimeter wave systems,” IEEE Access, vol. 6, pp. 37328–37348, 2018.
[13] Y. Long, Z. Chen, J. Fang, and C. Tellambura, “Data-driven-based analog beam selection for hybrid beamforming under mm-wave channels,” IEEE J. Sel. Topics Signal Process., vol. 12, no. 2, pp. 340–352, May 2018.
[14] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, “Deep-Learning-Based millimeter-wave massive MIMO for hybrid precoding,” IEEE Trans. Veh. Technol., vol. 68, no. 3, pp. 3027–3032, Mar. 2019.
[15] 5G: Study on Channel Model for Frequencies From 0.5 to 100 GHz, document TR38.901 Release 14, May 2017.
[16] R. Zi, X. Ge, H. Wang, J. Zhang, and C.-X. Wang, “Multiuser massive MIMO uplink performance with mutual coupling effects,” in Proc. IEEE Global Commun. Conf., Austin, TX, USA, Dec. 2014, pp. 3296–3301.
[17] T. Taga, “Analysis for mean effective gain of mobile antennas in land mobile radio environments,” IEEE Trans. Veh. Technol., vol. 39, no. 2, pp. 177–181, May 1990.
[18] S. Noh, M. D. Zo¨lowski, and D. J. Love, “Training sequence design for feedback assisted hybrid beamforming in massive MIMO systems,” IEEE Trans. Commun., vol. 64, no. 1, pp. 187–200, Jan. 2016.
[19] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” 2015, arXiv:1502.03167. [Online]. Available: http://arxiv.org/abs/1502.03167
[20] V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in Proc. ICML, Haifa, Isreal, Jun. 2010, pp. 1–8.
[21] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980
[22] T. Oh, C. Song, J. Jung, M. Ahn, J. Kim, and I. Lee, “A new RF-beam training method and asymptotic performance analysis for multi-user millimeter wave systems,” IEEE Access, vol. 6, pp. 48125–48135, 2018.
[23] Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks—Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band, Standard 802.11ad-2012, 2012.
[24] Y. Ren, Y. Wang, C. Qi, and Y. Liu, “Multiple-beam selection with limited feedback for hybrid beamforming in massive MIMO systems,” IEEE Access, vol. 5, pp. 13327–13335, Feb. 2017.
[25] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, “Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels,” IEEE Trans. Signal Process., vol. 52, no. 2, pp. 461–471, Feb. 2004.

LAKIU SUNG was born in Seoul, South Korea, in 1991. He received the B.S., M.S., and Ph.D. degrees in electronic engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2014, 2016, and 2020, respectively. His research interests include 5G wireless communications, millimeter-wave communications, and beamforming.

DONG-HO CHO (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from KAIST, in 1985. From 1987 to 1997, he was a Professor with the Department of Computer Engineering, Kyung Hee University. Since 1998, he has been a Professor with the School of Electrical Engineering, KAIST. He was the Director of the Institute for Information Technology Convergence, KAIST, from 2007 to 2011. He has also been the Director of the KAIST Online Electric Vehicle Project, since 2009. He was the Head of The Cho Chun Shik Graduate School for Green Transportation, from 2010 to 2015. He was also an IT Convergence Campus (ICC) Vice-President of KAIST, from 2011 to 2013. His research interests include mobile communications, online electric vehicle system based on wireless power transfer, and bioinformatics.