Machine Learning-Enabled Joint Antenna Selection and Precoding Design: From Offline Complexity to Online Performance

Thang X. Vu, *Member, IEEE*, Symeon Chatzinotas, *Senior Member, IEEE*, Van-Dinh Nguyen, *Member, IEEE*, Dinh Thai Hoang, *Member, IEEE*, Diep N. Nguyen, *Senior Member, IEEE*, Marco Di Renzo, *Fellow, IEEE*, and Björn Ottersten, *Fellow, IEEE*

***Abstract*—We investigate the performance of multi-user multiple-antenna downlink systems in which a base station (BS) serves multiple users via a shared wireless medium. In order to fully exploit the spatial diversity while minimizing the passive energy consumed by radio frequency (RF) components, the BS is equipped with M RF chains and N antennas, where $M < N$. Upon receiving pilot sequences to obtain the channel state information (CSI), the BS determines the best subset of M antennas for serving the users. We propose a joint antenna selection and precoding design (JASPD) algorithm to maximize the system sum rate subject to a transmit power constraint and quality of service (QoS) requirements. The JASPD overcomes the non-convexity of the formulated problem via a doubly iterative algorithm, in which an inner loop successively optimizes the precoding vectors, followed by an outer loop that tries all valid antenna subsets. Although approaching the (near) global optimality, the JASPD suffers from a combinatorial complexity, which may limit its application in real-time network operations. To overcome this limitation, we propose a learning-based antenna selection and precoding design algorithm (L-ASPA), which employs a deep neural network (DNN) to establish underlaying relations between the key system parameters and the selected antennas. The proposed L-ASPD is robust against the number of users and their locations, BS's transmit power, as well as the small-scale channel fading. With a well-trained learning model, it is shown that the L-ASPD significantly outperforms baseline schemes based on the block diagonalization [5] and a learning-assisted solution for broadcasting systems [29] and achieves higher effective sum rate than that of the JASPA under limited processing time. In addition, we observed that the proposed L-ASPD can reduce the computation complexity by 95% while retaining more than 95% of the optimal performance.**

***Index Terms*—Multiuser, precoding, antenna selection, machine learning, neural networks, successive convex optimization.**

I. Introduction

Multiple-input multiple-output (MIMO) is an enabling technology to deal with the rapidly increasing demand for data-hungry applications in current and future mobile networks. By using a large number of antennas, an MIMO base station is able to send multiple information streams to multiple users simultaneously with negligible inter-user interference. The advantages of MIMO systems, under a proper beamforming design, comprise not only high spectral efficiency but also improved energy efficiency [1]. When the number of antennas in MIMO systems becomes very large, antenna selection (AS) can be employed to improve the performance in terms of both hardware cost and technological aspects [2]. This is due to the fact that the radio frequency (RF) chains are usually much more expensive than antenna elements. More importantly, a proper AS strategy is capable of not only obtaining full spatial diversity but also considerably minimizing the RF chains' energy consumption, hence improving the system energy efficiency [3]. In general, AS is an NP-hard problem whose optimal solution is only guaranteed via exhaustive search, which tries all possible antenna combinations. The high complexity of AS may limit its potential in practice, especially in 5G services which usually have stringent latency and real-time decision making requirements [4].

Low-complexity solutions have become necessary to make AS practically feasible, especially for the BS of medium to large number of antennas. A block diagonalization-based algorithm is proposed in [5] for multiuser MIMO systems, that selects the best antennas to either minimize the symbol error rate (SER) upper bound or maximize the minimum capacity. This method consecutively eliminates one antenna at a time that imposes the most energy in the corresponding orthogonal beamformers. The authors of [6] propose a joint beamforming design and AS algorithm to minimize the multicasting transmit power. By using group sparsity-promoting $l_{1,2}$ norms instead of the l_0 norm, the selected antennas and beamformers can be obtained via an iterative algorithm. The application of $l_{1,2}$ norms is also employed in massive MIMO for minimizing the transmit power [7] and in cell-free MIMO downlink setups for joint access point selection and power allocation [8]. In [9], an AS algorithm based on mirror-prox successive convex approximation (SCA) is proposed for maximizing the minimum rate in multiple-input single-output (MISO) broadcasting systems. A similar SCA-based approach is proposed in [10], [11] for energy efficiency maximization.

Recently, the use of machine learning (ML) in communications systems has attracted much attention [12]–[24].

This work is supported by the European Research Council under project AGNOSTIC, grant 742648. Parts of this work were presented in the WiOPT 2019 [31].

T. X. Vu, S. Chatzinotas, V. D. Nguyen, and B. Ottersten are with the Interdisciplinary Centre for Security, Reliability and Trust (SnT) – University of Luxembourg, L-1855 Luxembourg. Email: {thang.vu, symeon.chatzinotas, dinh.nguyen, bjorn.ottersten}@uni.lu.

D. T. Hoang and D. N. Nguyen are with University of Technilogy Sydney, Sydney, Australia. Email: {hoang.dinh, diep.nguyen}@uts.edu.au.

M. Di Renzo is with the Laboratoire des Signaux et Systèmes, CNRS, CentraleSupélec, Université Paris-Saclay, 91192 Gif-sur-Yvette, France. Email: marco.direnzo@centralesupelec.fr.

The main advantage of ML-aided communications lies in the capability of establishing underlying relations between system parameters and the desired objective, hence being able to shift the computation burden in real-time processing to the offline training phase [25], [26]. The authors of [16] propose a beamforming neural network (BNN) for minimizing the transmit power of multiuser MISO systems, which employs convolutional neural networks (CNN) and a supervised-learning method to predict the magnitude and direction of the beamforming vectors. This method is extended in [17], [18] for unsupervised-learning to maximize the system weighted sum-rate. In [19], a deep learning-aided transmission strategy is proposed for single-user MIMO system with limited feed back, which is capable of addressing both pilot-aided training and channel code selection. The authors of [20] develop a deep learning-based beamforming design to maximize the spectral efficiency of a single-user millimeter wave (mmWave) MISO system, which achieves higher spectral efficiency than conventional hybrid beamforming designs. The application of Q-learning is developed in [21] to overcome the combinatorial-complexity task of selecting the best channel impulse response in vehicle to infrastructure communications. A similar Q-learning based method is proposed in [23] to solve the joint design of beamforming, power control, and interference coordination of cellular networks. In [22], the authors develop a deep reinforcement learning framework which can autonomously optimize broadcast beams in MIMO broadcast systems based on users' measurements. A common data set for training mmWave MIMO networks is provided in [24] regarding various performance metrics.

Towards the learning-aided physical layer design, the application of ML to AS is a promising way to tackle the high-complexity of AS [27]–[30]. A joint design for AS and hybrid beamformers for single-user mmWave MIMO is proposed in [27] based on two serial CNNs, in which one CNN is used to predict the selected antennas and another CNN is used to estimate the hybrid beamformers. The authors of [28] propose a multi-class classification approach to tackle the AS problem in single-user MIMO systems based on two classification methods, namely multiclass k-nearest neighbors and support vector machine (SVM). In [29], a neural network-based approach is proposed to reduce the computational complexity of AS for broadcasting. The neural network (NN) is employed to directly predict the selected antennas that maximize the minimum signal to noise ratio among the users. The authors of [30] propose a learning-based transmit antenna selection to improve the security in the wiretap channel. Therein, two learning-based SVM and naive-Bayes schemes are considered. Although being able to improve the secrecy performance with a reduced feedback overhead, the setup analyzed in [30] is limited to only a single antenna selection.

A. Contributions

In this paper, we investigate the performance of a multiuser MISO downlink system via a joint design of AS and precoding vectors to improve the system sum rate while guaranteeing the users' quality of service (QoS) requirements. Our contributions are as follows:

- First, we develop a joint antenna selection and beamforming design (JASPD) framework to maximize the *effective* system sum rate, which accounts for the time overhead spent on both channel estimation and computational processing, subject to users' QoS requirements and limited transmit power budget. The proposed JASPD works in an iterative manner, which first optimizes the beamforming vectors for a given antenna subset, and then selects the best antenna subset.
- Second, to tackle the non-convexity in optimizing the beamforming vectors of JASPD, we propose two iterative optimization algorithms based on semidefinite relaxation (SDR) and SCA methods. The convergence of the proposed iterative algorithms to at least a local optimum is theoretically guaranteed.
- Third, we propose a learning-based antenna selection and precoding design (L-ASPD) algorithm to overcome the high computational complexity of AS, which employs a deep neural network (DNN) to capture and reveal the relationship between the system parameters and the selected antennas via an offline training process. More importantly, our leaning model is robust against not only the channel fading but also the number of users and their locations. Compared to existing works, which either study single-user MIMO systems [27], [28], a single beamformer for broadcasting [29] or a single antenna selection [30], we consider a more general multi-user system.
- Finally, extensive simulation results show that, under the same limited processing time, the proposed L-ASPD outperforms the JASPD and significantly outperforms existing AS schemes on both model-based [5] and ML-aided [29] designs. We observed that the L-ASPD can achieve more than 95% of the optimal sum rate while reducing more than 95% of the computational time.

The rest of the paper is organized as follows. Section II presents the system model and key parameters. Section III develops two iterative optimization algorithms used in the JASPD. Section IV introduces a ML-aided joint design to accelerate real-time processing. Section V demonstrates the effectiveness of the proposed algorithms via simulation results. Finally, Section IV concludes the paper.

Notations: The superscript $(.)^T$, $(.)^H$ and $\mathrm{Tr}(.)$ stand for the transpose, Hermitian transpose, and trace operation, respectively. $\binom{n}{k}$ represents the binomial coefficients. $|.|$ and $\|.\|$ denote the cardinality and the l_2-norm of a set, respectively.

II. System Model

We consider a multiuser MISO downlink system operated in time division duplex (TDD) mode, in which a multi-antenna base station (BS) servers K single-antenna users in the same frequency resource[1], as depicted in Fig. 1. The BS is equipped with M RF chains and N antennas, where $N > M \geq K$. The motivation of having more antennas than the number of RF chains is that the BS can i) fully exploit spatial diversity

[1]In practice the whole bandwidth is divided into multiple sub-frequency bands. The proposed scheme is directly applied to each band.

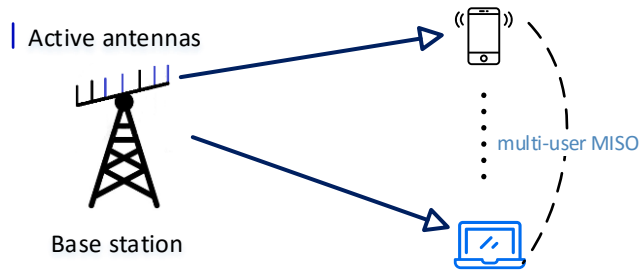


Figure 1. Diagram of multiuser MISO system. A subset of antennas is selected for data transmission.

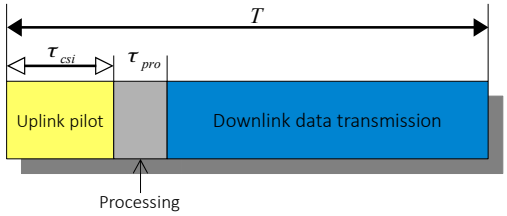


Figure 2. Block diagram of one transmission block.

gain and ii) minimize the static energy consumed by hardware components [3], e.g., RF chains and amplifiers. The system operates in a quasi-static block fading channel in which the channel gains are constant within on block and independently change from one block to another. Before sending data to the users, the BS needs to acquire the channel state information (CSI) via pilot-aided channel estimation[2] in order to perform reprocessing, e.g., beamforming and power allocation.

Fig. 2 illustrates the three phases in one transmission block. Let T and τ_{csi} denote the block duration and channel estimation time, both expressed in terms of channel use (c.u.), respectively. The block duration is determined by the system coherence time. Assuming mutually orthogonal pilot sequences across the users, the channel estimation time is $\tau_{csi} = K(\lfloor N/M \rfloor + 1)$ c.u., where $\lfloor x \rfloor$ denotes the largest integer not exceeding x. Unlike most of previous works that ignore the processing time, we consider the general case in which the processing time takes place in τ_{pro} (c.u.). In practice, the value of τ_{pro} largely depends on beamforming techniques and the hardware capability.

Let $\boldsymbol{h}_k \in \mathbb{C}^{1\times N}$ denote the channel vector from the BS's antennas to user k, including the pathloss. We assume that full CSIs are available at the BS. Because there are only $M < N$ RF chains, the BS has to determine an optimal subset of M antennas for sending data to the users. Let $\mathcal{A} = \{a_1, a_2, \ldots, a_M\}, a_m \in [N] \triangleq \{1, 2, \ldots, N\}$, be a subset of M antennas (out of N), and let $\boldsymbol{\mathcal{A}}$ be the collection of all possible antenna subsets. By definition, we have $|\mathcal{A}| = M$ and $|\boldsymbol{\mathcal{A}}| = \binom{N}{M}$.

Denote by $\boldsymbol{h}_{k,\mathcal{A}} \in \mathbb{C}^{1\times M}$ the channel vector from active antennas in a subset $\mathcal{A}$ to user k, i.e., $\boldsymbol{h}_{k,\mathcal{A}} = [h_k[a_1], h_k[a_2], \ldots, h_k[a_M]]$, where $a_m \in \mathcal{A}$ and $h_k[n]$ is the n-th element of $\boldsymbol{h}_k$. Before serving the users, the BS first precodes the data to suppress inter-user interference. Let $\boldsymbol{w}_{k,\mathcal{A}} \in \mathbb{C}^{M\times 1}$ be the precoding vector for user k corresponding to the selected antenna subset $\mathcal{A}$. The received signal at user k is

$$y_{k,\mathcal{A}} = \boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{k,\mathcal{A}}x_k + \sum\nolimits_{i\neq k}\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{i,\mathcal{A}}x_i + n_k, \quad (1)$$

where n_k is Gaussian noise with zero mean and variance σ^2. The first term in (1) is the desired signal, and the second term is the inter-user interference.

By considering interference as noise, the *effective achievable rate* of user k is

$$R_k(\mathcal{A}) = B\left(1 - \frac{\tau_{csi} + \tau_{pro}}{T}\right) \times \log_2\left(1 + \frac{|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{k,\mathcal{A}}|^2}{\sum_{i\neq k}|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{i,\mathcal{A}}|^2 + \sigma^2}\right), \forall k, \quad (2)$$

where B is the shared channel bandwidth and $1 - \frac{\tau_{csi}+\tau_{pro}}{T}$ accounts for actual time for data transmission. The total transmit power[3] is $\sum_{k=1}^{K}\|\boldsymbol{w}_{k,\mathcal{A}}\|^2$.

Remark 1: It is observed from (2) that the effective data rate is determined not only by the precoding vectors $\boldsymbol{w}_{k,\mathcal{A}}$ but also by the channel estimation and processing times. In particular, spending more time on either channel estimation or processing will degrade the effective transmission rate.

III. Optimal Antenna Selection and Precoding Design

In this section, we develop a joint antenna selection and precoding design to maximize the system sum rate while satisfying the minimum QoS requirements and limited power budget. The joint optimization problem can be formulated as follows:

$$\begin{aligned} \text{P0}: \ & \underset{\mathcal{A}\in\boldsymbol{\mathcal{A}},\{\boldsymbol{w}_{k,\mathcal{A}}\}}{\texttt{maximize}} \ \sum\nolimits_{k=1}^{K} R_k(\mathcal{A}) \\ \texttt{s.t.} \ & R_k(\mathcal{A}) \geq \eta_k, \forall k, \\ & \sum\nolimits_{k=1}^{K}\|\boldsymbol{w}_{k,\mathcal{A}}\|^2 \leq P_{tot}, \end{aligned} \quad (3)$$

where $R_k(\mathcal{A})$ is given in (2), P_{tot} is the total transmit power budget at BS, and η_k is the QoS requirement for user k. In problem (3), the first constraint is to satisfy the minimum user QoS requirement and the second constraint states that the total transmit power should not exceed the power budget. We note that the problem formulation in (3) can be directly extended to the weighted sum rate metric for given weight coefficients with the weights are used as parts of the training input.

In general, problem (3) is a mixed binary non-linear problem where the binary variables of the activated antennas are strongly coupled with the continuous variables of the precoding vectors. Because the precoding vectors are designed for a

[2]The system is assumed to operate above certain SNR levels in which the CSI can be efficiently estimated.

[3]The energy consumed by hardware components is excluded since it is constant and does not affect the precoding design.

given selected antenna subset, problem P0 can be reformulated in an iterative form as follows:

$$\underset{\mathcal{A}\in\boldsymbol{\mathcal{A}}}{\texttt{maximize}}\ \ \mathrm{P1}(\mathcal{A}), \tag{4}$$

where $\mathrm{P1}(\mathcal{A})$ is the precoding design problem for the candidate antenna subset $\mathcal{A}$, which is defined as follows

$$\mathrm{P1}(\mathcal{A}):\ \underset{\{\boldsymbol{w}_{k,\mathcal{A}}\}}{\texttt{Max}}\ \ \bar{B}\sum_{k=1}^{K}\log_2\Big(1+\frac{|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{k,\mathcal{A}}|^2}{\sum_{i\neq k}|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{i,\mathcal{A}}|^2+\sigma^2}\Big) \tag{5}$$

$$\texttt{s.t.}\ \ \bar{B}\log_2\Big(1+\frac{|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{k,\mathcal{A}}|^2}{\sum_{i\neq k}|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{i,\mathcal{A}}|^2+\sigma^2}\Big)\geq\eta_k,\forall k, \tag{5a}$$

$$\sum\nolimits_{k=1}^{K}\|\boldsymbol{w}_{k,\mathcal{A}}\|^2\leq P_{tot}, \tag{5b}$$

where $\bar{B}\triangleq B(1-\frac{\tau_{csi}+\tau_{pro}}{T})$ and we have used (2) for $R_k(\mathcal{A})$.

If problem $\mathrm{P1}(\mathcal{A})$ can be solved optimally, then the optimal solution of P0 can be obtained via an exhaustive search in (4), which tries all possible antenna subsets. Unfortunately, solving problem $\mathrm{P1}(\mathcal{A})$ is challenging due to the non-concavity of the objective function and the non-convexity of the first constraint.

In the following, we propose two solutions based on SDR and SCA methods to tackle the non-convexity of the beamforming vectors design in Section III-A. We then describe the proposed JASPD algorithm and analyze its complexity in Section III-B.

A. Near Optimal Beamforming Design for Selected Antennas

In this subsection, we design the beamforming vectors to maximize the system sum rate for a selected antenna subset. In the following, we propose two methods to solve (5).

1) Semidefinite Relaxation based Solution: Semidefinite-based formulation is an efficient method to design the beamforming vectors of wireless systems, which converts quadratic terms into linear ones by lifting the original variable domain into a higher-dimensional space. We adopt the semidefinite method to deal with the signal-to-noise-plus-interference-ratio (SINR) term in both the objective function and the first constraint. Define a new set of variables $\boldsymbol{W}_k=\boldsymbol{w}_{k,\mathcal{A}}\boldsymbol{w}_{k,\mathcal{A}}^H\in\mathbb{C}^{M\times M}$, and denote $\boldsymbol{H}_k\triangleq\boldsymbol{h}_{k,\mathcal{A}}^H\boldsymbol{h}_{k,\mathcal{A}}$. It is straightforward to verify that $|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{l,\mathcal{A}}|^2=\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{l,\mathcal{A}}\boldsymbol{w}_{l,\mathcal{A}}^H\boldsymbol{h}_{k,\mathcal{A}}^H=\mathrm{Tr}(\boldsymbol{H}_k\boldsymbol{W}_l)$ and $\|\boldsymbol{w}_{k,\mathcal{A}}\|^2=\mathrm{Tr}(\boldsymbol{W}_k)$.

By introducing arbitrary positive variables $\{x_k\}_{k=1}^K$, we can reformulate problem (5) as follows:

$$\underset{\boldsymbol{W},\boldsymbol{x}}{\texttt{maximize}}\ \ \frac{\bar{B}}{\log(2)}\sum\nolimits_{k=1}^{K}x_k \tag{6}$$

$$\texttt{s.t.}\ \ \log\Big(1+\frac{\mathrm{Tr}(\boldsymbol{H}_k\boldsymbol{W}_k)}{\sum_{i\neq k}\mathrm{Tr}(\boldsymbol{H}_k\boldsymbol{W}_i)+\sigma^2}\Big)\geq x_k,\forall k, \tag{6a}$$

$$x_k\geq\frac{\eta_k\log(2)}{\bar{B}},\forall k, \tag{6b}$$

$$\sum\nolimits_{k=1}^{K}\mathrm{Tr}(\boldsymbol{W}_k)\leq P_{tot}, \tag{6c}$$

$$\mathrm{rank}(\boldsymbol{W}_k)=1,\forall k,$$

where we use short-hand notations $\boldsymbol{W}$ and $\boldsymbol{x}$ for $(\boldsymbol{W}_1,\ldots,\boldsymbol{W}_K)$ and $(x_1,\ldots,x_K)$, respectively.

Algorithm 1 ITERATIVE ALGORITHM TO SOLVE (6)

1: Initialize $\boldsymbol{y}_0$, ϵ, X_{old} and `error`.
2: **while** `error` $>\epsilon$ **do**
3: Solve the SDR of (9) by dropping the rank-one constraint to obtain $\{\boldsymbol{W}_{\star k},x_{\star k},y_{\star k}\}_{k=1}^K$
4: Compute `error` $=\frac{\bar{B}}{\log(2)}|\sum_{k=1}^K x_{\star k}-X_{\mathrm{old}}|$
5: Update $X_{\mathrm{old}}\leftarrow\frac{\bar{B}}{\log(2)}\sum_{k=1}^K x_{\star k}$; $y_{0k}\leftarrow y_{\star k},\forall k$

The equivalence between (6) and (5) can be verified as the equality holds in (6a) at the optimum. It is observed that the objective is a linear function and constraints (6b) and (6c) are convex. Thus, the challenge in solving problem (6) lies in (6a) and the rank-one constraint. While the latter constraint can be efficiently coped with by using the relaxation method followed by randomization if needed [32], dealing with the former constraint is more struggling.

In the next step, we introduce slack variables $\{y_k\}_{k=1}^K$ and reformulate constraint (6a) as

$$\log\Big(\sigma^2+\sum\nolimits_{i=1}^{K}\mathrm{Tr}(\boldsymbol{H}_k\boldsymbol{W}_i)\Big)\geq x_k+y_k, \tag{7}$$

$$\sigma^2+\sum\nolimits_{i\neq k}\mathrm{Tr}(\boldsymbol{H}_k\boldsymbol{W}_i)\leq e^{y_k}. \tag{8}$$

Because the function $\log()$ is concave, constraint (7) is convex. However, since the function $\exp(.)$ is convex, constraint (8) is unbounded. To overcome this difficulty, we employ the inner approximation method, which uses the first-order approximation of e^{y_k} at the right hand side of (8). As a result, the approximated problem of (6) can be formulated as follows:

$$\mathrm{P2}(\boldsymbol{y}_0):\ \underset{\boldsymbol{W},\boldsymbol{x},\boldsymbol{y}}{\texttt{maximize}}\ \ \frac{\bar{B}}{\log(2)}\sum\nolimits_{k=1}^{K}x_k \tag{9}$$

$$\texttt{s.t.}\ \ (6\mathrm{b});(6\mathrm{c});(7);\ \mathrm{rank}(\boldsymbol{W}_k)=1,\forall k,$$

$$\sigma^2+\sum\nolimits_{i\neq k}\mathrm{Tr}(\boldsymbol{H}_k\boldsymbol{W}_i)\leq e^{y_{0k}}(y_k-y_{0k}+1),\forall k, \tag{9a}$$

where $\boldsymbol{y}\triangleq\{y_k\}_{k=1}^K$ and $\boldsymbol{y}_0$ is any feasible value of $\boldsymbol{y}$ that satisfies constraint (8).

It is evident that, for a given $\boldsymbol{y}_0$, the objective and constraints of problem (9) are convex except for the rank one constraint. This suggests to solve (9) by the semi-definite relaxation (SDR) method [32] which ignores the rank one constraint and can be solved in an efficient manner by standard solvers, e.g., CVX. Because $e^{y_0}(y-y_0+1)\leq e^y,\forall y_0$, the approximated problem (9) always gives a suboptimal solution of the original problem (6).

It is worth noting that the optimal solution of problem (9) is largely determined by the parameters $\boldsymbol{y}_0$. Thus, it is crucial to select proper values $\boldsymbol{y}_0$ such that the solution of (9) is close to the optimal solution of (6). As such, we propose an iterative optimization algorithm to improve the performance of problem (9), shown in Algorithm 1. The premise behind the proposed algorithm is to better estimate $\boldsymbol{y}_0$ through iterations.

Proposition 1 (Convergence of Algorithm 1): The sequence of the objective values generated by Algorithm 1 in solving the SDR of problem $\mathrm{P2}(\boldsymbol{y}_0)$ is non-decreasing.

The proof of Proposition 1 is shown in Appendix A. Although not guaranteeing the global optimum of problem (6), Propo-

sition 1 justifies the convergence to at least a local optimum of the proposed iterative algorithm[4].

Remark 2 (Initialization of Algorithm 1): The execution of Algorithm 1 requires initial values $y_{0k}, \forall k$. Therefore, it requires an efficient way to find these initial values before tackling problem (9). To this end, we start by solving the feasibility problem below:

$$\begin{aligned} &\texttt{Find} \quad \boldsymbol{W} \qquad (10)\\ &\texttt{s.t.} \quad \frac{\mathrm{Tr}(\boldsymbol{H}_k\boldsymbol{W}_k)}{2^{\eta_k/\bar{B}}-1} \geq \sum\nolimits_{i\neq k}\mathrm{Tr}(\boldsymbol{H}_k\boldsymbol{W}_i)+\sigma^2, \forall k,\\ &\qquad \sum\nolimits_{k=1}^{K}\mathrm{Tr}(\boldsymbol{W}_k) \leq P_{tot}, \end{aligned}$$

which is convex. Then the initial values are computed as $y_{0k} = \log(\sum_{i\neq k}\mathrm{Tr}(\boldsymbol{H}_k\boldsymbol{W}_i^*)+\sigma^2), \forall k$, where $\boldsymbol{W}_k^*$ is the solution of (10).

Remark 3 (Randomization): The solution in (9) is based on the SDR which sometimes violates the rank-one constraint. In such cases, Gaussian randomization can be adopted. Details on Gaussian randomization process are available in [32]. Our simulation results show that more than 99% of the times Algorithm 1 can output rank-one solutions.

2) Reformulation based on Difference of Convex: The SDR-based reformulation in the previous subsection leverages the original problem's non-convexity by working in a higher dimensional domain, which requires more memory. In this subsection, we solve (5) based on difference-of-convex (DC) reformulation directly on the original variable domain.

By introducing arbitrary positive variables $\boldsymbol{u} \triangleq \{u_k\}_{k=1}^K$, we can reformulate problem (5) as follows:

$$\begin{aligned} \underset{\boldsymbol{w},\boldsymbol{u}}{\texttt{Maximize}} \quad & \bar{B}\sum\nolimits_{k=1}^{K}\log_2(1+u_k) && (11)\\ \texttt{s.t.} \quad & \frac{|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{k,\mathcal{A}}|^2}{\sum_{i\neq k}|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{i,\mathcal{A}}|^2+\sigma^2} \geq u_k, \forall k, && (11a)\\ & u_k \geq \bar{\eta}_k, \forall k, && (11b)\\ & \sum\nolimits_{k=1}^{K}\|\boldsymbol{w}_{k,\mathcal{A}}\|^2 \leq P_{tot}, && (11c) \end{aligned}$$

where $\bar{\eta}_k \triangleq 2^{\eta_k/\bar{B}}-1$ and $\boldsymbol{w}$ is a short-hand notation for $(\boldsymbol{w}_{1,\mathcal{A}},\ldots,\boldsymbol{w}_{K,\mathcal{A}})$. The equivalence between (11) and (5) can be verified since constraint (11a) holds with equality at the optimum.

As the denominator of the left-hand-side of (11a) is positive, it can be rewritten as

$$\frac{|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{k,\mathcal{A}}|^2}{u_k} \geq \sum\nolimits_{i\neq k}|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{i,\mathcal{A}}|^2+\sigma^2. \qquad (12)$$

An important observation from (12) is that $\frac{|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{k,\mathcal{A}}|^2}{u_k}$ is a convex function of $\boldsymbol{w}_{k,\mathcal{A}}$ and u_k (see Appendix B). Therefore, (12) has a form of the DC representation, which suggests an efficient way to solve (11a). In particular, let $\hat{\boldsymbol{w}}_{k,\mathcal{A}}, \hat{u}_k$ be any feasible solution of (11), we can approximate (12) by using the first order approximation of the left-hand-side of (12), stated as

$$\begin{aligned} \sum_{i\neq k}\boldsymbol{w}_{k,\mathcal{A}}^H\boldsymbol{H}_k\boldsymbol{w}_{i,\mathcal{A}}+\sigma^2 \leq & \frac{\boldsymbol{w}_{k,\mathcal{A}}^H\left(\boldsymbol{H}_k+\boldsymbol{H}_k^T\right)\hat{\boldsymbol{w}}_{k,\mathcal{A}}}{\hat{u}_k}\\ & - u_k\frac{\hat{\boldsymbol{w}}_{k,\mathcal{A}}^H\boldsymbol{H}_k\hat{\boldsymbol{w}}_{k,\mathcal{A}}}{\hat{u}_k^2}+\frac{\hat{\boldsymbol{w}}_{k,\mathcal{A}}^H\left(\boldsymbol{H}_k-\boldsymbol{H}_k^T\right)\hat{\boldsymbol{w}}_{k,\mathcal{A}}}{\hat{u}_k}, \qquad (13) \end{aligned}$$

which is obviously convex in $\boldsymbol{w}_{k,\mathcal{A}}$ and u_k, where $\boldsymbol{H}_k = \boldsymbol{h}_{k,\mathcal{A}}^H\boldsymbol{h}_{k,\mathcal{A}}$.

By using (13) as an approximation of (11a), problem (11) can be approximated as

$$\begin{aligned} \mathrm{P3}(\hat{\boldsymbol{w}},\hat{\boldsymbol{u}}): \underset{\boldsymbol{w},\boldsymbol{u}}{\texttt{Maximize}} \quad & \bar{B}\sum\nolimits_{k=1}^{K}\log_2(1+u_k) && (14)\\ \texttt{s.t.} \quad & \text{(11b); (11c); (13).} \end{aligned}$$

For given $\hat{\boldsymbol{w}}_{k,\mathcal{A}}, \hat{x}_k$, the objective function in (14) is concave and the constraints are convex, hence it can be solved in an efficient manner by standard solvers, e.g., CVX. Because the right-hand-side of (13) is always less than or equal to $\frac{\boldsymbol{w}_{k,\mathcal{A}}^H\boldsymbol{H}_k\boldsymbol{w}_{k,\mathcal{A}}}{u_k}$, the approximated problem (14) always gives a suboptimal solution of the original problem (11).

In order to reduce the performance gap between the approximated problem (14) and the original problem (11), we propose Algorithm 2 which consists of solving a sequence of SCA problems. The premise behind the proposed algorithm is to better select the parameters $\hat{\boldsymbol{w}}_{k,\mathcal{A}}, \hat{u}_k$ through iterations.

Algorithm 2 ITERATIVE ALGORITHM TO SOLVE (11)

1: Initialize $\hat{\boldsymbol{w}}_{k,\mathcal{A}}, \hat{u}_k$, ϵ, X_{old} and `error`.
2: **while** `error` $> \epsilon$ **do**
3: Solve problem $\mathrm{P3}(\hat{\boldsymbol{w}}_{k,\mathcal{A}}, \hat{u}_k)$ in (14) to obtain $\boldsymbol{w}_k^\star, u_k^\star, \forall k$
4: Compute `error` $= |\bar{B}\sum_{k=1}^{K}\log_2(1+u_k^\star) - X_{\mathrm{old}}|$
5: Update $X_{\mathrm{old}} \leftarrow \bar{B}\sum_{k=1}^{K}\log_2(1+u_k^\star)$; $\hat{\boldsymbol{w}}_{k,\mathcal{A}} \leftarrow \boldsymbol{w}_k^\star$; $\hat{u}_k \leftarrow u_k^\star, \forall k$

Remark 4 (Initialization of Algorithm 2): Finding a feasible point is always essential in the SCA. Intuitively, one can think about the feasibility problem of (5), which is stated as

$$\begin{aligned} &\underset{\{\boldsymbol{w}_{k,\mathcal{A}}\}}{\texttt{Maximize}} \quad 1 && (15)\\ &\texttt{s.t.} \quad \frac{1}{\bar{\eta}_k}|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{k,\mathcal{A}}|^2 \geq \sum_{i\neq k}|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{i,\mathcal{A}}|^2+\sigma^2, \forall k, && (15a)\\ &\qquad \sum\nolimits_{k=1}^{K}\|\boldsymbol{w}_{k,\mathcal{A}}\|^2 \leq P_{tot}. && (15b) \end{aligned}$$

However, since both sides of (15a) are convex, this constraint is unbounded. Therefore, finding a feasible point by solving (15) is not efficient. Instead, we adopt (10) as the mean to find initial values $\hat{\boldsymbol{w}}, \hat{\boldsymbol{u}}$. In particular, from $\boldsymbol{W}_{\star k}, \forall k$, the solution of the convex problem (10), we obtain the corresponding feasible precoding vectors $\boldsymbol{w}_{\star k}$. Then, we assign $\hat{\boldsymbol{w}}_k = \boldsymbol{w}_{\star k}$ and $\hat{u}_k = \frac{|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{\star k}|^2}{\sum_{i\neq k}|\boldsymbol{h}_{k,\mathcal{A}}\boldsymbol{w}_{\star i}|^2+\sigma^2}$.

[4]The study of the performance gap to the global optimum is left for future work.

Algorithm 3 EXHAUSTIVE SEARCH BASED JOINT ANTENNA SELECTION AND PRECODING DESIGN

Inputs: $\boldsymbol{H}, P_{tot}, \{\eta_k\}_{k=1}^K$. **Outputs**: $C_{opt}, \mathcal{A}_{opt}, \boldsymbol{W}_{opt}$

1: Construct the super group $\boldsymbol{\mathcal{A}} = \{\mathcal{A} \mid \mathcal{A} \subset [N], |\mathcal{A}| = M\}$
2: Initialize $C_{opt} = 0$
3: **for** $i = 1 : |\boldsymbol{\mathcal{A}}|$ **do**
4: $\quad \mathcal{A} = \boldsymbol{\mathcal{A}}[i]$
5: $\quad$ Apply Algorithm 1 or Algorithm 2 on the current antenna subset $\mathcal{A}$ to obtain the optimal $X_{old}(\mathcal{A})$ and $\boldsymbol{W}_\star(\mathcal{A})$
6: `If` $C_{opt} < X_{old}(\mathcal{A})$
7: $\quad C_{opt} \leftarrow X_{old}(\mathcal{A})$; $\mathcal{A}_{opt} \leftarrow \mathcal{A}$; $\boldsymbol{W}_{opt} = \boldsymbol{W}_\star(\mathcal{A})$.

B. JASPD Algorithm and Complexity Analysis

Once the precoding vectors have been optimized for each antenna subset, i.e., problem (5) is solved, we can tackle the original optimization problem (3) via Algorithm 3.

The proposed JASPD algorithm consists of two loops: the outer loop tries all valid antenna subsets, and the inner loop optimizes the precoding vectors iteratively. While the complexity of the inner loop is relatively reasonable since (the SDR of) problem (9) (or problem (14)) is convex [36], the outer iteration's complexity increases combinatorially with the number of antennas. In fact, the JASPD has to examine all $\binom{N}{M}$ candidates for the selected antennas. As an example, for $N = 20, M = 8$, there are 125970 possible antenna subsets to be went through, each of which imposes an inner loop in Algorithm 1 or Algorithm 2. Although guaranteeing the maximal achievable rate, the proposed JASPD suffers an exponential complexity due to the selection process. Its high computation time may limit its applicability in practice and degrade the effective rate (see (2)). In the next section, we propose a low-complexity joint design to overcome the computation burden of the antenna selection process.

IV. ACCELERATING THE OPTIMIZATION: A DEEP LEARNING-BASED APPROACH

In this section, we exploit recent advances in machine learning to overcome the major high-complexity limitation of selection process by proposing a learning-based antenna selection and precoding design algorithm (L-ASPD). The premise behind the proposed L-ASPD is to exploit machine-learning based predictions to help the optimal algorithm to tackle the most difficult and time-consuming part in the optimization. In particular, the L-ASPD will first predict potential subsets of antennas, which will be much smaller than $\binom{N}{M}$.

We deploy DNN as the learning model to establish underlaying relations between the system parameters (inputs) and the selected antenna subset. The DNN consists of three main parts: one input layer, one output layer and hidden layers, as depicted in Fig. 3. Based on the labeled data, the DNN will optimize the learning parameters in order to minimize the prediction error, e.g., cost function. The L-ASPD is implemented via 3 steps: i) offline training data generation, ii) building the learning model, and iii) real-time prediction.

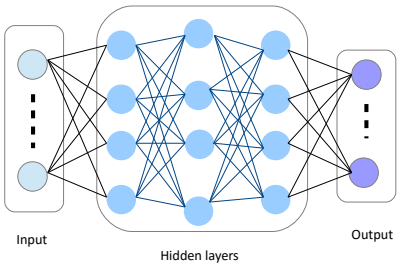


Figure 3. Illustration of a DNN with three hidden layers.

A. Training Data Generation

Since the communication between the BS and the users is specified by the channel gains, the transmit power budget and noise power, they are essential for the learning model. Let $\boldsymbol{H} = [\boldsymbol{h}_1^H, \ldots, \boldsymbol{h}_K^H]^H \in \mathbb{C}^{K\times N}$ denote the channel coefficients from the BS's antennas to all users. Since the number of users can be arbitrary between 1 and M (the number of RF chains), the channel matrix $\boldsymbol{H}$ is first zero-padded to obtain the standard size $\bar{\mathbf{H}} = [\boldsymbol{H}^H, \mathbf{0}_{N\times(M-K)}]^H \in \mathbb{C}^{M\times N}$. Because the NN accepts only real-value inputs, the original complex representation of the channel matrix is invalid. One can stack the real and imaginary parts of $\bar{\mathbf{H}}$ and use them as the training input to the NN [29]. However, we observe that such method is not efficient to our problem because it does not directly capture inter-user interference - the major limiting factor in multiuser systems. As the inter-user interference is determined by the cross-product of the channel vectors of two users, we choose $\boldsymbol{x} = \frac{P_{tot}}{\sigma^2}\text{abs}(\text{vec}(\bar{\mathbf{H}}\bar{\mathbf{H}}^H)) \in \mathbb{R}^{M^2\times 1}$ as the training input. It is worth noting that the training input $\boldsymbol{x}$ is robust against the number of users and pathloss, as well as the BS's transmit power. Last but not least, $\boldsymbol{x}$ should be normalized before being fed to the NN, i.e., $\boldsymbol{x} = \frac{\boldsymbol{x}}{\max(\boldsymbol{x})}$.

Once the input sample is given, we need to define the output, which is the selected antenna combination that provides the maximum objective function in (3). For each training input $\boldsymbol{x}$, we define an output vector $\boldsymbol{b} \in \{0,1\}^{\binom{N}{M}\times 1}$ that consists of all possible antenna subsets. $b[n] = 1$ if the n-th subset is selected, otherwise $b[n] = 0$. Because we are interested in selecting only one subset, we have $\|\boldsymbol{b}\|_0 = 1$. In order to compute $\boldsymbol{b}$, for each channel realization $\boldsymbol{H}$ (corresponding to $\boldsymbol{x}$), we run the proposed JASPD algorithm to find the best antenna subset $\mathcal{A}^\star$ and then assign the output element $b[n^\star] = 1$ corresponding to $\mathcal{A}^\star$. Denote by N_S the number of samples used to train the learning model. The total training input is aggregated in the input matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{N_S}]$, where $\boldsymbol{x}_t$ is the t-th input sample. Similarly, the training output matrix is $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{N_S}]$, where $\boldsymbol{b}_t$ is the t-th output sample corresponding to the input sample $\boldsymbol{x}_t$. The steps for generating the training samples are listed in Table I. We note that JASPD algorithm considered in Table I is used for generating training samples and is executed off-line. Once the NN is well-trained, it is used for only the selected antenna subsets in the real-time prediction phase.

Table I
STEPS TO GENERATE TRAINING SAMPLES FOR L-ASPD

1.	**For** $t = 1 : N_S$
2.	Generate a random number of users K between $[1, M]$.
3.	Generate random locations of these K users between 50 and 300m from the BS. Calculate the pathloss.
4.	Generate a channel matrix $\boldsymbol{H} \in \mathbb{C}^{K\times N}$, including the pathloss.
	Output sample generation
5.	Run JASPD algorithm to find the best antenna subset.
6.	Compute the binary output vector $\boldsymbol{b}_t$ with only a single non-zero element corresponding to the selected subset.
	Input sample generation
5.	Zero-padding: $\bar{\mathbf{H}} = [\boldsymbol{H}^H, \mathbf{0}_{N\times(M-K)}]^H$.
6.	Calculate $\boldsymbol{x}_t = \frac{P_{tot}}{\sigma^2}\text{abs}(\text{vec}(\bar{\mathbf{H}}^H\bar{\mathbf{H}}))$; $\boldsymbol{x}_t = \frac{\boldsymbol{x}_t}{\max(\boldsymbol{x}_t)}$.
7.	**Endfor**

B. Building the Learning Model

When the training data is available, it will be used to train the NN with the learning parameter $\boldsymbol{\Theta}$. For an L-layer NN, we have $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L]$, where $\boldsymbol{\theta}_l \in \mathbb{R}^{N_l\times 1}, 1 \le l \le L$, is the learning parameters in the l-th layer, and N_l is the number of nodes in the l-th layer. As the most popular and efficient candidate for classification problems, we employ a sigmoid-family $\texttt{tansig}(z) = 2(1 + e^{-2z})^{-1} - 1$ as the activation function for the hidden layers and the soft-max as the activation function for the output layer. The learning phase can be done via the minimization of prediction error

$$\Delta(\boldsymbol{\Theta}) = \frac{1}{N_S}\| - \text{Tr}(\boldsymbol{B}^T \log(f_{\boldsymbol{\Theta}}(\boldsymbol{X}))) \\ -\text{Tr}(\bar{\boldsymbol{B}}^T \log(1 - f_{\boldsymbol{\Theta}}(\boldsymbol{X}))) \|^2 + \frac{\lambda}{2N_S}\sum\nolimits_{l=1}^{L} \| \boldsymbol{\theta}_l \|^2, \tag{16}$$

where λ is the regulation parameter, $\bar{\boldsymbol{B}} = \mathbf{1} - \boldsymbol{B}$, and $f_{\boldsymbol{\Theta}}(\boldsymbol{X})$ is the prediction of the output layer.

C. Real-time Prediction

When the NN has been well trained, it is ready to provide real-time and highly accurate predictions. From the current channel coefficient matrix $\boldsymbol{H}$, we construct $\boldsymbol{x} = \frac{P_{tot}}{\sigma^2}\text{abs}(\text{vec}(\bar{\mathbf{H}}^H\bar{\mathbf{H}}))$, where $\bar{\mathbf{H}} = [\boldsymbol{H}^H, \mathbf{0}_{N\times(M-K)}]^H$, which is then normalized to obtain $\boldsymbol{x}_{\text{norm}} = \frac{\boldsymbol{x}}{\max(\boldsymbol{x})}$. Then $\boldsymbol{x}_{\text{norm}}$ is used as the input of the trained NN to output the prediction vector $\hat{\boldsymbol{b}}$. It is worth noting that the NN does not provide absolute prediction, e.g., 0 or 1, but probabilistic uncertainties, e.g., $-1 \le \hat{b}[n] \le 1, \forall n$. In general, the larger an element in $\hat{\boldsymbol{b}}$ is, the higher chance this element is the best antenna subset. Consequently, the subset $\mathcal{A}_{n^\star}$ corresponding to the largest output prediction, i.e., $n^\star = \arg\max_n \hat{b}[n]$, can be selected. However, the prediction is not always precise. Therefore, in order to improve the performance of L-ASPD, instead of choosing only one best candidate, we select K_S subsets, denoted by $\boldsymbol{\mathcal{K}}_S$, corresponding to the K_S largest elements in $\hat{\boldsymbol{b}}$. Then, we apply the precoding design (Algorithm 1 or 2) on these K_S subsets. Intuitively, larger values of K_S will increase the chance for the L-ASPD to select the best antenna subset at an expense of more computation complexity. The steps of the L-ASPD are listed in Algorithm 4. Compared with the JASPD, the L-ASPD significantly reduces the computational time since it tries only K_S promising candidates instead of $\binom{N}{M}$. Consequently, the L-ASPD is expected to achieve higher effective sum rate than that of the JASPD, especially when $K_S \ll \binom{N}{M}$.

Algorithm 4 Proposed L-ASPD Algorithm

Inputs: $\boldsymbol{\Theta}$, $\boldsymbol{H}, P_{tot}, \{\eta_k\}_{k=1}^K$. **Outputs**: $C_{opt}, \mathcal{A}_{opt}, \boldsymbol{w}_{opt}$

Construct $\boldsymbol{x} = \frac{P_{tot}}{\sigma^2}\text{abs}(\text{vec}(\boldsymbol{H}^H\boldsymbol{H}))^2$; $\boldsymbol{x}_{\text{norm}} = \frac{\boldsymbol{x}}{\max(\boldsymbol{x})}$
Apply $\boldsymbol{x}_{\text{norm}}$ to the learned model $\boldsymbol{\Theta}$ to predict $\boldsymbol{\mathcal{K}}_S$
Initialize $C_{opt} = 0$
for $\mathcal{A} \in \boldsymbol{\mathcal{K}}_S$
  Apply Algorithm 1 or 2 on the current subset $\mathcal{A}$ to
  obtain the optimal $X_{old}(\mathcal{A})$ and $\boldsymbol{w}_{\star,\mathcal{A}}$
  if $C_{opt} < X_{old}(\mathcal{A})$
    $C_{opt} = X_{old}(\mathcal{A})$; $\mathcal{A}_{opt} \leftarrow \mathcal{A}$; $\boldsymbol{w}_{opt} \leftarrow \boldsymbol{w}_{\star,\mathcal{A}}$.

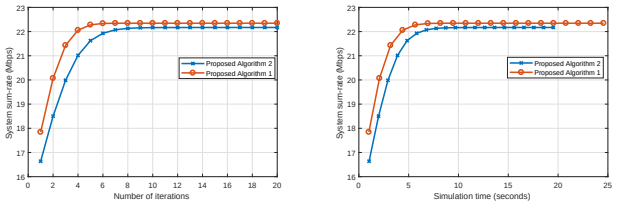
V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed algorithms via simulation results. The users are uniformly distributed in an area between 50 and 300 meters from the centered-BS. We employ the WINNER II line-of-sight pathloss model [33], which results in that the pathloss is uniformly distributed between -59.4 dB and -74.6 dB. All wireless channels are subject to Rayleigh fading. The channel bandwidth $B = 1$ MHz and the noise spectral density is -140 dBm/Hz. We adopt the LTE specifications [34] that one c.u. lasts in one symbol duration and is equal to 66.7 μs, and one block duration is spanned over 200 c.u.. The BS is assumed to spend 0.2 c.u. to solve one convex optimization problem [36]. As a result, it takes $0.2K_S$ c.u. to execute the proposed L-ASPD, where K_S is the number of predicted subsets. We employ an NN with two hidden layers to train the learning model for the L-ASPD, each layer consists of 100 nodes[5]. SVM can also be employed for its fast training phase, however, results in poorer performance compared to NN. This is because SVM results in hyperplanes to discriminate the data whereas the NN can discriminate data using more elaborate functions. The NN is trained using the scaled conjugate gradient method. Other simulation parameters are listed in Table II.

Table II
SIMULATION PARAMETERS

Parameters	Value
Cell radius	300 m
BS's transmit power	1 W - 5W
Number of RF chains M	4
Number of antennas N	Varies
Number of users K	Varies between 1 and M
QoS $\eta_k = \eta, \forall k$	2 Mbps
Training method	Scaled conjugate gradient
Activation function (hidden layers)	$\texttt{tansig}$
Activation function (output layer)	$\texttt{soft}-\texttt{max}$
Loss function	Cross-entropy

[5]We heuristically try a different number of hidden layers and find out that a NN with two hidden layers is sufficient for our problem



(a) Sum rate versus number of iterations (b) Sum rate versus simulation time

Figure 4. Performance comparison of the proposed Algorithm 1 and 2, $P_{tot} = 37$ dBm and $K = 4$. Both algorithms converge in less than 10 iterations.

A. Convergence of the Proposed Optimization Algorithms

We first evaluate the convergence performance of the proposed iterative Algorithm 1 and 2 presented in Section III. The results are obtained from 200 random realizations of channel fading coefficients and users' locations. For each realization, we run both Algorithm 1 and 2 until they converge. Fig. 4a compares the sum-rate obtained by the two proposed algorithms as a function of the iteration number. It is clearly shown that both algorithms converge quickly after less than 10 iterations, which demonstrates the effectiveness of the proposed iterative algorithms.

In order to provide insights on the computation performance of the proposed algorithms, we show in Fig 4b the sum-rate versus the simulation time. Both algorithms are carried out by SeDuMi solver integrated in Matlab 2017b, running on a personal laptop with the Intel i7-6820HQ CPU and 8GB RAM. It is observed that Algorithm 2 executes slightly faster than Algorithm 1, however, achieves a smaller sum-rate. The performance gain brought by Algorithm 1 results from the fact that it uses more memory than Algorithm 2, as shown in Table III. Due to superior performance, we will employ the proposed Algorithm 1 in the remaining comparisons.

Table III
NUMBER OF VARIABLES REQUIRED BY ALGORITHM 1 AND 2 FOR DIFFERENT SETUPS FOR $N = 8$.

M	2	3	4	5
Algorithm 1	267	400	533	666
Algorithm 2	55	94	141	196

B. Performance-complexity Trade-off of the L-ASPD

In this subsection, we examine the efficiency of the proposed L-ASPD via a performance-complexity gain trade-off. By confining the search space of the prediction output, i.e., $K_{\mathcal{S}}$ - the number of potential antenna subsets, we can manage the complexity of L-ASPD since it will work only on $K_{\mathcal{S}}$ candidates. The complexity gain of L-ASPD is defined as the relative time saving compared to the exhaustive search that tries every antenna subsets, calculated as:

$$\theta(K_{\mathcal{S}}) = \frac{\tau(\binom{N}{M} - K_{\mathcal{S}})}{\tau\binom{N}{M}} = 1 - \frac{K_{\mathcal{S}}}{\binom{N}{M}}, \tag{17}$$

where τ is the computational time spent on the optimization of the precoding vectors for a selected antenna subset. The

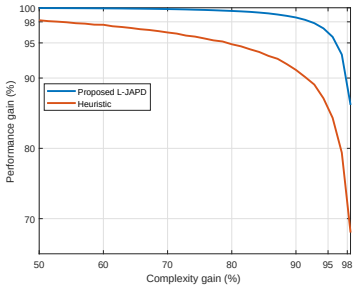


Figure 5. Performance-complexity tradeoff of the proposed L-ASPD. $M = 4, N = 8$.

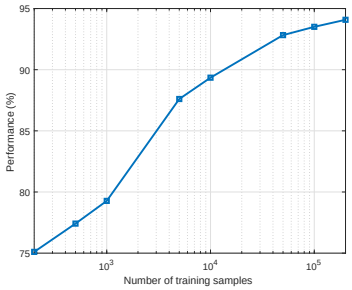


Figure 6. Learning (relative) performance versus the number of training samples. $M = 4, N = 8$.

performance gain is defined as the ratio between the sum rate obtained by L-ASPD divided by the optimal sum rate which is achieved by searching all possible antenna subsets.

Fig. 5 plots the performance-complexity tradeoff of the proposed L-ASPD with $M = 4$ RF chains and $N = 8$ total number of antennas. It is observed that the L-ASPD retains more than 96% of the optimal sum rate (which is obtained by exhaustive search) while saving more than 95% complexity. Even when spending only 2% the computational time, the L-ASPD still achieves 86% the optimal performance, which confirms the effectiveness of the proposed L-ASPD algorithm. Compared with the heuristic solution, the L-ASPD further reduces more than 13% the computational time at the 95% performance gain target.

Fig. 6 plots the relative performance in the real-time prediction of L-ASPD versus the number of training samples. The relative performance is measured as the ratio of the L-ASPD's sum rate divided by the one obtained by the JASPD. Each training sample is generated randomly and captures the randomness in both channel small-scale fading and user location. In general, having more training samples results in better prediction accuracy since the L-ASPD learns more about the intrinsic relation between the selected antennas and the input features. It is shown that 2×10^5 training samples are sufficient for the L-ASPD to achieve more than 94% of the optimal performance.

C. Online Performance Comparison

This subsection demonstrates the effectiveness of the proposed L-ASPD algorithm via performance comparisons with existing solutions in difference scenarios. The first baseline

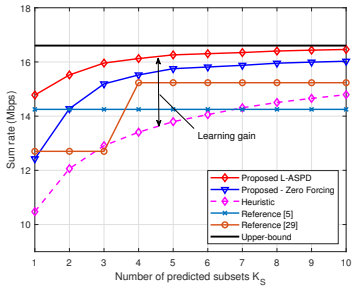


Figure 7. Sum rate performance of the proposed algorithms versus the number of predicted subsets K_S. $P_{tot} = 33$ dBm, $M = 4$ and $N = 8$.

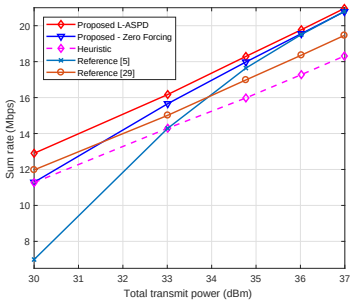


Figure 8. Sume rate performance of the proposed algorithms versus the total transmit power P_{tot}. $K_S = 7$ and $N = 8$ available antennas.

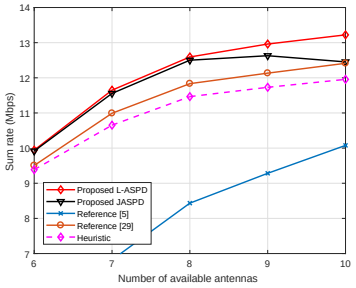


Figure 9. Effective sum rate comparison for various number of total antennas N. P_{tot} = 30 dBm, $M = 4, K_S = 10$.

scheme is proposed in [5], which employs block diagonalization to consecutively eliminate antennas that incur the largest transmit power cost. The second baseline is introduced in [29], which is a learning-assisted antenna selection for multicasting. In addition, a *Heuristic* search is also presented, which also applies the proposed beamforming design but it searches for the antenna subset heuristically. We note that comparison with [27], [28], [30] is not applicable because [27], [28] consider a single-user system and [30] selects only a single antenna.

Fig. 7 shows the achievable sum rate as a function of K_S - the most promising subsets predicted by the proposed L-ASPD algorithm. In order to reveal the benefit of proposed beamforming design in Algorithm 1, we also show a curve, which applies a zero-forcing based power control [35] on the antenna subsets predicted by Algorithm 4. This curve is named as *Proposed - Zero Forcing* in the figures. It is shown that the proposed L-ASPD significantly surpasses all schemes for all observed K_S values. In general, having more predicted subsets K_S results in a larger sum rate, which is in line with results in Fig. 5. In particular, by searching over the most five promising subsets, the proposed L-ASPD achieves 1 Mbps and 2 Mbps higher than schemes in [29] and [5], respectively. We note that the sum rate of the scheme in [5] is independent from K_S since it predicts the best antenna subset. Similarly, the performance curve of [29] has a step-shape because it uses the active antennas as the prediction outputs, hence it is only able to confine the original search space to $\binom{M+n}{M}$ subsets, with $0 \leq n \leq N - M$.

Fig. 8 plots the sum rate as a function of the transmit power. The effectiveness of the proposed learning-based method is shown via the largest sum rate achieved by the L-JAPD compared to other schemes. On average, the L-JAPD algorithm produces 1.5 Mbps and 2 Mbps more than the solution in [29] and heuristic scheme, respectively, proving that the NN has been well trained. Compared to the solution in [5], the L-ASPD achieves a relative sum rate gain of 5 Mbps and 2 Mbps at the transmit power equal to 30 dBm and 33 dBm, respectively. One interesting observation is that the Zero-forcing scheme and the solution in [5] approach the L-ASPD's performance when the total transmit power budget increases. This is because for large P_{tot}, the BS has sufficient power budget to fully mitigate inter-user interference. For small P_{tot}, the system resource becomes scarce, therefore completely eliminating inter-user interference is far from the optimum, which is shown in a big gap between the L-ASPD and these two schemes. In such high-load scenarios, employing the proposed design is highly beneficial.

Fig. 9 presents the effective sum rate for different total antennas numbers N. For a fair comparison, the total transmit power is kept constant at 30 dBm and the total overhead of channel estimation and computation is taken into account. For the former, it takes 8 c.u. to obtain the CSI when the total antenna number is $6, 7, 8$, and takes 12 c.u. when the number of antennas is 9 and 10. Consider the latter, the L-ASPD algorithm only searches over 10 most promising candidates, while the JASPD tries all $\binom{N}{M}$ antenna subsets. In general, having more antennas results in higher effective sum rate of all schemes, which confirms the benefit of antenna selection. Interestingly, the proposed L-ASPD algorithm achieves the best performance and surpasses the exhaustive search scheme, especially for large N, which is in contrast to common understanding that the exhaustive search achieves the best performance. This is because we take the computation time into account in the comparison, as shown in (2). As a result, the exhaustive search scheme spends too much time in searching for the best subset, particularly with large N, resulting in smaller effective rates. As an example for $N = 10$, the exhaustive search scheme requires a computation time which is 21 times more than that of the L-ASPD.

VI. Conclusions

We studied the joint design for antenna selection and precoding vectors in multi-user multi-antenna systems to fully exploit the spatial diversity. We first proposed a (near) optimal

joint antenna selection and precoding algorithm to maximize the system sum rate, subjected to the users' QoS and limited transmit power. The proposed joint design successively optimizes the precoding vectors via two proposed iterative optimization algorithms based on the semidefinite relaxation and successive convex approximation methods. In order to further improve the optimization efficiency, we then developed the machine learning-based solution to provide appropriate and time-stringent antenna predictions. The proposed learning-based algorithm is robust against the number of users and their locations, the BS's transmit power, as well as the channel fading. We showed via simulation results that the proposed learning-based solution significantly surpasses existing selection schemes and outperforms the exhaustive search-based solution.

Based on the outcome of this work, several research directions can be considered. The first problem is how to improve the training phase efficiency, which is especially important when the number of available antennas is very large. In such a case, a low-complexity precoding design, e.g., zero-forcing, can be used to quickly obtain sufficient training samples. The second problem lies in dealing with the network dynamics, which requires the learning model to frequently and timely adapted. Transfer leaning and reinforcement learning are promising solutions in this case to avoid retraining the whole network.

Appendix A
Proof of Proposition 1

Denote $\left(\boldsymbol{W}_\star^{(t)}, \boldsymbol{x}_\star^{(t)}, \boldsymbol{y}_\star^{(t)}\right)$ as the optimal solution of P2$(\boldsymbol{y}_0^{(t)})$ at iteration t. We will show that if $y_{\star k}^{(t)} < y_{0k}^{(t)}, \forall k$, then by using $y_{0k}^{(t+1)} = y_{\star k}^{(t)}$ in the $(t+1)$-th iteration, we will have $\sum_k x_{\star k}^{(t+1)} > \sum_k x_{\star k}^{(t)}$, where $\{x_{\star k}^{(t+1)}\}_{k=1}^K$ is the solution at iteration $t+1$. Indeed, by choosing a relatively large initial value $\boldsymbol{y}_0^{(1)}$, we always have $y_{\star k}^{(1)} < y_{0k}^{(1)}, \forall k$.

Denote $f(y;a) = e^a(y-a+1)$ as the first order approximation of the e^y function at a. At iteration $t+1$, we have $y_{0k}^{(t+1)} = y_{\star k}^{(t)}, \forall k$. Therefore, $f(y; y_{\star k}^{(t)})$ is used in the right-hand side of constraint (9a) at the $(t+1)$-th iteration. Consider a candidate $(y_1^{(t+1)}, \ldots, y_K^{(t+1)})$ for any $y_k^{(t+1)} \in (\hat{y}_k, y_{\star k}^{(t)})$, where $\hat{y}_k = y_{\star k}^{(t)} - 1 + e^{y_{0k}^{(t)} - y_{\star k}^{(t)}}(y_{\star k}^{(t)} - y_{0k}^{(t)} + 1)$. Because function $\exp()$ is convex and $y_k^{(t+1)} < y_{\star k}^{(t)}$, then we have $f(y_k^{(t+1)}; y_{\star k}^{(t)}) > f(y_{\star k}^{(t)}; y_{0k}^{(t)}), \forall k$. Therefore, there exits $\boldsymbol{W}_k^{(t+1)}$ and $x_k^{(t+1)} > x_{\star k}^{(t)}$ which satisfies constraints (7) and (9a). Consider a new set $\{\boldsymbol{W}_k^{(t+1)}, x_k^{(t+1)}, y_k^{(t+1)}\}_{k=1}^K$. This set satisfies all the constraints of problem P2$(\boldsymbol{y}_\star^{(t)})$, and therefore is a feasible solution of the optimization problem. As the result, the optimal objective at iteration $(t+1)$, $\frac{\bar{B}}{\log(2)}\sum_k x_{\star k}^{(t+1)}$, must satisfy $\frac{\bar{B}}{\log(2)}\sum_k x_{\star k}^{(t+1)} \geq \frac{\bar{B}}{\log(2)}\sum_k x_k^{(t+1)} > \frac{\bar{B}}{\log(2)}\sum_k x_{\star k}^{(t)}$, which completes the proof of Proposition 1.

Appendix B
Convexity of function $\frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{y}$

To prove the convexity of $F(\boldsymbol{x}, y) = \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{y}$ for any positive semi-definite matrix $\boldsymbol{A}$, we need to show that the Hessian matrix of $F(\boldsymbol{x}, y)$ is positive semidefinite. Indeed, the Hessian matrix of $F(\boldsymbol{x}, y)$ is

$$\boldsymbol{H}_F = \begin{bmatrix} \frac{\boldsymbol{A}+\boldsymbol{A}^T}{y} & -\frac{(\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{x}}{y^2} \\ -\frac{\boldsymbol{x}^T(\boldsymbol{A}+\boldsymbol{A}^T)}{y^2} & \frac{2\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{y^3} \end{bmatrix}.$$

For arbitrary vector $\boldsymbol{c} = [\boldsymbol{a}^T b]^T$, where $\boldsymbol{a} \in \mathbb{R}^{N\times 1}$, consider a function

$$\begin{aligned} \boldsymbol{c}^T\boldsymbol{H}_F\boldsymbol{c} &= \frac{\boldsymbol{a}^T(\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{a}}{y} - \frac{\boldsymbol{a}^T(\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{x}b}{y^2} \\ &\quad - \frac{\boldsymbol{x}^T(\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{a}b}{y^2} + \frac{2\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}b^2}{y^3} \\ &\overset{(*)}{=} \frac{\boldsymbol{a}^T(\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{a}}{y} - 2\frac{\boldsymbol{a}^T(\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{x}b}{y^2} + \frac{\boldsymbol{x}^T(\boldsymbol{A}+\boldsymbol{A}^T)\boldsymbol{x}b^2}{y^3} \\ &= \frac{\boldsymbol{a}^T\tilde{\boldsymbol{A}}\boldsymbol{a} - 2\boldsymbol{a}^T\tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}}}{y}, \end{aligned} \tag{18}$$

where $\tilde{\boldsymbol{A}} \triangleq \boldsymbol{A}^T + \boldsymbol{A}$, $\tilde{\boldsymbol{x}} \triangleq \boldsymbol{x}b/y$ and $(*)$ results from the fact that $\boldsymbol{A}$ is symmetric and $\boldsymbol{a}^T\tilde{\boldsymbol{A}}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}^T\tilde{\boldsymbol{A}}\boldsymbol{a}$. It is obvious that the RHS of (18) is always non-negative for $y > 0$ and positive semi-definite matrix $\tilde{\boldsymbol{A}}$, which concludes the positive semi-definite of the Hessian matrix of $F(\boldsymbol{x}, y)$.

References

[1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[2] R. Heath and A. Paulraj, "Antenna selection for spatial multiplexing systems based on minimum error rate," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2001, pp. 2276–2280.

[3] Y. Pei, T.-H. Pham, and Y. Liang, "How many RF chains are optimal for large-scale MIMO systems when circuit power is considered?" in *Proc. IEEE Global Commun. Conf.*, Dec. 2012, pp. 3868–3873.

[4] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[5] R. Chen, J. G. Andrews, and R. W. Heath, "Efficient transmit antenna selection for multiuser MIMO systems with block diagonalization," in *Proc. IEEE Global Telecommunications Conference*, Nov. 2007, pp. 3499–3503.

[6] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, May 2013.

[7] S. Qin, G. Li, G. Lv, G. Zhang, and H. Hui, "L1/2-regularization based antenna selection for RF-chain limited massive MIMO systems," in *Proc. IEEE Vehicular Technology Conference (VTC-Fall)*, Sep. 2016, pp. 1–5.

[8] T. X. Vu, S. Chatzinotas, S. ShahbazPanahi, and B. Ottersten, "Joint power allocation and access point selection for cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun.*, May 2020, pp. 1–6.

[9] M. S. Ibrahim, A. Konar, M. Hong, and N. D. Sidiropoulos, "Mirror-prox SCA algorithm for multicast beamforming and antenna selection," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, June 2018, pp. 1–5.

[10] O. Tervo, L. Tran, H. Pennanen, S. Chatzinotas, B. Ottersten, and M. Juntti, "Energy-efficient multicell multigroup multicasting with joint beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4904–4919, Sep. 2018.

[11] S. He, Y. Huang, J. Wang, L. Yang, and W. Hong, "Joint antenna selection and energy-efficient beamforming design," *IEEE Signal Process. Lett.*, vol. 23, no. 9, pp. 1165–1169, Sep. 2016.

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

[13] A Zappone, M Di Renzo, M Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?", *IEEE Trans. Commun.*,vol. 67, no. 10, pp. 7331–7376, Oct. 2019.

[14] A Zappone, M Di Renzo, M Debbah, TT Lam, X Qian, "Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization," *IEEE Veh. Techno. Mag.*, vol. 14, no. 3, pp. 60–69, Sept. 2019.

[15] L. Lei, L. You, G. Dai, T. X. Vu, D. Yuan, and S. Chatzinotas, "A deep learning approach for optimizing content delivering in cache-enabled HetNet," in *Proc. IEEE Int. Symp. Wireless Commun. Syst.*, Aug. 2017, pp. 449–453.
[16] W. Xia, G. Zheng, Y. Zhu, J. Zhang, J. Wang, and A. P. Petropulu, "A deep learning framework for optimization of MISO downlink beamforming," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1866–1880, March 2020.
[17] H. Huang, W. Xia, J. Xiong, J. Yang, G. Zheng, and X. Zhu, "Unsupervised learning-based fast beamforming design for downlink MIMO," *IEEE Access*, vol. 7, pp. 7599–7605, 2019.
[18] H. Huang, Y. Peng, J. Yang, W. Xia, and G. Gui, "Fast beamforming design via deep learning," *IEEE Trans. Veh. Techno.*, vol. 69, no. 1, pp. 1065–1069, Jan 2020.
[19] J. Jang, H. Lee, S. Hwang, H. Ren, and I. Lee, "Deep learning-based limited feedback designs for MIMO systems," *IEEE Wireless Commun. Lett.*, pp. 1–1, 2019.
[20] T. Lin and Y. Zhu, "Beamforming design for large-scale antenna arrays using deep learning," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 103–107, Jan 2020.
[21] T. E. Bogale, X. Wang, and L. Le, "Adaptive channel prediction, beamforming and scheduling design for 5G V2I network: Analytical and machine learning approaches," *IEEE Trans. Veh. Techno.*, pp. 1–1, 2020.
[22] R. Shafin, H. Chen, Y. H. Nam, S. Hur, J. Park, J. Zhang, J. Reed, and L. Liu, "Self-tuning sectorization: Deep reinforcement learning meets broadcast beam optimization," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2020.
[23] F. B. Mismar, B. L. Evans, and A. Alkhateeb, "Deep reinforcement learning for 5G networks: Joint beamforming, power control, and interference coordination," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1581–1592, March 2020.
[24] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. Info. Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2019, pp. 1–8.
[25] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for wireless resource management," in *IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Jul. 2017, pp. 247—-252.
[26] L. Lei, T. X. Vu, L. You, S. Fowler, and D. Yuan, "Efficient minimum-energy scheduling with machine-learning based predictions for multiuser MISO systems," in *Proc. IEEE Int. Conf. Commun.*, May 2018, pp. 1–6.
[27] A. M. Elbir and K. V. Mishra, "Joint antenna selection and hybrid beamformer design using unquantized and quantized deep learning networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1677–1688, Mar. 2020.
[28] J. Joung, "Machine learning-based antenna selection in wireless communications," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2241–2244, Nov. 2016.
[29] M. S. Ibrahim, A. S. Zamzam, X. Fu, and N. D. Sidiropoulos, "Learning-based antenna selection for multicasting," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2018, pp. 1–5.
[30] D. He, C. Liu, T. Q. S. Quek, and H. Wang, "Transmit antenna selection in MIMO wiretap channels: A machine learning approach," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 634–637, Aug. 2018.
[31] T. X. Vu, L. Lei, S. Chatzinotas, and B. Ottersten, "Machine learning based antenna selection and power allocation in multi-user MISO systems," in *Proc. Int. Symp. on Modeling and Opt. in Mobile, Ad Hoc, and Wireless Netw.*, Jun. 2019, pp. 1–6.
[32] Z.-Q. Luo, W. K. Ma, A. M. C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, Mar. 2010.
[33] P. Kyosti and et al., "WINNER II channel models," 2007, tech. rep. D1.1.2 V1.2.
[34] T. Innovations, "LTE in a nutshell: The physical layer," 2010, white paper.
[35] T. X. Vu, L. Lei, S. Vuppala, A. Kalantari, S. Chatzinotas, and B. Ottersten, "Latency minimization for content delivery networks with wireless edge caching," in *Proc. IEEE Int. Conf. Commun.*, Kansas City, MO, 2018, pp. 1–6.
[36] J. Mattingley and S. Boyd, "Real-time convex optimization in signal processing," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 50–61, May 2010.

Thang X. Vu (S'11–M'15) received the B.S. and the M.Sc., both in Electronics and Telecommunications Engineering, from the VNU University of Engineering and Technology, Vietnam, in 2007 and 2009, respectively, and the Ph.D. in Electrical Engineering from the University Paris-Sud, France, in 2014. In 2010, he received the Allocation de Recherche fellowship to study Ph.D. in France. From September 2010 to May 2014, he was with the Laboratory of Signals and Systems (LSS), a joint laboratory of CNRS, CentraleSupelec and University Paris-Sud XI, France. From July 2014 to January 2016, he was a postdoctoral researcher with the Information Systems Technology and Design (ISTD) pillar, Singapore University of Technology and Design (SUTD), Singapore. Currently, he is a research scientist at the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg. His research interests are in the field of wireless communications, with particular interests of wireless edge caching, cloud radio access networks, machine learning for communications and cross-layer resources optimization. He was a recipient of the SigTelCom 2019 best paper award.

Symeon Chatzinotas (S'06—M'09—SM'13) is currently the Deputy Head of the SIGCOM Research Group, Interdisciplinary Centre for Security, Reliability, and Trust, University of Luxembourg, Luxembourg and Visiting Professor at the University of Parma, Italy. He received the M.Eng. degree in telecommunications from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2003, and the M.Sc. and Ph.D. degrees in electronic engineering from the University of Surrey, Surrey, U.K., in 2006 and 2009, respectively. He was involved in numerous Research and Development projects for the Institute of Informatics Telecommunications, National Center for Scientific Research Demokritos, the Institute of Telematics and Informatics, Center of Research and Technology Hellas, and the Mobile Communications Research Group, Center of Communication Systems Research, University of Surrey. He has over 250 publications, 2000 citations, and an H-Index of 25 according to Google Scholar. His research interests include multiuser information theory, co-operative/cognitive communications, and wireless networks optimization. He was a co-recipient of the 2014 Distinguished Contributions to Satellite Communications Award, and the Satellite and Space Communications Technical Committee, the IEEE Communications Society, and the CROWNCOM 2015 Best Paper Award.

Van-Dinh Nguyen (S'14—M'19) received the B.E. degree in electrical engineering from Ho Chi Minh City University of Technology, Vietnam, in 2012 and the M.E. and Ph.D. degrees in electronic engineering from Soongsil University, Seoul, South Korea, in 2015 and 2018, respectively. He is currently a Research Associate with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg. He was a Postdoc Researcher and a Lecturer with Soongsil University, a Postdoctoral Visiting Scholar with University of Technology Sydney, AUS (July-August 2018) and a Ph.D. Visiting Scholar with Queen's University Belfast, U.K. (June-July 2015 and August 2016). His current research activity is focused on fog/edge computing, Internet of Things, 5G networks and machine learning for wireless communications.

Dr. Nguyen received several best conference paper awards, IEEE Transaction on Communications Exemplary Reviewer 2018 and IEEE GLOBECOM Student Travel Grant Award 2017. He has authored or co-authored in some 40papers published in international journals and conference proceedings. He has served as a reviewer for many top-tier international journals on wireless communications, and has also been a Technical Programme Committee Member for several flag-ship international conferences in the related fields. He is an Editor for the IEEE Open Journal of the Communications Society and IEEE Communications Letters.

Dinh Thai Hoang (M'16) is currently a faculty member at the School of Electrical and Data Engineering, University of Technology Sydney, Australia. He received his Ph.D. in Computer Science and Engineering from the Nanyang Technological University, Singapore, in 2016. His research interests include emerging topics in wireless communications and networking such as machine learning, ambient backscatter communications, IRS, mobile edge intelligence, cybersecurity, IoT, and 5G/6G networks. He has received several awards including Australian Research Council. Currently, he is an Editor of IEEE Wireless Communications Letters and IEEE Transactions on Cognitive Communications and Networking.

Diep N. Nguyen (SM'19) received the M.E. and Ph.D. degrees in electrical and computer engineering from University of California San Diego (UCSD) and The University of Arizona (UA), respectively. He is a Faculty Member of the Faculty of Engineering and Information Technology, University of Technology Sydney (UTS). Before joining the UTS, he was a DECRA Research Fellow at Macquarie University, and a Member of Technical Staff at Broadcom, CA, USA, and ARCON Corporation, Boston, consulting the Federal Administration of Aviation on turning detection of UAVs and aircraft, U.S. Air Force Research Lab on anti-jamming. His current research interests include computer networking, wireless communications, and machine learning applications, with an emphasis on systems performance and security/privacy. He has received several awards from LG Electronics, UCSD, UA, U.S. National Science Foundation, and Australian Research Council. He is an Editor, Associate Editor of the Transactions on Mobile Computing, IEEE Access, Sensors journal, and IEEE Open Journal of the Communications Society (OJ-COMS).

Marco Di Renzo (S'05–AM'07–M'09–SM'14-F'19) was born in LAquila, Italy, in 1978. He received the Laurea (cum laude) and Ph.D. degrees in electrical engineering from the University of LAquila, Italy, in 2003 and 2007, respectively, and the Habilitation a Diriger des Recherches (Doctor of Science) degree from University Paris-Sud, France, in 2013. Since 2010, he has been with the French National Center for Scientific Research (CNRS), where he is a CNRS Research Director (CNRS Professor) in the Laboratory of Signals and Systems (L2S) of Paris-Saclay University - CNRS, CentraleSupelec, Univ Paris Sud, Paris, France. He serves as the Editor-in-Chief of IEEE Communications Letters, and as an Editor of IEEE Transactions on Communications, and IEEE Transactions on Wireless Communications. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society and IEEE Communications Society.

He is a recipient of several awards, including the 2013 IEEE-COMSOC Best Young Researcher Award for Europe, Middle East and Africa, the 2013 NoENEWCOM Best Paper Award, the 2014-2015 Royal Academy of Engineering Distinguished Visiting Fellowship, the 2015 IEEE Jack Neubauer Memorial Best System Paper Award, the 20152018 CNRS Award for Excellence in Research and Ph.D. Supervision, the 2016 MSCA Global Fellowship (declined), the 2017 SEE-IEEE Alain Glavieux Award, the 2018 IEEECOMSOC Young Professional in Academia Award, and 8 Best Paper Awards at IEEE conferences (2012 and 2014 IEEE CAMAD, 2013 IEEE VTC-Fall, 2014 IEEE ATC, 2015 IEEE ComManTel, 2017 IEEE SigTelCom, EAI 2018 INISCOM, IEEE ICC 2019). He is a highly cited researcher according to Clarivate Analytics and Web of Science, and a Fellow of the IEEE.

Bjorn Ottersten (S'87–M'89–SM'99–F'04) received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1990. He has held research positions with the Department of Electrical Engineering, Linköping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, Leuven, Belgium, and the University of Luxembourg, Luxembourg. From 1996 to 1997, he was the Director of Research with ArrayComm, Inc., a start-up in San Jose, CA, USA, based on his patented technology. In 1991, he was appointed Professor of signal processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. Dr. Ottersten has been Head of the Department for Signals, Sensors, and Systems, KTH, and Dean of the School of Electrical Engineering, KTH. He is currently the Director for the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg.

He is a recipient of the IEEE Signal Processing Society Technical Achievement Award, the EURASIP Group Technical Achievement Award, and the European Research Council advanced research grant twice. He has co-authored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, 2013, and 2019, and 8 IEEE conference papers best paper awards. He has been a board member of IEEE Signal Processing Society, the Swedish Research Council and currently serves of the boards of EURASIP and the Swedish Foundation for Strategic Research. Dr. Ottersten has served as Editor in Chief of EURASIP Signal Processing, and acted on the editorial boards of IEEE Transactions on Signal Processing, IEEE Signal Processing Magazine, IEEE Open Journal for Signal Processing, EURASIP Journal of Advances in Signal Processing and Foundations and Trends in Signal Processing. He is a fellow of EURASIP.