Data-driven Based Analog Beam Selection for Hybrid Beamforming under Mm-Wave Channels

Yin Long, Zhi Chen, *Senior Member, IEEE,* Jun Fang, *Member, IEEE,* Chintha Tellambura, *Fellow, IEEE*

***Abstract*—Hybrid beamforming is a promising low-cost solution for large multiple-input multiple-output (MIMO) systems, where the base station (BS) is equipped with fewer radio frequency chains. In these systems, the selection of codewords for analog beamforming is essential to optimize the uplink sum-rate. In this paper, based on machine learning, we propose a data-driven method of analog beam selection to achieve a near-optimal sum-rate with low complexity, which is highly dependent on training data. Specifically, we take the beam selection problem as a multiclass-classification problem, where the training data set consists of a large number of samples of the millimeter-wave channel. Using this training data, we exploit the support vector machine (SVM) algorithm to obtain a statistical classification model, which maximizes the sum rate. For real-time transmissions, with the derived classification model, we can select, with low complexity, the optimal analog beam of each user. We also propose a novel method to determine the optimal parameter of Gaussian kernel function via McLaughlin expansion. Analysis and simulation results reveal that, as long as the training data is sufficient, the proposed data-driven method achieves a near-optimal sum-rate performance, while the complexity reduces by several orders of magnitude, compared to the conventional method.**

***Index Terms*—hybrid beamforming, data-driven solution, mm-wave, beam selection, SVM**

I. Introduction

Although the fifth generation (5G) mobile communications standards are still very much evolving, the aims for higher data rates, lower latency, and higher energy-efficient performance are firmly clear [1]. These aims bring about the demands for wider bandwidth spectrum. Currently, available bandwidth in the spectrum up through 6 GHz is not sufficient to satisfy these requirements. This shortage, in turn, has helped us move the target operating frequency bands up into the millimeter-wave (mm-wave) [2] range for the next generation of wireless communication systems [3] [4]. The shorter wavelengths at these higher frequency bands enable implementations with many more antenna elements per system within a super-small space [5] [6]. However, it also increases the signal-path and propagation challenges associated with operating at these frequencies. For example, due to the gas absorption, the attenuation for a 60 GHz waveform is more than 10 dB/km, while a 700 MHz waveform experiences an attenuation on the order of 0.01 dB/km.

These losses can be compensated with the elaborate array design and the application of spatial signal-processing techniques, including beamforming. Beamforming can be enabled by large antenna arrays and can be applied directly to provide higher transmit gains to cope with the path loss and harmful interference signals.

To achieve a desirable flexibility and controllability with beamforming in the design of antenna array, adopting an independent weighting control over each antenna-array element is a feasible method. This requires a transmit or receive component dedicated to each antenna-array element. However, for large multiple-input multiple-output (MIMO) systems [7] [8] whose array size is over a hundred antennas, such an architecture is rather difficult to build due to cost, space, and power limitations. For example, implementing a high performance analog-to-digital converter (ADC) and digital-to-analog converter (DAC) for each channel can drive the cost and power beyond an affordable budget. Similarly, having variable gain amplifiers in the radio frequency (RF) chain for each channel can increase system cost.

Hybrid beamforming [9] [10] [11] is a popular technique that can be used to partition beamforming into digital domain and RF domain. Therefore, hybrid beamforming can be implemented to balance tradeoffs between cost and flexibility, while still fielding a system that meets the required performance parameters. Hybrid-beamforming designs are developed by combining multiple array elements into subarray modules. A transmit or receive module can be dedicated to multiple elements in the array. Thus, the system will need fewer transmit or receive components (i.e., RF chains). The number of elements in each subarray can be selected to ensure that system performance is met across the range of steering angles. Using the transmit path as an example, each element within a subarray can have a phase shift applied directly in the RF domain, while digital beamforming techniques based on complex weighting vectors can be applied on the signals that feed each subarray. Digital beamforming is able to conduct the control of the signal for both amplitude and phase on signals aggregated at the subarray level. Consequently, a cost-efficient MIMO system architecture for low-cost deployment is proposed, which is called hybrid MIMO.

In hybrid beamforming, each RF chain is equipped with a bunch of phase shifters to conduct analog beamforming. Thus, to ensure a high performance in terms of sum-rate or bit

This work was supported in part by the Important National Science and Technology Specific Projects of China under Grant 2017ZX03001010, and by the Science and Technology on Communication Networks Laboratory fund.

Yin Long is with Science and Technology on Communication Networks Laboratory, University of Electronic Science and Technology of China, China, e-mail: (yinnuoloong@163.com); Zhi Chen and Jun Fang are with National Key Lab of Science and Technology on Communications, University of Electronic Science and Technology of China, China, e-mail: (chenzhi@uestc.edu.cn); (JunFang@uestc.edu.cn); Chintha Tellambura is with Department of Electrical and Computer Engineering, University of Alberta, Canada, e-mail: (chintha@ece.ualberta.ca)

error rate for hybrid MIMO, choosing suitable analog beams for each RF chain plays a key role. Thus, recently, a plenty of works have focused on the selection of analog beams. In [12], a low-complexity analog beam selection scheme under point-to-point scenarios is proposed. When the number of candidates of analog beams is small, the proposed scheme is able to achieve a near-optimal spectral efficiency at high SNR regime. Literature [13] presents two beam selection algorithms for analog beamforming based on rotman lens theory, which is able to achieve higher BER performance. In [14], an exhaustive method is proposed to select the analog beams that make SNR or SINR maximum. However, so far, all the related works try to find the optimal combination of analog beams by evaluating the design metric over all possible combinations. Nevertheless, evaluating the design metric is a high-complexity task, thus choosing suitable analog beams for each RF chain is a high complexity-cost procedure, which poses an unacceptable delay on real-time communications. Therefore, developing a low-complexity method is motivated.

Recently, big data [15] [16], which is an emerging technology about extracting meaningful value from large volume of data, has attracted a plenty of interests in various fields. Big data enables us to harness the volume, variety, and velocity of data and deduce actionable insight from data. In the study of cellular networks, big data would bring us huge opportunities to innovate cellular networks, since big data is able to provide novel efficient solutions to the design or optimization of cellular networks. For example, cellular networks embracing big data have been studied in [17]. A self-optimizing 5G networking based on big data is proposed in [18]. Furthermore, as mentioned in [17] and [18], machine learning [19] is a powerful tool in big data, which is able to dig hidden insights from training data and make a judgment for a new data set.

In this paper, to solve the analog beam selection problem in a low-complexity way, we propose a data-driven solution by resorting to support vector machine (SVM) [20]. SVM is a preferred multi-class classification algorithm [21] in machine learning, which is good at handling linearly inseparable dataset of samples and avoiding over-fitting. To begin with, we consider the beam selection problem as a multi-class classification problem, where a large number of samples of mm-wave channel are taken as training data. Based on these training data, we adopt SVM algorithm to obtain a statistical classification model in terms of maximizing sum-rate performance. By using the derived classification model, we can choose the optimal analog beam for each user with low complexity in the middle of real-time transmission. Analysis and simulation results reveal that, if training data can be provided sufficiently, the proposed data-driven method is able to achieve a near-optimal sum-rate performance, while the complexity would reduce by several orders of magnitude, compared with the conventional exhaustive method.

To the best of our knowledge, this paper is the first attempt to solve the problem of beam selection by the data-driven method. Our main contributions are as follows:

1) As we know, directly calculating sum-rate for zero-forcing (ZF) digital beamformer is involved with several matrix inversion operations, which is a high-complexity manipulation. Thus, in this paper, by using the vector-combined manipulation, we derive a low-complexity metric to measure sum-rate. Moreover, we take the derived metric as the key performance indicator (KPI) of each possible combination of analog beams.
2) In machine learning, training data is presented as feature vector whose dimensionality is proportional to the complexity of classification. In order to reduce the complexity of classification, we take the direction of arrival (DoA) and angle Of arrival (AoA) of each path of mm-wave channels as entries of feature vectors. Due to the sparsity of mm-wave channels [22], the number of transmission path of mm-wave channels is few. Hence, the dimensionality of feature vector of training data is very small, which is able to suppress the complexity of classification.
3) Generally, in hybrid beamforming, a codebook of analog beams provides more than two candidates for beams selection, which brings about the imbalance of training data for an one-vs-the-rest classifier [20]. However, the regular SVM does not perform well on imbalanced data. Therefore, we propose a biased-SVM where the major training data and minor training data use different error penalty, respectively.
4) The classification performance of SVM primarily depends on the design parameter of kernel functions. To achieve a high classification performance, we propose a new method to determine the optimal design parameter of the Gaussian kernel function in virtue of McLaughlin expansion. The experiment results indicate that the proposed method can achieve a better classification performance than conventional cross-validation method.

The reminder of this paper is organised as follows. In Section II, we introduce the system model. In Section III, a low-complexity metric for sum-rate is obtained. In Section IV, a data-driven solution for analog beam selection is proposed. In Section V, the complexity analysis is conducted. Simulation results are presented in Section VI. Finally, the paper is summarized in Section VII.

Notations: x, $\mathbf{x}$ and $\mathbf{X}$ denote scalar, vector and matrix, respectively. $\mathbf{x}^T$ represents the transpose of vector $\mathbf{x}$. $diag[x_1, x_2, \cdots, x_k]$ denotes a matrix whose diagonal elements are composed by $x_1, x_2, \cdots, x_k$ while the rest of elements are zero. $\|\cdot\|$ represents Forbinius norm. $\mathcal{S}$ denotes a set, and $|\mathcal{S}|$ is the cardinality of set $\mathcal{S}$.

II. SYSTEM MODEL

We introduce the system model in this section. As shown in Fig. 1, an uplink massive MIMO system with hybrid beamforming is considered. In the given system, the BS employs N_B antennas to serve K mobile users, while each user is equipped with N_u antennas. As illustrated in Fig. 1, the hybrid beamforming architecture comprises digital beamforming part and analog beamforming part. In the analog beamforming part, there are N_s RF chains at the BS, where each RF chain is equipped with N_R antennas. Consequently, for the number of antennas employed at the BS, we have $N_B = N_s \times N_R$. Here, we assume that the number of users is no more than the number of RF chains, i.e., $K \leq N_s$.

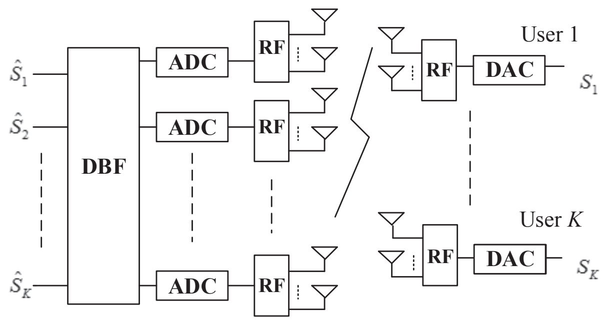

Fig. 1. System model

A. Analog beamforming

Basically, as we know, analog beamforming is aim to adjust the phase of the transmitted or received signal at antennas in the RF domain by using phase shifters in front of each antenna.

To begin with, we assume only a single data stream needs to be transmitted for each user. Hence, when the kth ($k \in \{1, 2, \cdots, K\}$) user transmits uplink signal to the BS through an analog beam, the uplink signal of the kth user can be written as

$$\mathbf{x}_k = \mathbf{c}_k s_k \tag{1}$$

where $s_k \in \mathbb{C}^{1\times 1}$ is the data symbol of the kth user, $\mathbf{c}_k \in \mathbb{C}^{N_u\times 1}$ is the analog beam for the kth user, and the ith entry of $\mathbf{c}_k$ is the value of phase shift on the ith antenna, denoted by $e^{j\theta_i^k}$ with $\theta_i^k \in [0, 2\pi]$. For each user, the maximum transmit power of the uplink signal is P, namely, $\mathbb{E}\left[\|\mathbf{x}_k\|^2\right] \le P$.

Then, the received data steams at the BS is expressed by

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{H}_i \mathbf{x}_i + \mathbf{n} = \sum_{i=1}^{K} \mathbf{H}_i \mathbf{c}_i s_i + \mathbf{n} \tag{2}$$

where $\mathbf{H}_i \in \mathbb{C}^{N_B\times N_u}$ is the uplink channel matrix of user i, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_B})$ represents the additive white Gaussian noise (AWGN) at the BS.

On the other hand, the receive phase shifter vector for the lth RF chain of the BS can be given by

$$\mathbf{g}_l = \left[e^{j\theta_1^l}, \cdots, e^{j\theta_{N_R}^l}\right]^T \tag{3}$$

where the ith entry of $\mathbf{g}_l$, $e^{j\theta_i^l}$, is the value of phase shifter on ith antenna of the lth RF chain. Hence, based on the system model mentioned above, the receive phase shifter matrix at the BS, $\mathbf{G}$, can be written as a $N_s \times N_B$ block diagonal matrix which is consisted of the N_s receive phase shifter vectors and can be expressed as

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_1^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{g}_1^T & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{g}_{N_s}^T \end{bmatrix}. \tag{4}$$

Then, after being processed by the receive phase shifter matrix, the received signal can be given by

$$\begin{aligned} \bar{\mathbf{y}} &= \mathbf{G} \sum_{i=1}^{K} \mathbf{H}_i \mathbf{x}_i + \mathbf{G}\mathbf{n} = \sum_{i=1}^{K} \mathbf{G}\mathbf{H}_i \mathbf{c}_i s_i + \mathbf{G}\mathbf{n} \\ &= \sum_{i=1}^{K} \bar{\mathbf{h}}_i s_i + \mathbf{G}\mathbf{n} \\ &= \left[\bar{\mathbf{h}}_1 \bar{\mathbf{h}}_2 \cdots \bar{\mathbf{h}}_K\right] \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_K \end{bmatrix} + \mathbf{G}\mathbf{n} \\ &= \bar{\mathbf{H}}\mathbf{s} + \mathbf{G}\mathbf{n} \end{aligned} \tag{5}$$

where $\bar{\mathbf{h}}_i \triangleq \mathbf{G}\mathbf{H}_i\mathbf{c}_i$ is the equivalent channel vector for the uplink channel of the ith user and $\bar{\mathbf{H}} \triangleq \left[\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2, \cdots \bar{\mathbf{h}}_K\right]$.

B. Digital beamforming

In the baseband process, the ZF beamforming is considered to detect the each user's uplink signal. Based on the criterion of ZF, the receive digital-beamforming matrix is the pseudo inverse of $\bar{\mathbf{H}}$, which is given by

$$\mathbf{W} = \left((\bar{\mathbf{H}})^H \bar{\mathbf{H}}\right)^{-1} (\bar{\mathbf{H}})^H. \tag{6}$$

The detected signal by using ZF beamforming can be expressed as

$$\begin{aligned} \tilde{\mathbf{y}} &= \mathbf{W}\bar{\mathbf{y}} = \left((\bar{\mathbf{H}})^H \bar{\mathbf{H}}\right)^{-1} (\bar{\mathbf{H}})^H \bar{\mathbf{y}} \\ &= \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_K \end{bmatrix} + \left((\bar{\mathbf{H}})^H \bar{\mathbf{H}}\right)^{-1} (\bar{\mathbf{H}})^H \mathbf{G}\mathbf{n}. \end{aligned} \tag{7}$$

C. Mm-wave Channel

Although hybrid beamforming is able to be operated in Rayleigh fading conditions, we need to adopt the mm-wave frequency band due to the demands for wider bandwidth spectrum. Hence, in this paper, we adopt the most widely applicable geometric channel models. As a mm-wave channel model, the geometric channel model has L limited scattering path. Consequently, the uplink channel of user k, $\mathbf{H}_k$, can be written as

$$\mathbf{H}_k = \sqrt{\frac{N_B N_u}{L\rho_k}} \sum_{i=1}^{L} \alpha_{k,i} \mathbf{a}_{BS}(\theta_{k,i}^{BS}) \mathbf{a}_{user}^H(\theta_{k,i}^{user}) \tag{8}$$

where $\alpha_{k,i}$ is the complex gain of the ith path with $\mathbb{E}\left[|\alpha_{k,i}|\right] = 1$, ρ_k is the path loss between the BS and the kth user and the variables $\theta_{k,i}^{user} \in [0, 2\pi]$ and $\theta_{k,i}^{BS} \in [0, 2\pi]$ are the AoDs of user k and AoAs of the BS of the ith path, respectively. Regardless of the elevation, we consider the azimuth only, which implies that both BS and users conduct horizontal beamforming only. Consequently, $\mathbf{a}_{user}^H(\theta_{k,i}^{user})$ and $\mathbf{a}_{BS}(\theta_{k,i}^{BS})$ are the antenna array response vectors at the user and the BS, respectively. Here, we adopt uniform linear arrays. Thus, $\mathbf{a}_{user}^H(\theta_{k,i}^{user})$ and $\mathbf{a}_{BS}(\theta_{k,i}^{BS})$ can be written as

$$\begin{aligned} &\mathbf{a}_{user}(\theta_{k,i}^{user}) \\ &= \frac{1}{\sqrt{N_B}}[1, e^{j\frac{2\pi}{\lambda} d\sin(\theta_{k,i}^{user})}, \cdots e^{j(N_u-1)\frac{2\pi}{\lambda} d\sin(\theta_{k,i}^{user})}], \end{aligned} \tag{9}$$

$$\begin{aligned} &\mathbf{a}_{BS}(\theta_{k,i}^{BS}) \\ &= \frac{1}{\sqrt{N_B}}[1, e^{j\frac{2\pi}{\lambda} d\sin(\theta_{k,i}^{BS})}, \cdots e^{j(N_B-1)\frac{2\pi}{\lambda} d\sin(\theta_{k,i}^{BS})}], \end{aligned} \tag{10}$$

respectively, where λ is the signal wavelength and d is the distance between antenna elements. The channel model in (8) can be written in a more compact form as

$$\mathbf{H}_k = \mathbf{A}_{BS} diag(\mathbf{b}) \mathbf{A}_{user}^H \tag{11}$$

where $\mathbf{b} = \sqrt{\frac{N_B N_u}{L\rho_k}}[\alpha_{k,1}, \alpha_{k,2}, \cdots, \alpha_{k,L}]$. The matrices $\mathbf{A}_{user}$ and $\mathbf{A}_{BS}$ contain the user and the BS array response vectors, respectively, which are given by

$$\mathbf{A}_{user} = [\mathbf{a}_{user}(\theta_{k,1}^{user}), \mathbf{a}_{user}(\theta_{k,2}^{user}), \cdots \mathbf{a}_{user}(\theta_{k,L}^{user})], \tag{12}$$

$$\mathbf{A}_{BS} = [\mathbf{a}_{user}(\theta_{k,1}^{BS}), \mathbf{a}_{user}(\theta_{k,2}^{BS}), \cdots \mathbf{a}_{user}(\theta_{k,L}^{BS})]. \tag{13}$$

The channel model in (8) turns out to be Rayleigh fading channel when L is very large. Based on the channel state information, the BS selects the analog beams for both the BS and users.

Note that: Since channel estimation is beyond the scope of this study, we consider that the parameters of the L channel paths, such as AoA, AoD, and the complex gain of each path, can be estimated perfectly and known to the BS.

D. Analog beam set

We assume each user chooses a transmit analog beam from a codebook $\mathcal{F}$ which is a set consisted of $|\mathcal{F}|$ predefined analog beams. The predefined codecook of transmit analog beams can be represented as

$$\mathcal{F} = \left\{\mathbf{c}^1, \mathbf{c}^2, \cdots \mathbf{c}^{|\mathcal{F}|}\right\} \tag{14}$$

where $\mathbf{c}^i \in \mathbb{C}^{N_u \times 1}$ $(i \in \{1, 2, \cdots, |\mathcal{F}|\})$ is a possible option of the transmit analog beam for a given user. The nth entry of $\mathbf{c}^i$ is $e^{j\theta_n^i}$ which is the value of phase shift on the corresponding antenna for a given user. Similarly, the predefined codecook of receive analog beams can be represented as

$$\mathcal{G} = \left\{\mathbf{G}_1, \mathbf{G}_2, \cdots \mathbf{G}_{|\mathcal{G}|}\right\} \tag{15}$$

where $\mathbf{G}_m \in \mathbb{C}^{N_S \times N_B}$ $(m \in \{1, 2, \cdots, |\mathcal{G}|\})$ is a possible option of the receive analog beam for the BS.

If the kth user takes $\mathbf{c}^n$ as the transmit analog beam, based on (5), the equivalent uplink channel vector for the kth user can be expressed as

$$\bar{\mathbf{h}}_k^{n,m} = \mathbf{G}_m \mathbf{H}_k \mathbf{c}^n. \tag{16}$$

Here, we assume that each user shares a same predefined codecook of transmit analog beams, which is known to the BS.

E. Uplink Sum-rate

Based on (7), the sum-rate of the uplink MIMO system can be given by

$$R = \sum_{i=1}^{K} \log_2(1 + \gamma_i) \tag{17}$$

where γ_i is the signal-to-interference-plus-noise ratio (SINR) of ith user and can be written as [23] [24]

$$\gamma_i = \frac{P}{N_u N_R \sigma^2 \left[\left(\left(\bar{\mathbf{H}}\right)^H \bar{\mathbf{H}}\right)^{-1}\right]_{i,i}}. \tag{18}$$

According to (17) and (18), it is worth noting that the SINR of the ith user is dependent on the equivalent channel. Based on the definition of the the equivalent channel, we know that the equivalent channel is involved with analog beams. Therefore, each user needs to select an optimal analog beam from the predefined codecook of transmit analog beams to maximize the uplink sum-rate. Specifically, the optimization problem of analog beam selection can be formulated as

$$\left\{\bar{\mathbf{G}}, \mathbf{c}_1, \mathbf{c}_2 \cdots \mathbf{c}_K\right\} = \max_{\mathbf{G} \in \mathcal{G}, \mathbf{c}_i \in \mathcal{F}} \sum_{i=1}^{K} \log_2(1 + \gamma_i). \tag{19}$$

Intuitively, we can obtain an optimal solution for the above problem by exhaustive search, such as [12] and [14]. However, the exhaustive search makes the complexity rather high, especially when the number of antennas or the number of candidates beams is very large. Hence, it is very meaningful to develop a low-complexity method to solve this problem. In the following subsection, we discuss a sub-optimal solution for analog-beam selection.

III. Sub-optimal Solution for Analog-beam Selection

To begin with, we know that, directly calculating SINR (18) for ZF digital beamforming is involved with matrix inversion operation which is a high-complexity manipulation. Thus, we need to derive a low-complexity metric to measure sum-rate.

A. Novel metric for sum-rate

Firstly, we conduct an analysis of sum-rate under a special case where $K = 2$. And then, we would expand the analysis to general multi-user cases. According to (18), the uplink sum-rate under $K = 2$ can be given as

$$R = \sum_{i=1}^{2} \log_2 \left(1 + \frac{P \det\left(\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2\right)}{N_u N_R \sigma^2 \left\|\bar{\mathbf{h}}_{\bar{i}}\right\|^2}\right) \tag{20}$$

where $\bar{i} = \{1, 2\}/i$ and

$$\det\left(\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2\right) = \left\|\bar{\mathbf{h}}_1\right\|^2 \left\|\bar{\mathbf{h}}_2\right\|^2 - \left(\bar{\mathbf{h}}_1\right)^H \bar{\mathbf{h}}_2 \left(\bar{\mathbf{h}}_2\right)^H \bar{\mathbf{h}}_1. \tag{21}$$

Proof: See Appendix A. ∎

By resorting to some math manipulations, (20) can be rewritten as

$$R = \log_2\left(1 + P f\left(\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2\right)\right) \tag{22}$$

where

$$\begin{aligned} f\left(\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2\right) = & \frac{\det(\bar{\mathbf{h}}_1 \bar{\mathbf{h}}_2)}{N_u N_R \sigma^2 \left\|\bar{\mathbf{h}}_1\right\|^2} + \frac{\det(\bar{\mathbf{h}}_1 \bar{\mathbf{h}}_2)}{N_u N_R \sigma^2 \left\|\bar{\mathbf{h}}_2\right\|^2} \\ & + \frac{P\left(\det(\bar{\mathbf{h}}_1 \bar{\mathbf{h}}_2)\right)^2}{N_u^2 (N_R)^2 \sigma^4 \left\|\bar{\mathbf{h}}_1\right\|^2 \left\|\bar{\mathbf{h}}_2\right\|^2}. \end{aligned} \tag{23}$$

Due to the logarithmic function in (22), a positive correlation exists between the uplink sum-rate R and $f\left(\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2\right)$. Therefore, we consider the function $f\left(\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2\right)$ as an evaluating metric to select an optimal analog beam for each user, which would derive the optimum solution for maximizing the uplink sum-rate in a low-complexity way.

In order to generalize the metric to the general case where $K \geq 3$, we are able to combine $K - 1$ equivalent channel vectors into a new equivalent channel vector. Specifically, the evaluating metric can be written as $f\left(\bar{\mathbf{h}}_{\bar{k}}, \bar{\mathbf{h}}_k\right)$ where $\bar{k}$ is the complementary set of k, $\bar{\mathbf{h}}_{\bar{k}}$ is the $(K-1)$-combined equivalent channel vector and can be written as

$$\bar{\mathbf{h}}_{\bar{k}} = \sum_{i \in \bar{k}} a_i \mathbf{G}_m \mathbf{H}_i \mathbf{c}_i \tag{24}$$

with a_i being the normalized coefficient.

Thus, the multi-user case is turned into a two-user case.

B. Sub-optimal solution

According to the metric derived above, the details of the analog beam selection algorithm would be presented in the following.

1) To begin with, we fix the receive analog beam by choosing one of beams in $\mathcal{G}$ as receive analog beam. Then, we represent the set of users who have already chosen analog beams by $\mathcal{S}$, while denote the set of users who have not chosen analog beams by Ω. Then, compute the Forbinius norm of equivalent channel vectors for each combination of $|\mathcal{F}|$ transmit analog beams and K users, then we find out the pair of user and beam which can obtain the maximum Forbinius norm among $K\,|\mathcal{F}|$ combinations. This procedure can be expressed as
$$(i, \mathbf{c}_i) = arg \max_{(k \in \Omega, \mathbf{c}^n \in \mathcal{F})} \left\|\bar{\mathbf{h}}_k^{n,m}\right\|^2. \tag{25}$$
2) Compute the combined channel vector over the equivalent channel vectors of users from set $\mathcal{S}$. This procedure can be expressed as
$$\tilde{\mathbf{h}} = \sum_{i \in \mathcal{S}} a_i \mathbf{G}_m \mathbf{H}_i \mathbf{c}_i \tag{26}$$
where $a_i = \mathbf{G}_m \mathbf{H}_i \mathbf{c}_i / \sum_{i \in \mathcal{S}} \mathbf{G}_m \mathbf{H}_i \mathbf{c}_i$ is the weighted factor of the ith user whose optimal analog beam is $\mathbf{x}_i$. Similarly, this procedure can be formulated as
$$(i, \mathbf{c}_i) = arg \max_{(k \in \Omega, \mathbf{c}_n \in \mathcal{F})} f\left(\tilde{\mathbf{h}}, \bar{\mathbf{h}}_k^{n,m}\right). \tag{27}$$
3) continue the procedures above until the optimal analog beams of all users under current receive analog beam are determined.
4) repeat the procedures above until all receive analog beams in $\mathcal{G}$ are took. And take the analog beam $\mathbf{G}_m$ with maximum value of metric (23) as the optimal receive analog beam $\bar{\mathbf{G}}$.

Algorithm 1: Suboptimal algorithm

Input: $\Omega = \{1, 2, \cdots, K\}, \mathcal{S} = \emptyset, m = 1$
Output: $\mathbf{c}_k, k = 1, 2, \cdots K$
step 1: **For all** $k \in \Omega$
  For all $\mathbf{c}^n \in \mathcal{F}$
   $\bar{\mathbf{h}}_k^n = \mathbf{G}_m \mathbf{H}_k \mathbf{c}^n$
step 2: $\{i, \mathbf{c}_i\} = \arg \max_{k \in \Omega, \mathbf{c}^n \in \mathcal{F}} \left\{\left\|\bar{\mathbf{h}}_k^{n,m}\right\|^2, \forall k, \forall n\right\}$
 $\Omega = \Omega - \{i\}$, $\mathcal{S} = \mathcal{S} + \{i\}$
step 3: calculate $\tilde{\mathbf{h}} = \sum_{i \in \mathcal{S}} a_i \mathbf{G}_m \mathbf{H}_i \mathbf{c}_i$ based on (29)
step 4: **For all** $k \in \Omega$
  For all $\mathbf{c}^n \in \mathcal{F}$
   $f(\tilde{\mathbf{h}}, \bar{\mathbf{h}}_k^n)$
Step 5: $\{i, \mathbf{c}_i\} = \arg \max_{k \in \Omega, \mathbf{c}^n \in \mathcal{F}} \left\{f(\tilde{\mathbf{h}}, \bar{\mathbf{h}}_k^{n,m}), \forall k, \forall n\right\}$
 $G_m = \max_{k \in \Omega} f(\tilde{\mathbf{h}}, \mathbf{G}_m \mathbf{H}_i \mathbf{c}_i)$
 $\Omega = \Omega - \{i\}$, $\mathcal{S} = \mathcal{S} + \{i\}$
Step 6: **If** $|\Omega| = 0$
  If $m < |\mathcal{G}|$
   $m = m + 1$, go to step 1;
  Else
   $\bar{\mathbf{G}} = \arg \max_{\mathbf{G} \in \mathcal{G}} \{G_1, G_2, \cdots G_{|\mathcal{G}|}\}$
 Else
 go to Step 3.

C. Algorithm procedure

The detailed step of sub-optimization analog beam selection algorithm is illustrated in Alg.1.

One may note that, the analog beam selection method described above avoids searching over all candidates of analog beams from codebook $\mathcal{F}$. Consequently, the complexity can reduce significantly compared with exhaustive search, as will be demonstrated in section V.

IV. Data-driven analog beam selection

Although the sub-optimization method of selecting analog beam avoids the exhaustive search, it still involves some high-complexity operations, such as Forbenius norm and matrix multiplication. For reducing the complexity further, in this section, we adopt machine learning to solve this problem in a low-complexity way. To be more specific, we exploit SVM to classify the uplink channels of each user to several different types, where each type corresponds to a candidate of analog beam. SVM is a supervised machine learning algorithm, which is mostly used in classification problems. Especially, compared with other classification algorithms, SVM has advantages on both handling linearly inseparable set of samples and avoiding over-fitting since the kernel trick is adopt. In SVM algorithm, we represent each data item as a point in n-dimensional space (where n is the dimensionality of feature vectors) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the separating hyper-plane which differentiates the two classes very well. By the hyper-plane, when a new input data (current channels of users) comes up, we can predict the class (optimal analog beam) of the new input data.

A. Training Samples Set

1) Generating Training Samples: In a supervised machine learning, training data is indispensable for obtaining the classifying criterion. Here, we assume that M channel samples are generated for training. Based on the model of mm-wave channel, each channel sample can be presented by $4L + 1$ real-value features including the path loss, L complex gain of paths (including $2L$ real-value features), L angles of departure of users and L angles of arrival of the BS. To guarantee the effectiveness of training, the features of each sample should be randomly generated based on their corresponding statistical character. Furthermore, since high-value feature would bring about bias, we need to normalize each feature as
$$\bar{e}_m^l = \frac{e_m^l - Mean(e_m^l)}{e_{\max}^l - e_{\min}^l} \tag{28}$$
where e_m^l is the value of the lth feature of the mth sample, $Mean(e_m^l)$ represents the mean of the lth feature of M samples, $e_{\max}^l$ donates the maximum value of the lth feature among M samples, while $e_{\min}^l$ represents the minimum value of the lth feature among M samples.

Then each channel sample can be represented as a feature vector $\mathbf{t}_m \in \mathbb{R}^{1 \times (4L+1)}$ consisted of $4L + 1$ normalized features.

2) KPI Function: KPI is used to evaluate the objective metric, such as BER, SNR and SINR. Based on the analysis in the above section, for problem (25), we consider $\left\|\mathbf{G}_m\mathbf{H}_k\mathbf{c}^i\right\|^2$ as the KPI.

3) Labeling: There are $|\mathcal{F}|$ choices of analog beams for each user, hence we evaluate the KPI for all possible combinations of a given sample and all choices of analog beams. And we label the training simple with $\mathbf{c}_n$ which is able to let this training sample obtain the maximum KPI. The label of all training samples can be presented as vector $\mathbf{r} \in \mathbb{R}^{1\times M}$ consisted of the index of optimal analog beam for M training samples.

B. Regular SVM based classifier

For each feature vector $\mathbf{t}_m$ ($m \in \{1, 2, \cdots, M\}$), we have its corresponding class-label $\mathbf{r}[m]$. By using the M labeled training samples, we are able to develop a multi-class classifier where the input is the channel feature vector and the output is the optimal analog beam which can maximize KPI for the input channel. Generally, in hybrid beamforming, a codebook of analog beams provides more than two candidates for beams selection. Hence, in order to classify $|\mathcal{F}|$ classes, we exploit $|\mathcal{F}|$ one-vs-the-rest SVM classifiers, each of which classifies a channel feature vector into one category or the other categories. Let us take the nth ($n \in \{1, 2, \cdots, |\mathcal{F}|\}$) classifier as an example, where we classify the sample labeled by n into one category but classify the other samples into another category. For the ith sample, we set $y_i = +1$ if the label $\mathbf{r}[i] = n$, while set $y_i = -1$ if $\mathbf{r}[i] \neq n$. And $\mathbf{w}$ is the vector consisting of parameters for the separating hyper-plane. In SVM algorithm, the optimization problem of training the separating hyper-plane of the nth classifier can be formulated as

$$\begin{aligned} &\min_{\mathbf{w},b,\zeta} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{M}\zeta_i \\ s.t.\quad & y_i\left(\mathbf{w}^T\phi(\mathbf{t}_i) + b\right) \geq 1 - \zeta_i, i = 1, 2, \cdots, M \\ & \zeta_i \geq 0, i = 1, 2, \cdots, M \end{aligned} \tag{29}$$

where ϕ is the mapping function by which the sample data $\mathbf{t}_i$ can be mapped into high-dimensional space, b is the threshold, C is the penalty constant, ζ_i is the value of error caused by misclassification for sample $\mathbf{t}_i$.

However, in our problem, the number of samples labeled by $\mathbf{c}^n$ is much smaller than that of other categories. When faced with imbalanced datasets where the number of negative instances far outnumbers the positive instances, the performance of regular SVM drops significantly. A popular approach towards solving these problems is to preprocess the data by oversampling the majority class or undersampling the minority class in order to create a balanced dataset. However, by this approach, the data structure is destroyed, which results in a inaccurate separating hyper-plane.

C. Biased-SVM based classifier

In this paper, for solving these problems, we propose an approach which pays more attention to the positive instances. This can be done, for instance, by increasing the penalty associated with misclassifying the positive class relative to the negative class. Specifically, we choose two different plenty constants for positive samples and negative samples, respectively. The optimization can be reformulated as

$$\begin{aligned} &\min_{\mathbf{w},b,\zeta} \frac{1}{2}\mathbf{w}^T\mathbf{w} + C_+\sum_{\{i|y_i=+1\}}\zeta_i + C_-\sum_{\{i|y_i=-1\}}\zeta_i \\ s.t.\quad & y_i\left(\mathbf{w}^T\phi(\mathbf{t}_i) + b\right) \geq 1 - \zeta_i, i = 1, 2, \cdots, M \\ & \zeta_i \geq 0, i = 1, 2, \cdots, M \end{aligned} \tag{30}$$

where C_+ and C_- are the plenty constants for positive samples and negative samples, respectively.

The Lagrange duality problem of (30) can be written as

$$\begin{aligned} &\min_{\mathbf{w},b,\zeta} \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} a_i a_j y_i y_j K(\mathbf{t}_i, \mathbf{t}_j) - \sum_{i=1}^{M} a_i \\ s.t.\quad & \sum_{i=1}^{N} y_i a_i = 0, \\ & 0 \leq a_i \leq C_+, \quad y_i = +1, \\ & 0 \leq a_j \leq C_-, \quad y_j = -1, \end{aligned} \tag{31}$$

where a_i and a_j are Lagrange multipliers, $K(\mathbf{t}_i, \mathbf{t}_j) = \langle\phi(\mathbf{t}_i), \phi(\mathbf{t}_j)\rangle$ is the Gaussian radial-based kernel function and can be further written as

$$K(\mathbf{t}_i, \mathbf{t}_j) = e^{\left(-\|\mathbf{t}_i - \mathbf{t}_j\|^2/(2\sigma^2)\right)} \tag{32}$$

with σ ($\breve{\sigma} \leq \sigma \leq \widehat{\sigma}$) being the design parameter.

In regular SVM, the optimal set of Lagrange multipliers of optimization problem can be solved by sequential minimal optimization algorithm (SMO) [25] with the fast and reliable convergence. SMO is an iterative algorithm for solving the optimization problem described in (31). SMO can break this problem into a series of smallest possible sub-problems, which are then solved analytically. However, in the proposed SVM, since the constraint conditions of Lagrange multipliers are different with the regular SVM, we need to analyze new constraint conditions of Lagrange multipliers for SMO algorithm. After each iteration of SMO algorithm, the new Lagrange multipliers must be within the constraint area. For example, in the case where $y_i \neq y_j, C_i > C_j, a_i^{old} > a_j^{old}, \left(a_i^{old} - a_j^{old}\right) < (C_i - C_j)$, based on the relationship $a_i^{old}y_i + a_j^{old}y_j = a_i^{new}y_i + a_j^{new}y_j$, we have:

$$\begin{cases} if\ a_j < 0, & \begin{cases} a_i^{new} = a_i^{old} - a_j^{old} \\ a_j^{new} = 0 \end{cases} \\ if\ 0 < a_j < C_j, & \begin{cases} a_i^{new} = a_i \\ a_j^{new} = a_j \end{cases} \\ if\ a_j > C_j, & \begin{cases} a_i^{new} = C_j + a_i^{old} - a_j^{old} \\ a_j^{new} = C_j \end{cases} \end{cases}. \tag{33}$$

D. Parameter Optimization

As shown above, the classification performance of the proposed SVM method primarily depends on the parameter σ of kernel functions and the penalty constant C_+ and C_-. In this subsection, we discuss how to determine the parameters to achieve an optimal classification performance.

1) Parameter σ: Because SVM is a kind of machine learning method based on the kernel function, the selection of the corresponding parameter σ would bring about great

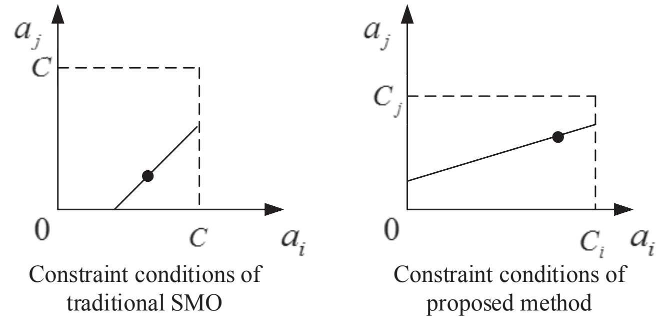

Fig. 2. Constraint Conditions of Lagrange Multipliers

influences on the generalization performance of SVM. At present, there has been a few researches on the parameter selection for a given kernel. Among them, cross-validation [26] is considered to be more precise, however it needs to train SVM for many times, which results in high-complexity tasks.

In this paper, we propose a novel method of determining the optimal σ in according to spatial distance. For obtaining high accuracy of classification results, we hope the spatial distance between samples of same class is smaller, while the spatial distance between samples of different classes is farther. Thus, we are able to determine the optimal σ based on this criterion. Specifically, the problem of determining the optimal σ can be written as

$$\begin{cases} \min\limits_{\sigma} \|\phi(\mathbf{t}_i) - \phi(\mathbf{t}_j)\|^2, y_i y_j = 1 \\ \max\limits_{\sigma} \|\phi(\mathbf{t}_i) - \phi(\mathbf{t}_j)\|^2, y_i y_j = -1. \end{cases} \tag{34}$$

Based on the fact $\|\phi(\mathbf{t}_i) - \phi(\mathbf{t}_j)\|^2 = 2 - 2K(\mathbf{t}_i, \mathbf{t}_j)$ whose detailed proof is given in Appendix B, the problem (34) can be rewritten as

$$y_i y_j ||\phi(\mathbf{t}_i) - \phi(\mathbf{t}_j)||^2 = \begin{cases} \min\limits_{\sigma} \left[2 - 2K\left(\phi(\mathbf{t}_i), \phi(\mathbf{t}_j)\right)\right], \ y_i y_j = 1 \\ \max\limits_{\sigma} \left[-2 + 2K\left(\phi(\mathbf{t}_i), \phi(\mathbf{t}_j)\right)\right], \ y_i y_j = -1. \end{cases}$$

Therefore, the problem (34) can be reformulated as

$$\max_{\sigma} \sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j K(\mathbf{t}_i, \mathbf{t}_j) = \max_{\sigma} \sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j e^{\tau \varepsilon_{ij}} \tag{35}$$

where $\varepsilon_{ij} = \|\mathbf{t}_i - \mathbf{t}_j\|^2$ and $\tau = -1/2\sigma^2$. Now, by using Mclaughlin expansion [27], $\sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j e^{\tau \varepsilon_{ij}}$ in (35) can be represented as

$$\begin{aligned} \sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j e^{\tau \varepsilon_{ij}} &= \sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j (1 + \tau \varepsilon_{ij} + \frac{1}{2} \tau^2 \varepsilon_{ij}^2) \\ &= \sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j + \tau \sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j \tau \varepsilon_{ij} + \tau^2 \sum_{i=1}^{M} \sum_{j=1}^{i-1} \frac{1}{2} y_i y_j \varepsilon_{ij}^2. \end{aligned} \tag{36}$$

When $\sum\limits_{i=1}^{M} \sum\limits_{j=1}^{i-1} \frac{1}{2} y_i y_j \varepsilon_{ij}^2 < 0$, the optimal τ^* can be determined as

$$\tau^* = -\sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j \tau \varepsilon_{ij} \Bigg/ \sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j \varepsilon_{ij}^2. \tag{37}$$

When $\sum\limits_{i=1}^{M} \sum\limits_{j=1}^{i-1} \frac{1}{2} y_i y_j \varepsilon_{ij}^2 > 0$, the optimal τ^* can be determined as

$$\tau^* = \max_{\tau = \{\breve{\tau}, \widehat{\tau}\}} L(\tau) \tag{38}$$

where $L \triangleq \sum_{i=1}^{M} \sum_{j=1}^{i-1} y_i y_j e^{\tau \varepsilon_{ij}}$.

2) Parameter C_+ and C_-: As mentioned in section IV, for the accuracy of the separating hyperplane in imbalanced datasets, we choose the larger penalty constant for positive samples, while we choose the smaller penalty constant for negative samples. In this way, as shown in the optimization problem (30), the misclassification for the fewer positive class would result in larger penalty, thereby improving the accuracy of the SVM classifier. Therefore, based on this criterion, we adopt the reciprocal of the number of positive samples and negative samples as the C_+ and C_-, respectively.

E. Classifying Stage

When $\mathbf{w}$, b and design parameters are determined, the classifier of analog beam $\mathbf{c}^n \in |\mathcal{F}|$ can be presented as

$$g_1^n\left(\bar{\mathbf{t}}_k\right) = \sum_{\mathbf{s}_i \in \mathcal{V}_1} a_i y_i K(\mathbf{s}_i, \bar{\mathbf{t}}_k) + b \tag{39}$$

where $\bar{\mathbf{t}}_k$ is a new feature vector needed to be classified, $\mathbf{s}_i$ is a support vector, $\mathcal{V}_1$ is the set of support vectors.

If $g_1^n\left(\bar{\mathbf{t}}_k\right) > 0$, we consider $\|\mathbf{G}\mathbf{H}_k\mathbf{c}^n\|^2$ outperforms $\left\|\mathbf{G}\mathbf{H}_k\mathbf{c}^i\right\|^2, i \neq n$. Thus the optimal beam for user k is $\mathbf{c}^n$.

F. Optimization problem (27)

1) Generating Training Samples: In the problem (27), the aim is to find the optimal combination of k and $\mathbf{c}^n$ in terms of maximizing the objective function $f\left(\tilde{\mathbf{h}}, \bar{\mathbf{h}}_k^{n,m}\right)$ whose input vectors are $\tilde{\mathbf{h}}$ and $\bar{\mathbf{t}}_k$. Basically, the combined channel vector $\tilde{\mathbf{h}}$ is also an equivalent channel vector, so it is justified to take the training samples of equivalent channel vector as the training samples of $\tilde{\mathbf{h}}$. Since there are M training samples for the mm-wave channel $\mathbf{H}$ and $|\mathcal{F}|$ candidate analog beams, we are able to generate $M\,|\mathcal{F}|$ training samples for the combined channel vectors $\tilde{\mathbf{h}}$ based on (16). Since another input is feature vector $\bar{\mathbf{t}}_k$ which has M training samples, thus there are $M^2\,|\mathcal{F}|$ training samples for the input of (27) by combining $\tilde{\mathbf{h}}$ and $\bar{\mathbf{t}}_k$. And each training sample is presented as $(N_s + 4L + 1)$-dimension feature vector. Similarly, each feature entry should be normalized in case of bias.

2) KPI Function and Labeling: For the problem (27), we set $f\left(\mathbf{a}, \mathbf{b}\right)$ in (23) as KPI, and each training sample is labeled with the reference number of the analog beam which is able to obtain maximum KPI. Thus, the $M^2\,|\mathcal{F}|$ samples are divided into $|\mathcal{F}|$ classes. Based on the training of SVM, the separating hyperplane for each class is obtained. Therefore, when a new feature vector comes up, we are able to derive the optimal analog beam with which the maximum $f\left(\mathbf{a}, \mathbf{b}\right)$ can be obtained.

3) Classifying Stage: Similarly, the classifier of analog beam $\mathbf{c}_n$ can be represented as

$$g_2^n\left(\left[\tilde{\mathbf{h}}, \bar{\mathbf{t}}_k\right]\right) = \sum_{\mathbf{s}_i \in \mathcal{V}_2} a_i y_i K\left(\mathbf{s}_i, \left[\tilde{\mathbf{h}}, \bar{\mathbf{t}}_k\right]\right) + b \tag{40}$$

where $\bar{\mathbf{t}}_k$ is a new feature vector needed to be classified, $\left[\tilde{\mathbf{h}}, \bar{\mathbf{t}}_k\right]$ is a vector combined $\tilde{\mathbf{h}}$ and $\bar{\mathbf{t}}_k$, $\mathbf{s}_i$ is a support vector, $\mathcal{V}_2$ is the set of support vectors.

If $g_2^n\left(\left[\tilde{\mathbf{h}}, \bar{\mathbf{t}}_k\right]\right) > 0$, the current analog beam $\mathbf{c}^n$ can maximize the KPI function $f\left(\tilde{\mathbf{h}}, \bar{\mathbf{h}}_k^{n,m}\right)$ for the new feature vector $\bar{\mathbf{t}}_k$.

G. Algorithm procedure

The detailed step of data-driven analog beam selection algorithm is illustrated in Alg.2.

Algorithm 2: Data-driven analog beam selection algorithm

Input: $\Omega = \{1, 2, \cdots, K\}, \mathcal{S} = \emptyset$,
$\bar{\mathbf{t}}_k$, $k \in \Omega$, $m = 1$,
$\mathbf{w}_{c_n}^1, b_{c_n}^1, \mathbf{w}_{c_n}^2, b_{c_n}^2$, a_i

Output: $\mathbf{c}_k, k = 1, 2, \cdots K$

Step 1: **For all** $k \in \Omega$
  For all $\mathbf{c}^n \in \mathcal{F}$
    If $g_1^n(\bar{\mathbf{t}}_k) > 0$
      $\mathbf{c}_k = \mathbf{c}^n$, $k = k + 1$, end;
    Else
      $n = n + 1$.

Step 2: $i = \arg\max\limits_{k\in\Omega}\left\{\left\|\bar{\mathbf{h}}_k^{n,m}\right\|^2, \forall k\right\}$
  $\Omega = \Omega - \{i\}$, $\mathcal{S} = \mathcal{S} + \{i\}$

Step 3: calculate $\tilde{\mathbf{h}} = \sum\limits_{i\in\mathcal{S}} a_j \mathbf{G}_m \mathbf{H}_i \mathbf{c}_i$ based on (25)

Step 4: **For all** $\left[\tilde{\mathbf{h}}, \bar{\mathbf{t}}_k\right]$, $k \in \Omega$
  For all $\mathbf{c}^n \in \mathcal{F}$
    If $g_2^n\left(\left[\tilde{\mathbf{h}}, \bar{\mathbf{t}}_k\right]\right) > 0$
      $\mathbf{c}_k = \mathbf{c}^n$, $k = k + 1$, end;
    Else
      $n = n + 1$.

Step 5: $i = \arg\max\limits_{k\in\Omega}\left\{f(\tilde{\mathbf{h}}, \bar{\mathbf{h}}_k^{n,m}), \forall k\right\}$
  $G_m = \max\limits_{k\in\Omega}\{f(\tilde{\mathbf{h}}, \mathbf{G}_m\mathbf{H}_i\mathbf{c}_i)\}$
  $\Omega = \Omega - \{i\}$, $\mathcal{S} = \mathcal{S} + \{i\}$

Step 6: **If** $|\Omega| = 0$
    If $m < |\mathcal{G}|$
      $m = m + 1$, go to Step 1;
    Else
      $\bar{\mathbf{G}} = \arg\max\limits_{\mathbf{G}\in\mathcal{G}}\{G_1, G_2, \cdots G_{|\mathcal{G}|}\}$
  Else
  go to Step 3.

V. Complexity Analysis

In this section, we analyze the complexity of the exhaustive search, sub-optimization method and data-driven method for analog beam selection.

A. Exhaustive search

To begin with, we know that complexity of calculating inversion of a matrix $\mathbf{X} \in \mathbb{C}^{t\times t}$ is $\mathcal{O}(t^3)$ [28]. Consequently, the complexity of exhaustive search is

$$O\left(|G|\,|F|^K K^5 N_S^4 N_u^2 N_R\right). \tag{41}$$

B. Sub-optimization method

Firstly, we know that the complexity of calculating Forbenius norm of a vector $\mathbf{x} \in \mathbb{C}^{t\times 1}$ is $\mathcal{O}(t)$ [28]. Since there are K users and each user has $|\mathcal{F}|$ possible analog beams, we need to consider $K\,|\mathcal{F}|$ equivalent channels and to find the first user and its corresponding optimal analog beam which provides the maximum vector norm (25) among $K\,|\mathcal{F}|$ equivalent channel vectors. And then the complexity of calculating $\bar{\mathbf{h}}_k = \mathbf{G}\mathbf{H}_k\mathbf{t}_k$ is $N_S^3 N_u^2 N_R$. Consequently, the complexity of (25) is $N_S^3 N_u^2 N_R N_S K\,|\mathcal{F}|$.

Since the analog beams for the other $K-1$ users are selected by using the metric in (23) which is involved with two operations of Forbenius norm for vector and two operations of inner product for vectors, the complexity of selecting the optimal analog beams for the other $K-1$ users is

$$\begin{aligned}
&\mathcal{O}\left((K-1)\,|\mathcal{F}|\,N_S^3 N_u^2 N_R\,(2N_S)\right)\\
&+\mathcal{O}\left((K-2)\,|\mathcal{F}|\,N_S^3 N_u^2 N_R\,(2N_S)\right)\\
&\cdots+\mathcal{O}\left(|\mathcal{F}|\,N_S^3 N_u^2 N_R\,(2N_S)\right)\\
&=\mathcal{O}\left(\frac{(K-1)\,K}{2}\,|\mathcal{F}|\,N_S^3 N_u^2 N_R\,(2N_S)\right).
\end{aligned} \tag{42}$$

Therefore, the complexity for the sub-optimization analog beam selection algorithm can be estimated as

$$\mathcal{O}\left(\begin{array}{c} N_S^3 N_u^2 N_R N_S K\,|\mathcal{F}|\,|\mathcal{G}| \\ +\frac{(K-1)K}{2}\,|\mathcal{F}|\left(N_S^3 N_u^2 N_R 2N_S\,|\mathcal{G}|\right)\end{array}\right). \tag{43}$$

C. Data-driven method

In the data-driven method, the algorithm complexity should exclude the training complexity of SVM, due to that the training stage is performed offline. Hence, only complexity of classifying should be taken into account. Since the kernel function requires a Forbenius norm of a feature vector, this needs $\mathcal{O}(N_S^3 N_u^2 N_R K + |\mathcal{V}_1|\,(4L+1)K\,|\mathcal{F}|)$.

Since the analog beams for the other $K-1$ users are selected by using the metric in (23), the complexity of selecting the analog beams for the other $K-1$ users is

$$\begin{aligned}
&\mathcal{O}\left(\begin{array}{c}(K-1)\,|\mathcal{F}|\,|\mathcal{V}_2|\,((N_S+4L+1))\\ +(K-1)\,N_S^3 N_u^2 N_R\end{array}\right)\\
&+\mathcal{O}\left(\begin{array}{c}(K-2)\,|\mathcal{F}|\,|\mathcal{V}_2|\,((N_S+4L+1))\\ +(K-2)\,N_S^3 N_u^2 N_R\end{array}\right)\cdots\\
&+\mathcal{O}\left(|\mathcal{F}|\,|\mathcal{V}_2|\,((N_S+4L+1))+N_S^3 N_u^2 N_R\right)\\
&=\mathcal{O}\left(\begin{array}{c}\frac{(K-1)K}{2}\,|\mathcal{F}|\,|\mathcal{V}_2|\,((N_S+4L+1))\\ +\frac{(K-1)K}{2}N_S^3 N_u^2 N_R\end{array}\right).
\end{aligned} \tag{44}$$

Therefore, the complexity for the data-driven method can be estimated as

$$\mathcal{O}\left(\begin{array}{c}\left(N_S^3 N_u^2 N_R K + |\mathcal{V}_1|\,(4L+1)K\,|\mathcal{F}|\right)|\mathcal{G}|\\ +\frac{(K-1)K}{2}\left(|\mathcal{F}|\,|\mathcal{V}_2|\,(N_S+4L+1)\right)|\mathcal{G}|\\ +\frac{(K-1)K}{2}\left(N_S^3 N_u^2 N_R\right)|\mathcal{G}|\end{array}\right). \tag{45}$$

Since the poor scattering nature of the mm-wave channel, the number of scattering path L is rather few. On the other side, the number of support vectors, i.e., $|\mathcal{V}_1|$, $|\mathcal{V}_2|$, is very few. Thus, the complexity of data-driven method (45) can reduce dramatically, compared with sub-optimization method (43).

Remark 1. *Similar to [29], the algorithm complexity of the data-driven method should exclude the training complexity,* *since the training stage is performed offline. Hence, only classifying complexity is taken into account. Besides, only when the statistical characters (such as the probability distribution of DOA, AOA and complex gain of each path) of channels change, we have to take a new training stage to obtain the classifying model for the new channel conditions.*

VI. Simulations

In this section, numerical results are presented to verify the proposed data-driven analog beam selection method. We model path loss of k-user as $\rho_k = D_k^{-\beta/2}$ where D_k is the distance between the BS and the kth user and β is the path loss exponent. Here, we set $\beta = 3.76$. The number of users in each cell K is set to 10. For simplicity, the distance between each user and the BS D_k is a variable uniformly distributed with the interval $[10, 15]$. Besides, we set $N_s = K$, $N_R = 5$ and $N_u = 5$. For the mm-wave channel, we set the number of scattering path $L = 4$ and the azimuth angles of departure or arrival of user and the BS are uniformly distributed between 0 and 2π, signal wavelength λ=5 mm and the antenna spacing distance $\lambda/2$. We assume there are 5 candidates of transmit analog beams, which can be represented as

$$\mathcal{F} = \left\{\mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3, \mathbf{c}^4, \mathbf{c}^5\right\} \tag{46}$$

where codeword $\mathbf{c}^n$ can be expressed as

$$\mathbf{c}^n = \frac{1}{\sqrt{N_u}}\left[1, e^{-j2\pi 1 n^*}, e^{-j2\pi 2 n^*}, \cdots, e^{-j2\pi(N_u-1)n^*}\right]^T \tag{47}$$

with $n^* = \frac{n-1}{|\mathcal{F}|}$. Besides, we assume there are three candidates of receive analog beams in $\mathcal{G}$, and each codeword is also structured by the rule (47).

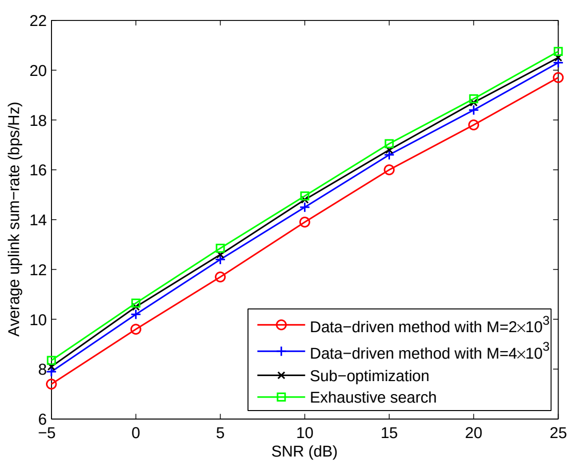

Fig. 3. Average uplink sum-rate versus SNR

Fig. 3 shows the average uplink sum-rate under different SNR, where the average sum-rate is obtained over 10000 channel realizations. One may note that the sum-rate of data-driven analog beam selection method with $M = 4\times10^3$ is very close to the sub-optimization method. However, as shown in Fig.3, the data-driven method with $M = 2\times10^3$ brings about an obvious degression of uplink sum-rate.

Fig. 4 shows the average uplink sum-rate versus number of training samples. One may note that the sum-rate of data-driven analog beam selection method tends to close to the sum-rate of sub-optimization method as the number of training

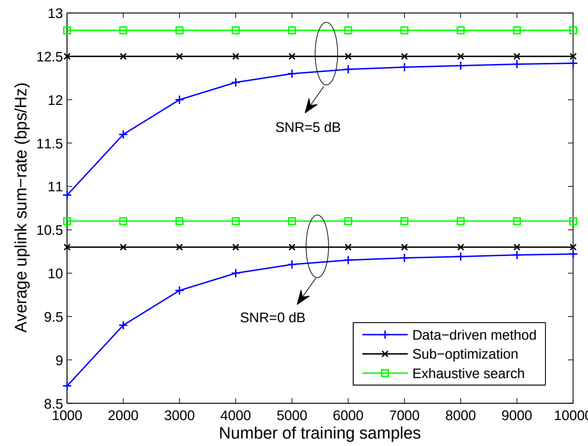

Fig. 4. Average uplink sum-rate versus the number of training samples

samples increases. For example, when the number of training samples is 1000, the sum-rate of data-driven method is 15% lower of the counterpart of the sub-optimization method, which is unacceptable. Nevertheless, when the number of training samples is 10000, the sum-rate of data-driven method is almost the same with the counterpart of the sub-optimization method. Therefore, Fig. 3 and Fig. 4 reveal that the data-driven method is highly dependent on the number of training data. As long as the the number of training data is large enough, the data-driven method is able to provide a near-optimal performance.

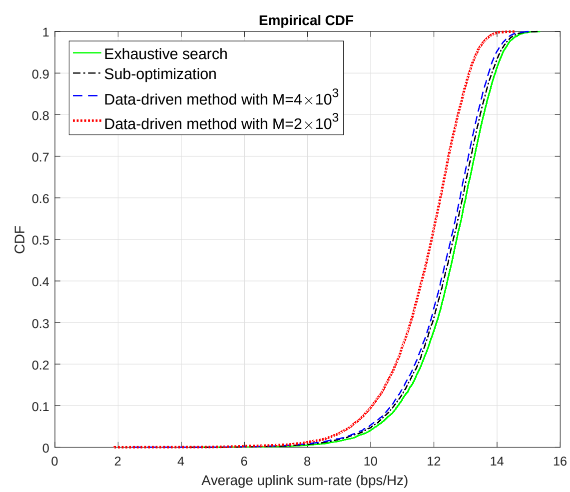

Fig. 5. Effectiveness of data-driven method

The effectiveness of the data-driven analog beam selection method under SNR=5 dB is verified in Fig.5 where the cumulative distribution functions (CDFs) over 10000 times realization of uplink channel are shown. The result for the exhaustive search is provided as a base line. One may notice that the CDF curve of data-driven analog beam selection method with $M = 4 \times 10^3$ approaches the curve of the sub-optimization method and exhaustive search, while the CDF curve of data-driven method with $M = 2\times10^3$ lags far behind that of the data-driven method with $M = 4 \times 10^3$.

The effectiveness of the proposed SVM solution under SNR=5 dB is verified in Fig.6 where the cumulative distribution functions (CDFs) over 10000 times realization of uplink

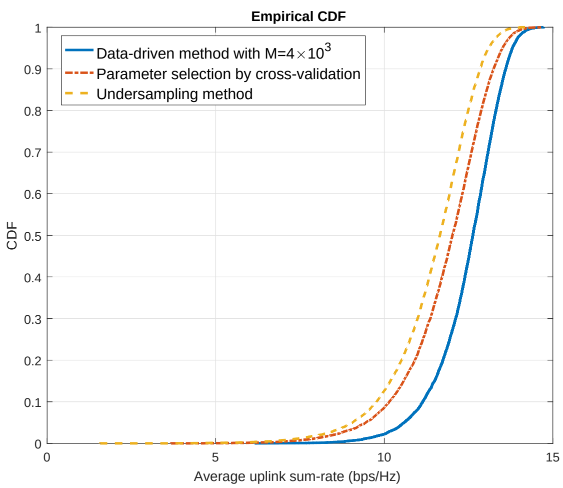

Fig. 6. Effectiveness of proposed parameter optimization

channel are shown. The CDF curves of parameter selection by cross-validation and undersampling method are provided for comparison. Especially, we adopt 10-fold method for cross-validation, the original samples of channel are randomly split into 10 same-size subsamples. Each one of the 10 subsamples is taken as the testing data for validating the obtained model, and the remaining 9 subsamples are taken as training data. The same process is then repeated 10 times, with each of the 10 subsamples used exactly once as the testing data. The 10 results of 10 repeats can then be averaged to obtain a single estimation. As shown in Fig.6, we can note that the CDF curve of data-driven analog beam selection method apparently outperforms the curves of parameter selection by cross-validation and undersampling method. Therefore, specific to the analog beam selection, the proposed biased-SVM method and parameter optimization method is superior to traditional undersampling method and cross-validation method, respectively.

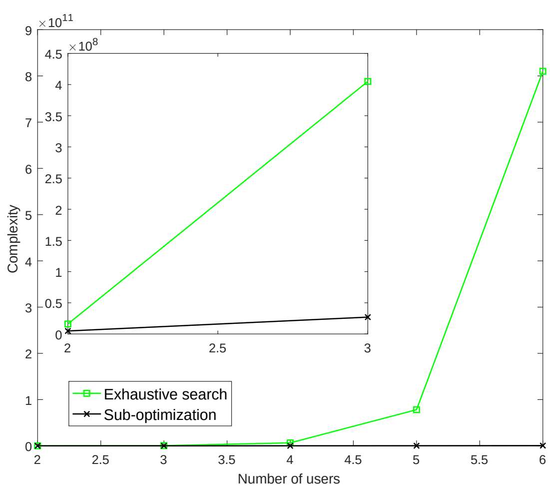

Fig. 7. Complexity of exhaustive search and sub-optimization

In Fig. 7 and Fig. 8, the complexity is compared. As shown in Fig. 8, compared with sub-optimization method, the complexity of proposed data-driven method reduces dramatically. For example, when the number of users is 2, the complexity of data-driven method is only one-tenth of the complexity of sub-optimization method. Moreover, one may note that, as the number of users increases, the complexity of sub-optimization method grows dramatically, while the growth of the complexity of data-driven method is very little. Thus, when the number of users grows large, the advantage of the data-driven method on complexity tends to be more obvious.

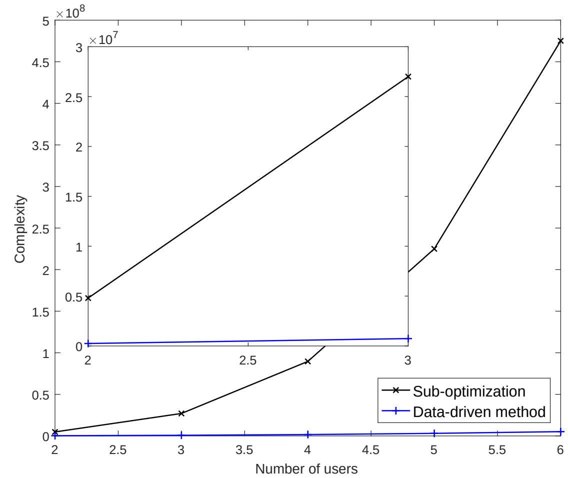

Fig. 8. Complexity of sub-optimization and data-driven method

VII. CONCLUSION

In hybrid MIMO systems, beam selection for analog beamforming is essential to achieve an optimal performance. In this paper, based on machine learning, we proposed a data-driven method of analog beam selection, which is highly dependent on training data. Specifically, we considered the beam selection problem as a multiclass-classification problem, and we exploited SVM to solve this problem. Normally, a codebook provides more than two candidates for beams selection, which brings about the imbalance of training data. However, regular SVM solves the problem with imbalanced training data not ideally. For overcoming it, we proposed a biased-SVM to solve the classification problem with imbalanced training data. Based on the biased-SVM, we obtained a statistical classification model in terms of maximizing sum-rate. During the real-time transmission, by using the derived classification model, we are able to select the optimal analog beam for each user with low complexity. Besides, we proposed a novel method to determine the optimal design parameter by McLaughlin expansion of the Gaussian kernel function. As long as the training data is large enough, the proposed method is able to achieve the same sum-rate with conventional method, while the complexity would reduce dramatically. Moreover, when the number of users grows large, the advantage of the data-driven method on complexity tends to be more obvious.

APPENDIX A
THE PROOF OF EQUATION (21)

Proof: Based on the property of inverse matrix, we have

$$\left(\left(\bar{\mathbf{H}}\right)^H \bar{\mathbf{H}} \right)^{-1} = \frac{\operatorname{adj}\left(\left(\bar{\mathbf{H}}\right)^H \bar{\mathbf{H}} \right)}{\det \left| \left(\bar{\mathbf{H}}\right)^H \bar{\mathbf{H}} \right|} \tag{48}$$

where $\mathrm{adj}\left(\left(\bar{\mathbf{H}}\right)^H\bar{\mathbf{H}}\right)$ represents the adjugate of matrix $\left(\bar{\mathbf{H}}\right)^H\bar{\mathbf{H}}$.

To begin with, $\mathrm{adj}\left(\left(\bar{\mathbf{H}}\right)^H\bar{\mathbf{H}}\right)$ can be written as

$$\mathrm{adj}\left(\left(\bar{\mathbf{H}}\right)^H\bar{\mathbf{H}}\right) = \begin{bmatrix} \left\|\bar{\mathbf{h}}_2\right\|^2 & -\left(\bar{\mathbf{h}}_1\right)^H\bar{\mathbf{h}}_2 \\ -\left(\bar{\mathbf{h}}_2\right)^H\bar{\mathbf{h}}_1 & \left\|\bar{\mathbf{h}}_1\right\|^2 \end{bmatrix}. \quad (49)$$

Then, $\det\left|\left(\bar{\mathbf{H}}\right)^H\bar{\mathbf{H}}\right|$ can be written as

$$\det\left|\left(\bar{\mathbf{H}}\right)^H\bar{\mathbf{H}}\right| = \left\|\bar{\mathbf{h}}_1\right\|^2\left\|\bar{\mathbf{h}}_2\right\|^2 - \left(\bar{\mathbf{h}}_1\right)^H\bar{\mathbf{h}}_2\left(\bar{\mathbf{h}}_2\right)^H\bar{\mathbf{h}}_1. \quad (50)$$

Substituting (49) and (50) into (48), we arrive at

$$\begin{aligned}\left(\left(\bar{\mathbf{H}}\right)^H\bar{\mathbf{H}}\right)^{-1} =& \frac{1}{\left\|\bar{\mathbf{h}}_1\right\|^2\left\|\bar{\mathbf{h}}_2\right\|^2 - \left(\bar{\mathbf{h}}_1\right)^H\bar{\mathbf{h}}_2\left(\bar{\mathbf{h}}_2\right)^H\bar{\mathbf{h}}_1} \\ &\times \begin{bmatrix} \left\|\bar{\mathbf{h}}_2\right\|^2 & -\left(\bar{\mathbf{h}}_1\right)^H\bar{\mathbf{h}}_2 \\ -\left(\bar{\mathbf{h}}_2\right)^H\bar{\mathbf{h}}_1 & \left\|\bar{\mathbf{h}}_1\right\|^2 \end{bmatrix}.\end{aligned} \quad (51)$$

According to (18), the SINR of user 1 and 2 can be written as

$$\begin{aligned}\gamma_1 &= \frac{P}{N_u N_R \sigma^2\left[\left(\left(\bar{\mathbf{H}}\right)^H\bar{\mathbf{H}}\right)^{-1}\right]_{1,1}} \\ &= \frac{P\left(\left\|\bar{\mathbf{h}}_1\right\|^2\left\|\bar{\mathbf{h}}_2\right\|^2 - \left(\bar{\mathbf{h}}_1\right)^H\bar{\mathbf{h}}_2\left(\bar{\mathbf{h}}_2\right)^H\bar{\mathbf{h}}_1\right)}{N_u N_R \sigma^2\left\|\bar{\mathbf{h}}_2\right\|^2}, \\ \gamma_2 &= \frac{P}{N_u N_R \sigma^2\left[\left(\left(\bar{\mathbf{H}}\right)^H\bar{\mathbf{H}}\right)^{-1}\right]_{2,2}} \\ &= \frac{P\left(\left\|\bar{\mathbf{h}}_1\right\|^2\left\|\bar{\mathbf{h}}_2\right\|^2 - \left(\bar{\mathbf{h}}_1\right)^H\bar{\mathbf{h}}_2\left(\bar{\mathbf{h}}_2\right)^H\bar{\mathbf{h}}_1\right)}{N_u N_R \sigma^2\left\|\bar{\mathbf{h}}_1\right\|^2},\end{aligned} \quad (52)$$

respectively.

The proof is completed by substituting (52) into (7). ■

Appendix B
Equation (21)

Proof: Firstly, based on the definition of Forbinius norm, we have,

$$\begin{aligned}&||\phi(\mathbf{t}_i) - \phi(\mathbf{t}_j)||^2 \\ &= \langle\phi(\mathbf{t}_i), \phi(\mathbf{t}_i)\rangle + \langle\phi(\mathbf{t}_j), \phi(\mathbf{t}_j)\rangle - 2\langle\phi(\mathbf{t}_i), \phi(\mathbf{t}_j)\rangle.\end{aligned} \quad (53)$$

By resorting to the fact $K(\mathbf{t}_i, \mathbf{t}_j) = \langle\phi(\mathbf{t}_i), \phi(\mathbf{t}_j)\rangle$, the above formula can be further written as

$$\begin{aligned}&||\phi(\mathbf{t}_i) - \phi(\mathbf{t}_j)||^2 \\ &= K\left(\phi(\mathbf{t}_i), \phi(\mathbf{t}_i)\right) + K\left(\phi(\mathbf{t}_j), \phi(\mathbf{t}_j)\right) - 2K\left(\phi(\mathbf{t}_i), \phi(\mathbf{t}_j)\right).\end{aligned} \quad (54)$$

Based on Gaussian radial-based kernel function, i.e., $K(\mathbf{t}_i, \mathbf{t}_j) = e^{\left(-\|\mathbf{t}_i - \mathbf{t}_j\|^2/(2\sigma^2)\right)}$, we have

$$K\left(\phi(\mathbf{t}_i), \phi(\mathbf{t}_i)\right) = K\left(\phi(\mathbf{t}_j), \phi(\mathbf{t}_j)\right) = 1. \quad (55)$$

By substituting the above results into (54), we arrive at

$$||\phi(\mathbf{t}_i) - \phi(\mathbf{t}_j)||^2 = 2 - 2K\left(\phi(\mathbf{t}_i), \phi(\mathbf{t}_j)\right). \quad (56)$$

Thus, the proof of the result is completed. ■

References

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

[2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Commun. Mag.*, vol. 49, no. 6, 2011.

[3] 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, 2013.

[4] W. Roh, J.-Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, 2014.

[5] C. Wang and H.-M. Wang, "Physical layer security in millimeter wave cellular networks," *IEEE Trans. on Wireless Commu.*, vol. 15, no. 8, pp. 5569–5585, 2016.

[6] H.-M. W. Y. Ju and T.-X. Zheng, "Secure transmissions in millimeter wave systems," *IEEE Trans. on Commu.*, vol. 65, no. 5, pp. 2114–2127, 2017.

[7] 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, 2014.

[8] 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, 2014.

[9] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave mimo systems with large antenna arrays," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 998–1009, 2016.

[10] H. Lin, F. Gao, S. Jin, and G. Y. Li, "A new view of multi-user hybrid massive MIMO: Non-orthogonal angle division multiple access," *IEEE J. Sel. Areas Commun.*, vol. PP, no. 99, pp. 1–1, 2017.

[11] J. Zhao, F. Gao, W. Jia, S. Zhang, S. Jin, and H. Lin, "Angle domain hybrid precoding and channel tracking for mmwave massive MIMO systems," *IEEE Trans. on Wireless Commu.*, vol. PP, no. 99, pp. 1–1, 2017.

[12] Y. Niu, Z. Feng, Y. Li, Z. Zhong, and D. Wu, "Low complexity and near-optimal beam selection for millimeter wave MIMO systems," in *Wireless Communications and Mobile Computing Conference, 2017 13th International*. IEEE, 2017, pp. 634–639.

[13] Y. Gao, M. Khaliel, T. Kaiser *et al.*, "Rotman lens based hybrid analog-digital beamforming in massive MIMO systems: Array architectures, beam selection algorithms and experiments," *IEEE Trans. Veh. Technol.*, 2017.

[14] Y. Ren, Y. Wang, C. Qi, and Y. Liu, "Multiple-beam selection with limited feedback for hybrid beamforming in massive MIMO systems," *IEEE Access*, 2017.

[15] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big Data*, vol. 1, no. 1, pp. 51–59, 2013.

[16] C. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[17] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 190–199, 2015.

[18] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5g: how to empower SON with big data for enabling 5G," *IEEE Network*, vol. 28, no. 6, pp. 27–33, 2014.

[19] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[20] C. Cortes and V. Vapnik, "Support vector machine," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[21] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for eeg-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 4, no. 2, p. R1, 2007.

[22] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[23] M. Matthaiou, C. Zhong, and T. Ratnarajah, "Novel generic bounds on the sum rate of MIMO ZF receivers," *IEEE Trans. on Signal Processing*, vol. 59, no. 9, pp. 4341–4353, 2011.

[24] M. Matthaiou, C. Zhong, M. R. Mckay, and T. Ratnarajah, "Sum rate analysis of zf receivers in distributed MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 180–191, 2013.

[25] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.

[26] S. Arlot, A. Celisse *et al.*, "A survey of cross-validation procedures for model selection," *Statistics surveys*, vol. 4, pp. 40–79, 2010.

[27] X.-R. Cao, "The maclaurin series for performance functions of markov chains," *Advances in Applied Probability*, vol. 30, no. 3, pp. 676–692, 1998.

[28] R. Raz, "On the complexity of matrix product," *In Proceedings of the thirty-fourth annual ACM symposium on Theory of computing*, vol. 32, no. 5, 2002.

[29] J. Joung, "Machine learning-based antenna selection in wireless communications," *IEEE Comm. Lett.*, vol. 20, no. 11, pp. 2241–2244, 2016.

Yin Long received the M.S. degree in communication engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2013. He is currently pursuing the Ph.D. degree in communication with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu, China. He was a Visiting Student with the University of Alberta, Edmonton, Alberta, Canada, from 2016 to 2017. His current research interests include massive MIMO, random matrix theory and data mining. He has served as a Reviewer for various international journals and conferences, including the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Zhi Chen (M'08-SM'16) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1997, 2000, and 2006, respectively. In 2006, he joined the National Key Laboratory of Science and Technology on Communications, UESTC, where he has been a Professor since 2013. He was a Visiting Scholar with the University of California at Riverside, Riverside, CA, USA, from 2010 to 2011. His current research interests include 5G mobile communications, tactile Internet, and terahertz communication. He has served as a Reviewer for various international journals and conferences, including the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Jun Fang (M'08) received the B.S. and M.S. degrees from Xidian University, Xian, China, in 1998 and 2001, respectively, and the Ph.D. degree from the National University of Singapore, Singapore, in 2006, all in electrical engineering. During 2006, he was a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University. From 2007 to 2010, he was a Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology. Since 2011, he has been with the University of Electronic of Science and Technology of China. His current research interests include sparse theory and compressed sensing, statistical learning, millimeter Wave, and massive MIMO communications. Dr. Fang received the IEEE Jack Neubauer Memorial Award in 2013 for the best systems paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He is an Associate Technical Editor for the IEEE Communications Magazine, and an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS.

Chintha Tellambura (F'11) received the B.Sc. degree (with first-class honor) from the University of Moratuwa, Sri Lanka, the MSc degree in Electronics from Kings College, University of London, United Kingdom, and the PhD degree in Electrical Engineering from the University of Victoria, Canada.

He was with Monash University, Australia, from 1997 to 2002. Presently, he is a Professor with the Department of Electrical and Computer Engineering, University of Alberta. His current research interests include the design, modelling and analysis of cognitive radio, heterogeneous cellular networks and 5G wireless networks.

Prof. Tellambura served as an editor for both IEEE Transactions on Communications (1999-2011) and IEEE Transactions on Wireless Communications (2001-2007) and for the latter he was the Area Editor for Wireless Communications Systems and Theory during 2007-2012. He has received best paper awards in the Communication Theory Symposium in 2012 IEEE International Conference on Communications (ICC) in Canada and 2017 ICC in France. He is the winner of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. In 2011, he was elected as an IEEE Fellow for his contributions to physical layer wireless communication theory. In 2017, he was elected as a Fellow of Canadian Academy of Engineering. Prof. Tellambura has authored or coauthored over 500 journal and conference papers with total citations more than 14,000 and an h-index of 64 (Google Scholar).