Received 23 May 2021; revised 25 June 2021; accepted 28 June 2021. Date of publication 6 July 2021; date of current version 19 July 2021.

Digital Object Identifier 10.1109/OJCOMS.2021.3094823

Joint Analog Beam Selection and Digital Beamforming in Millimeter Wave Cell-Free Massive MIMO Systems

CENK M. YETIS [1] (Member, IEEE), EMIL BJÖRNSON [2] (Senior Member, IEEE), AND PONTUS GISELSSON [1]

[1] Department of Automatic Control, Lund University, 22100 Lund, Sweden

[2] Department of Electrical Engineering (ISY), Linköping University, 58183 Linköping, Sweden

CORRESPONDING AUTHOR: C. M. YETIS (e-mail: cenkmyetis@ieee.org)

This work was supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

ABSTRACT Cell-free massive MIMO systems consist of many distributed access points with simple components that jointly serve the users. In millimeter wave bands, only a limited set of predetermined beams can be supported. In a network that consolidates these technologies, downlink analog beam selection stands as a challenging task for the network sum-rate maximization. Low-cost digital filters can improve the network sum-rate further. In this work, we propose low-cost joint designs of analog beam selection and digital filters. The proposed joint designs achieve significantly higher sum-rates than the disjoint design benchmark. Supervised machine learning (ML) algorithms can efficiently approximate the input-output mapping functions of the beam selection decisions of the joint designs with low computational complexities. Since the training of ML algorithms is performed off-line, we propose a well-constructed joint design that combines multiple initializations, iterations, and selection features, as well as beam conflict control, i.e., the same beam cannot be used for multiple users. The numerical results indicate that ML algorithms can retain 99-100% of the original sum-rate results achieved by the proposed well-constructed designs.

INDEX TERMS Cell-free, millimeter wave, hybrid architecture, analog beamforming, digital beamforming, beam training.

I. INTRODUCTION

CELL-FREE massive MIMO (mMIMO) networks thrive on the idea of jointly and coherently serving a proportionally small number of users by a large number of simple multi-antenna access points (APs). Compared to cellular mMIMO networks, cell-free networks can provide a more uniform service performance for the users in the network since the antennas are distributed. For example, the 95%-likely per-user spectral and energy efficiencies of cell-free networks are five and ten times higher than cellular networks, respectively [1].

Millimeter wave (mm-wave) communications can achieve multi-Gbps data rates by exploiting underutilized wide bandwidths in the mm-wave spectrum. Therefore, the consolidation of mm-wave communications and cell-free networks is a promising direction for the next generation wireless networks [2]. In particular, the macro-diversity achieved by having many distributed APs compensate for the spotty coverage that otherwise limits the practical use of mm-wave spectrum. The hardware complexity of the AP is critical to reduce the costs of deployment and power consumption of APs that are deployed in large numbers in the cell-free network. Thus, low complexity hybrid analog and digital beamforming designs need to be adopted.

In this work, we propose low-cost joint design algorithms for analog beam selection and digital beamforming in the downlink transmission of a mm-wave cell-free mMIMO system consisting of multiple antenna APs and single antenna users. We assume APs are equipped with uniform linear arrays (ULAs) and the number of radio frequency (RF) chains at an AP is equal to the number of transmitted streams. The analog beam selection process does not rely on an external aid, e.g., location information of the users, but only on the sum-rate metrics of the users. Furthermore, we incorporate multiple initializations, iterations, and selection features, as well as beam conflict control (BCC), i.e., a selected beam for a user cannot be reselected for another user. We refer to a joint design with multiple features as a well-constructed design. The proposed joint design solutions achieve significant network sum-rate gains compared to the naive disjoint design of analog beam selection, which is based on the direct link (DL) power metrics of the users, and digital precoder. The disjoint approach first completes the analog beam selection and then completes the digital precoder design. On the other hand, the proposed joint approach iteratively updates the beam selections and digital precoder designs until convergence.

Finally, for analog beam selection, we propose machine learning (ML) algorithms that are trained off-line by the proposed well-constructed designs. Online beam selection by an ML algorithm which is succeeded by digital precoder designs are one-time-only, i.e., no for-loops, and this approach can mirror our proposed well-constructed designs. The proposed ML based approach can achieve 99-100% of the original sum-rate results achieved by the well-constructed designs.

II. RELATED WORKS AND CONTRIBUTIONS

Beamforming designs with digital precoders for microwave communications in cell-free networks are investigated in [3], [4]. ZF, minimum mean squared error (MMSE), maximum-ratio transmitter/combiner (MRT/MRC) digital filters can be preferred in the baseband processing of mm-wave networks [5]–[7] for their low computational complexities.

In mm-wave networks, proper selections of beams are achieved by transmitting the candidate beams and measuring their performances by a network metric. This process is known as beam training. The conventional beam training is time and power inefficient since it sweeps all beam directions exhaustively. The hierarchical codebook approach lowers the search complexity by implementing a tree-structured codebook design [8]. Nevertheless, the complexity of hierarchical search is still high [9].

In [10], [11] compressive sensing is implemented to lower the beam selection complexity. By exploiting the spatial sparsity of mm-wave channels, beam selection is achieved implicitly in the reduced-dimensional beamspace CSI domain. In [12], the computational complexity is reduced by combining probabilistic framework based simulated annealing and the 2-D numerical method based Rosenbrock search procedure. In [13], a new antenna array that can adjust the gain and phase of each antenna is used to support multi-stream transmission with fewer number of RF chains. In [14], two partial beam training strategies are proposed with reduced computational complexities. In [15], the computational complexity is reduced by a two-stage beam training procedure. In [16], location assistance is proposed to eliminate the need for beam sweeping at the cost of increased power consumption due to continuous global positioning system (GPS) connectivity. In [17], a multi-user scalable and channel variation robust beam searching method is proposed. In [18], a branch-and-bound based beam searching algorithm is proposed for a hybrid design with a subarray architecture.

Ideally, the beam searching complexity should be linear in the network parameters, i.e., the number of APs, users, and the number of antennas at APs, to achieve scalability with the network size. In this work, given a beamforming codebook, we propose two beam searching algorithms, namely semilinear and linear search algorithms. The search complexity of the former is exponential with respect to the number of users but linear with the number of APs whereas the search complexity of the latter is linear both with the number of APs and users.

The assignment of the same beam to multiple users results in beam conflict in mm-wave networks. Beam conflict causes effective channels to have low-ranks which can significantly reduce the sum-rate of the network [14]. In this work, we propose BCC to entirely eliminate the beam conflict in the network. In cell-free networks, BCC is more challenging since an assigned beam to a user can be decided to be used by another AP for the same user as well. On the contrary, for instance, in interference networks, where a transmitter sends a different stream to each user, a different beam is assigned between each AP and a user. Thus, the assigned beam can be immediately removed from the common codebook of the network. As a solution to this challenge in cell-free networks, in this work, we propose to keep a codebook log for each user which is updated and announced in the network after each beam assignment. BCC implementation can also reduce the simulation duration due to the reduced number of beam combinations. In this work, we also propose BCC initializations, i.e., random initializations that satisfy BCC, for the algorithms without the BCC implementation.

Another major challenge in cell-free networks is that an AP needs to equip as many RF chains as the number of users since APs jointly serve all users. However, there can be opportunities in the varying channel conditions to shut off RF chains to gain significant power savings at low sum-rate losses. In this work, we also propose to adaptively shut off the RF chains that leads to 30% power gain at the cost of 5% sum-rate loss on average.

As the network size grows, even low complexity beam selection algorithms can be too costly. Next, we propose fast and efficient supervised ML algorithms which can be trained by the proposed well-constructed algorithms. In particular, three ML algorithms motivate us, the support vector machine (SVM), multi-layer perceptron (MLP), and random forest (RFt) algorithms. BCC introduces patterns on the outputs of the mapping functions in the beam selection problems. Hence, the labels that are input to ML algorithms become highly correlated. Classifier chains are proven to be effective in exploiting the correlations in multi-label ML problems [19]. Our numerical results indicate that the RFt algorithm with the classifier chains can retain 99% of the original sum-rate results that are achieved by the beam selection algorithms. This is a sharp contrast to 63% achieved without the classifier chains.

Since the training of ML algorithms is carried out off-line, in this work, a well-constructed joint analog beam selection and digital beamforming algorithm is proposed. In particular, we propose a joint design with multiple initializations and iterations. Furthermore, due to BCC, the assignments of beams to the users in the earlier search segments impinge the assignments in the later search segments. In addition to multiple initializations and iterations, we integrate a selection feature to the proposed joint design as well. Numerical results indicate that the linear search algorithm with the selection feature can achieve higher sum-rate at a lower complexity even than the semilinear algorithm that has no selection feature.

The rest of the paper is organized as follows. In Section III, we introduce the system model of mm-wave cell-free mMIMO networks. In Section IV, the problem formulation of joint design is introduced. In Section V, the proposed low complexity beam searching algorithms are introduced, and the challenge of BCC in cell-free networks is addressed and a solution is proposed. In Section VI, the solution to multi-output classification problem under BCC is provided and ML implementation is outlined. In Section VII, the numerical results of joint design and ML algorithms are presented, and finally, the paper is concluded in Section VIII.

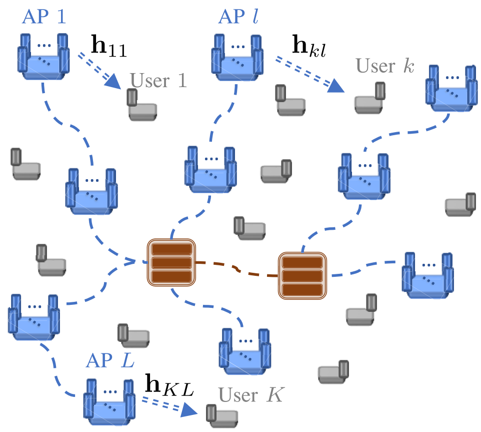
Notations: Throughout the paper, $(.)^H$ and $(.)^{-1}$ denote the conjugate transpose and the inverse operations of a matrix, respectively. $\|.\|_2$ denotes the L_2 vector norm operator. $\mathcal{CN}(0, x)$ denotes the complex Gaussian distribution with zero mean and variance x. $E\{.\}$ and $|.|$ are the expectation and absolute value operators, respectively. Finally, ! denotes the factorial operator.

III. SYSTEM MODEL

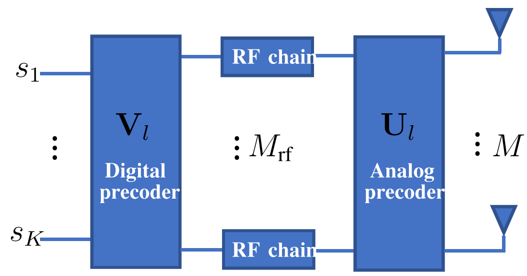
We consider the downlink transmission of a mm-wave cell-free mMIMO system consisting of L APs and K users. Each AP has M antennas, while each user has a single antenna. In contrast to cell-free systems built for sub-6 GHz systems, the number AP antennas can be very large in mm-wave systems. The channel between AP l and user k is $\mathbf{h}_{kl} \in \mathbb{C}^M$, and it is assumed to be frequency-flat. The system and AP models are shown in Fig. 1.

A. RECEIVED SIGNAL

Consider a hybrid beamforming case where M_{rf} is the number of radio frequency chains at an AP and, in general, $M \geq M_{\text{rf}} \geq K$. The radio frequency (analog) and baseband (digital) precoders at AP l, $\forall l \in \mathcal{L} \triangleq \{1, \ldots, L\}$ are given by $\mathbf{U}_l \in \mathbb{C}^{M \times M_{\text{rf}}}$ and $\mathbf{V}_l = [\mathbf{v}_{1l} \ldots \mathbf{v}_{Kl}] \in \mathbb{C}^{M_{\text{rf}} \times K}$, respectively. We assume $M_{\text{rf}} = K$ so the system can generate a different signal for each user. The precoding vector $\mathbf{u}_{kl} \in \mathbb{C}^M$, i.e., the k^{th} column of $\mathbf{U}_l$, is chosen from a unitary codebook with B predefined beams $\underline{\mathcal{U}} = \{\underline{\mathbf{u}}_1, \ldots, \underline{\mathbf{u}}_b, \ldots, \underline{\mathbf{u}}_B\}$. The APs collaboratively transmit the unit-power data signals



(a) Cell-free network model.



(b) AP model.

FIGURE 1. Cell-free and AP models.

s_k, $\forall k \in \mathcal{K} \triangleq \{1, \ldots, K\}$ to the users. The received signal at user k during the data transmission phase is given by

$$y_k = \sqrt{\frac{p_{\text{T}}}{K}} \sum_{l=1}^{L} \sum_{i=1}^{K} \mathbf{h}_{kl}^H \tilde{\mathbf{u}}_{il} s_k + n_k, \tag{1}$$

where $\tilde{\mathbf{u}}_{il} \triangleq \mathbf{U}_l \mathbf{v}_{il}$ and $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ is the noise at user k, $p_{\text{T}} > 0$ is the transmit power of an AP. Equal power allocation is assumed at the APs. The sum-rate performances can be improved by another fixed but unequal power allocation solution [20]. Since the main goal of this paper is centered around joint design solutions with low complexity and effective beam searching and ML algorithms, the extension in this direction is omitted for the sake of simplicity.

The received signal (1) can be rewritten as

$$y_k = D_k + I_k + n_k, \tag{2a}$$

where

$$D_k = \sqrt{\frac{p_{\text{T}}}{K}} \sum_{l=1}^{L} \mathbf{h}_{kl}^H \tilde{\mathbf{u}}_{kl} s_k \text{ and} \tag{2b}$$

$$I_k = \sqrt{\frac{p_{\text{T}}}{K}} \sum_{l=1}^{L} \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{h}_{kl}^H \tilde{\mathbf{u}}_{jl} s_j \tag{2c}$$

are the desired and interference signals, respectively.

B. CHANNEL MODEL

The channels are characterized by using the extended Saleh-Valenzuela geometric channel model with P scatterers per user [21], [22]. The channel vector is defined by

$$\mathbf{h}_{kl} = \sqrt{\frac{M}{P}} \sum_{p=1}^{P} \frac{\beta_{kl}^{p}}{\sqrt{\alpha_{kl}^{p}}} \mathbf{a}\left(\theta_{kl}^{p}\right), \tag{3}$$

where β_{kl}^{p} is the channel gain with a random but fixed complex value, $\mathbf{a}(\theta_{kl}^{p})$ is the array response vector for a given angle of departure (AoD), $\theta_{kl}^{p} \in [-\pi, \pi)$ is the AoD for the path p, and α_{kl}^{p} is the path loss given by

$$\alpha_{kl}^{p}(\text{dB}) = 20\log_{10}(4\pi f_c/c) + 10n\log_{10}(d) + X_{\sigma}, \tag{4}$$

where c, f_c, n, d, and X_{σ} denote the speed of light (m/s), the carrier frequency (Hz), the path loss exponent, the distance (m), and the shadow fading following a normal distribution with mean 0 and standard deviation σ (dB).

Assuming each AP is equipped with ULA, the element y, $y = 1, \ldots, M$ of the array response vector for AP l is given by

$$\mathbf{a}\left(\theta_{kl}^{p}, y\right) = \sqrt{\frac{1}{M}} e^{(y-1)j2\pi(d/\lambda)\sin\left(\theta_{kl}^{p}\right)}. \tag{5}$$

IV. PROBLEM FORMULATION

In this section, the joint design of beam selection and digital precoder design problem in cell-free networks under BCC formulated. The beam selection metric affects the complexity of the searching algorithm. In this section, two beam selection metric options are introduced as well.

A. JOINT DESIGN

The rate of user k is given as

$$\mathrm{R}_k = \log_2(1 + \mathrm{SINR}_k), \tag{6a}$$

where

$$\begin{aligned}\mathrm{SINR}_k &= \frac{E\left\{|D_k|^2\right\}}{E\left\{|I_k|^2\right\} + \sigma_n^2} \\ &= \frac{\frac{p_{\mathrm{T}}}{K}\left|\sum_{l=1}^{L} \mathbf{h}_{kl}^{H}\tilde{\mathbf{u}}_{kl}\right|^2}{\frac{p_{\mathrm{T}}}{K}\sum_{\substack{j=1\\ j\neq k}}^{K}\left|\sum_{l=1}^{L}\mathbf{h}_{kl}^{H}\tilde{\mathbf{u}}_{jl}\right|^2 + \sigma_n^2}\end{aligned} \tag{6b}$$

is the signal-to-interference-plus-noise-ratio (SINR) of user k.

The joint design problem of beam selection and digital precoder under BCC for sum-rate maximization is given as

$$\begin{aligned}&\underset{\{\mathbf{u}_{kl}\}}{\arg\max}\ \max_{\{\mathbf{v}_{kl}\}} \sum_{k=1}^{K} \mathrm{R}_k && (7a)\\ &\text{s.t.}\ \ \mathbf{u}_{kl} \in \underline{\mathcal{U}}, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}, && (7b)\\ &\qquad \mathbf{u}_{kl} \neq \mathbf{u}_{jm}, \forall k, \forall j, j \neq k \in \mathcal{K}, \forall l, \forall m \in \mathcal{L}, && (7c)\\ &\qquad \|\mathbf{U}_l\mathbf{v}_{kl}\|_2^2 = 1, \forall k \in \mathcal{K}, \forall l \in \mathcal{L}. && (7d)\end{aligned}$$

The first constraint (7b) implies that the analog precoders are chosen from a given codebook. The second constraint (7c) asserts BCC on the solution, i.e., the same beam cannot be assigned to multiple users. Note that in cell-free networks, all APs collaboratively transmit the same data signal to a user. Therefore, as seen in (7c), APs are allowed to choose the same beam to the same user. Finally, the last constraint (7d) asserts that the digital precoder does not provide a power gain.

B. SELECTION METRICS

A beam selection algorithm that is based on the maximum received signal power at each user can be achieved distributively with a low computational complexity and with low signalling overhead. The selection metric as the received signal power between AP l and user k is given by

$$E\left\{|D_{kl}|^2\right\} = \frac{p_{\mathrm{T}}}{K}\left|\mathbf{h}_{kl}^{H}\tilde{\mathbf{u}}_{kl}\right|^2. \tag{8}$$

However, the sum-rate reflects the trade-off between strong signal and weak interference which is the essence of sum-rate maximization problem given in (7) [23], [24]. ML algorithms provide the needed succor for the increased costs due to using the sum-rate as the beam selection metric. In other words, since the training of ML algorithms is performed off-line, the increased costs can be tolerated.

V. PROPOSED JOINT DESIGN ALGORITHMS

The beam selection problem is distinctively challenging. In a cell-free network, there is only one large cell where there are far more APs than users, and all APs serve all users. An exhaustive algorithm centrally selects the optimal beams by evaluating the sum-rates achieved by all stream combinations between all APs and users, and then by choosing the beam combination that results in the maximum sum-rate. The complexity of an exhaustive algorithm is B^{KL}. Assuming the time slot of beam transmission is normalized to 1, the delay of a beam selection algorithm is equal to its complexity. For instance, an exhaustive algorithm awaits B^{KL} time slots to evaluate all beam combinations and centrally select the best beam combination. Therefore, as the cell-free network size grows, an exhaustive algorithm becomes intractable in its computational complexity and initial access delay.

In this section, we propose two low complexity, thus low delay, beam selection algorithms, which are coined as semilinear and linear search algorithms. The search processes are iterated segment by segment as illustrated in Fig. 2. While all possible beam combinations are being searched in a segment, beams in other segments are held fixed. In the semilinear algorithm, a segment is formed between an AP and all users. In other words, compared to the centralized algorithm where the search segment is the whole network, the searching process is segmented into APs. To reduce the complexity further, in the linear search algorithm, the searching process is segmented into APs as well. Hence, the complexities of semilinear and linear search algorithms become linear in only L, and in both L and K, respectively.

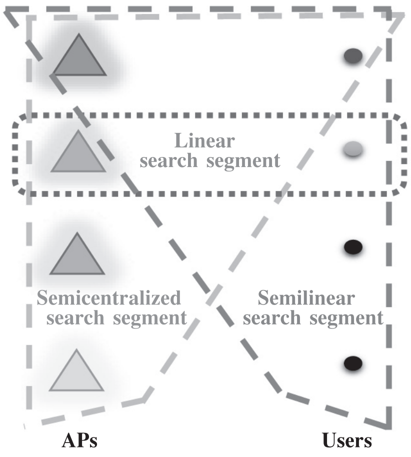


FIGURE 2. The search segments of semi-centralized, semilinear and linear search algorithms. The search segment of centralized, i.e., exhaustive, algorithm covers the whole network.

A. SEMILINEAR AND SEMICENTRALIZED SEARCHES

In the semilinear search algorithm, the selection process is proceeded between each AP and all users. After the transmissions of B^K possible beam combinations from an AP to all users, the beam combination that achieves the maximum sum-rate is selected. Then, the algorithm proceeds with the selection of the best beam combination from the next AP to all users. The selection of beams from an AP to all users is carried out by fixing the beams from all other APs $\forall l'$, $l' \in \{1, \ldots, L\}\backslash\{l\}$ in the network. The complexity of this algorithm is $L(B^K)$, which is linear in the number of APs.

On the other hand, in the semi-centralized search algorithm, the selection process is proceeded between all APs and a user. After the transmissions of B^L possible beam combinations from all APs to user k, the beam combination that achieves the maximum rate at user k is selected. Then, the algorithm proceeds with the selection of the best beam combination for the next user. The selection of beams from all APs to user k is carried out by fixing the beams to other users $\forall k'$, $k' \in \{1, \ldots, K\}\backslash\{k\}$ in the network. The complexity of this algorithm is $K(B^L)$, which is linear in the number of users.

For the joint design problem (7), semilinear is advantageous over the semi-centralized search algorithm from two aspects. First of all, as mentioned earlier, in cell-free networks, $L \gg K$. Hence, a search algorithm which is linear in L is preferable. Secondly, when $B < L$, the semi-centralized algorithm can run out of available beams for other users k' after the first search segment in case each AP wishes to transmit a different beam to user k. Note that, due to BCC, for both semilinear and semi-centralized algorithms $B \geq K$ must be satisfied.

Next, a linear search algorithm is proposed to achieve a more fundamental complexity reduction.

B. LINEAR SEARCH

In the linear search algorithm, a pair of one AP and one user forms a search segment. The selection process is proceeded pair by pair. The selection of beam from AP l to user k is carried out by fixing all other the beams in the network. In particular, the beams between AP l to users $\forall k'$, and the beams between APs $\forall l'$, $l' \in \{1, \ldots, L\}\backslash\{l\}$ and users $\forall k$ are fixed. The complexity of this algorithm is LKB, which is linear in the number of APs, users, and antennas at each AP.

C. BEAM CONFLICT CONTROL

Searching processes at the segments of the proposed searching algorithms can be executed in parallel or series. The latter option improves the search quality, but it introduces a delay. However, due to BCC, parallel search can introduce even more delays and it can make the search process even more complex as follows. The beam selection results of the parallel search segments need to be immediately announced in the network. The segments that receive the announcements before the start of the search process can update the codebook log for each user $\mathcal{U}_k$, i.e., remove a codeword from the codebook of user k, $\mathcal{U}_k$, if that codeword is assigned to another user k'. If the searching process has already started, then the result of the process can be waited. If the result leads a codeword assignment that is already assigned to another user, then the search process needs to be restarted after updating $\mathcal{U}_k$. Hence, whether synchronous or asynchronous parallel search process is applied, the delay of parallel process can be even more than the serial process under BCC.

Note that BCC is different in cell-free networks than in other networks where the same user is not coherently served by the APs, i.e., APs transmit different data signals to the same user. In other networks, between each AP and a user, a different analog beam should be used. In this case, once a beam is selected, it can be removed from the common codebook $\mathcal{U}$. In cell-free networks, all APs want to coherently transmit to each user. Therefore, between all APs and a user, the same analog beam can be used if APs choose to, but the same beam cannot be used for another user. In this case, once a beam is selected, it should not be removed from the codebook in case another AP wants to use it for the same user. This means that a codebook log for each user $\mathcal{U}_k$ must be updated and announced after each decision in the network. This makes BCC a more challenging task in cell-free networks.

However, under BCC, the number of beam combinations is reduced from LB^K to $LB!/(B-K)!$ for the linear search algorithm. Hence, without BCC, it is more costly to try all combinations as B and K increase. As demonstrated numerically in Section VII, the algorithms with and without BCC, and the algorithms with BCC initialization result in interesting trade-offs between the network sum-rate and simulation duration.

D. DIGITAL PRECODERS

In this work, we utilize ZF digital precoders in the baseband process of APs. At AP l, ZF precoder is given as

$$\mathbf{V}_l^{\text{ZF}} = \mathbf{H}_l^H\left(\mathbf{H}_l\mathbf{H}_l^H\right)^{-1}, \tag{9a}$$

where

$$\mathbf{H}_l = \begin{bmatrix} \mathbf{h}_{1l}^H\mathbf{U}_l \\ \cdots \\ \mathbf{h}_{Kl}^H\mathbf{U}_l \end{bmatrix} \tag{9b}$$

is the effective channel from AP l to all users.

The MMSE precoder can be preferred in the lower signal-to-noise ratio (SNR) regime where it can achieve higher sum-rate results than the ZF precoder. Furthermore, when BCC is not used, $\mathbf{H}_l$ can be a low-rank matrix. Thus, MMSE precoder can be used instead of ZF precoder since it can support the low-rank issue to some extent. However, BCC is worthwhile since the sum-rate results without BCC are lower as demonstrated by numerical results in Section VII. The MMSE precoder at AP l is given as

$$\mathbf{V}_l^{\text{MMSE}} = \mathbf{H}_l^H\left(\frac{p_{\text{T}}}{K}\mathbf{H}_l\mathbf{H}_l^H + \sigma_n^2\mathbf{I}_K\right)^{-1}, \tag{10}$$

where $\mathbf{I}_K$ is the identity matrix of size $K \times K$.

The digital precoders in (9) and (10) require local baseband CSI, i.e., the effective channel vectors from AP l to all users, $\mathbf{h}_{kl}^{\text{eff}} = \mathbf{h}_{kl}^H\mathbf{U}_l$, $\forall k \in \mathcal{K}$. Here, $(\mathbf{h}_{kl}^{\text{eff}})^T \in \mathbb{C}^{M_{\text{rf}}}$, where $(.)^T$ is the transpose operator. Note that the channel $\mathbf{h}_{kl} \in \mathbb{C}^M$ cannot be directly estimated due to the constraint in the number of RF chains, $M_{\text{rf}} \leq M$, in general [25]. Different analog beamforming vectors result in different effective channels. Therefore, as widely proposed in the literature and standards, after the beam selections are concluded, the effective channels are estimated to design the digital precoders in the final stages of the hybrid designs [5]–[7], [14], [23]. We refer this approach as a disjoint design. On the other hand, as detailed in the next section, our proposed joint designs require the effective channel estimation and digital precoder design at AP l each time AP l tests a beam. Similar to multiple initializations, iterations, and selection features, the mentioned requirement can be impractical in fast varying channel conditions. However, as detailed in Section VI and numerically demonstrated in Section VII, ML algorithms can overcome these impractical challenges. By training the ML algorithms off-line with our proposed well-constructed designs to do the beam selection, online execution of the ML algorithms for the beam selection followed by a digital precoder design mimics the proposed well-constructed designs and can achieve nearly the same sum-rate results at significantly lower computational complexities.

E. PSEUDOCODES

The pseudocodes of the proposed semilinear and linear search algorithms are given in Algorithms 1-3. The common lines of the proposed algorithms are given in Algorithm 1. At step 9 of Algorithm 1, either the semilinear or linear search in Algorithm 2 or 3 is chosen, respectively. In Algorithm 4, the pseudocode of the proposed algorithm for the prioritization of search segments is given. For brevity, based on the proposed joint design of beam selection and digital precoder, the algorithms that apply multiple initializations and iterations are denoted by II in Algorithm 1 and the algorithm that applies the selection method as well is denoted by IIS in Algorithm 4. Algorithms that use the DL power in (8) and the sum-rate as the beam selection metrics are noted by DL and rate, respectively. The pseudocodes of algorithms that

Algorithm 1 Pseudocodes of the Proposed Semilinear-II-Rate and Linear-II-Rate Algorithms

1: $\text{R}_{\text{sum}}^{\max} = 0$
2: **for** $i = 1$: Initializations **do**
3:   Random initializations of analog precoders $\mathbf{u}_{kl}^0 \in \underline{\mathcal{U}}$, $\forall k, \forall l$ $\mathbf{u}_{kl}^0 \neq \mathbf{u}_{k'j}^0, \forall l, \forall j, \forall k, \forall k', k \neq k'$ (with BCC)
4:   $\mathbf{u}_{kl} = \mathbf{u}_{kl}^0$, $\forall k, \forall l$
5:   $\underline{\mathbf{U}}^{\text{comb}}$ (for the semilinear search algorithm)
6:   $\mathcal{U}_k = \underline{\mathcal{U}}$, $\forall k$ (for the linear search algorithm)
7:   **for** $s = 1$: Iterations **do**
8:     **for** $l = 1 : L$ **do**
9:       *Choose semilinear or linear search algorithm*
10:     **end for**
11:     R_{sum}'
12:     **if** $\text{R}_{\text{sum}}' > \text{R}_{\text{sum}}^{\max}$ **then**
13:       $\mathbf{U}_l^{\text{opt}} = \mathbf{U}_l^{\star}$, $\forall l$ (for the semilinear search algorithm)
14:       $\mathbf{u}_{kl}^{\text{opt}} = \mathbf{u}_{kl}^{\star}$, $\forall k, \forall l$ (for the linear search algorithm)
15:       $\text{R}_{\text{sum}}^{\max} = \text{R}_{\text{sum}}'$
16:     **end if**
17:   **end for**
18: **end for**

Algorithm 2 The Proposed semilinear Search Algorithm

for $c = 1 : C$ **do**
  $\mathbf{U}_l = \underline{\mathbf{U}}_c^{\text{comb}}$
  $\mathbf{V}_l$
  $\text{R}_{\text{sum}}(c)$
end for
$[\sim, c^\star] = \text{maxvec}[\text{R}_{\text{sum}}(1) \ldots \text{R}_{\text{sum}}(c) \ldots \text{R}_{\text{sum}}(C)]$
$\mathbf{U}_l^\star = \underline{\mathbf{U}}_{c^\star}^{\text{comb}}$
Update $\underline{\mathbf{U}}^{\text{comb}}$ (with BCC)
Update C (with BCC)

Algorithm 3 The Proposed Linear Search Algorithm

for $k = 1 : K$ **do**
  for $b = 1 : B_k$ **do**
    $\mathbf{u}_{kl} = \underline{\mathbf{u}}_b$
    $\mathbf{V}_l$
    $\text{R}_{\text{sum}}(b)$
  end for
  $[\sim, b^\star] = \text{maxvec}\big[\text{R}_{\text{sum}}(1) \ldots \text{R}_{\text{sum}}(b) \ldots \text{R}_{\text{sum}}(B_k)\big]$
  $\mathbf{u}_{kl}^\star = \underline{\mathbf{u}}_{b^\star}$
  $\mathcal{U}_{k'} \leftarrow \mathcal{U}_{k'} \setminus \underline{\mathbf{u}}_{b^\star}$, $\forall k', k' \neq k$ (with BCC)
  Update $B_{k'}$, $\forall k', k' \neq k$ (with BCC)
end for

Algorithm 4 Pseudocode of the Proposed Linear-IIS-Rate Algorithm

```
1: R_sum^max = 0
2: for i = 1 : Initializations do
3:     Random initializations of analog precoders u_kl^0 ∈ 𝒰, ∀k, ∀l
       u_kl^0 ≠ u_k'j^0, ∀l, ∀j, ∀k, ∀k', k ≠ k' (with BCC)
4:     𝒰_k = 𝒰, ∀k
5:     L_vec = 1 : L
6:     for j = 1 : L do
7:         u_kl = u_kl^0, ∀k, ∀l
8:         R_sum^max' = 0
9:         for s = 1 : Iterations do
10:            for l ∈ L_vec do
11:                Apply linear search algorithm
12:            end for
13:            Same lines 11-16 in Algorithm 1
14:        end for
15:        if R_sum^max' > R_sum^max then
16:            u_kl^opt = u_kl^opt', ∀k, ∀l
17:            R_sum^max = R_sum^max'
18:        end if
19:        L_vec = circshift(L_vec)
20:    end for
21: end for
```

use DL metric are same as in Algorithms 1–4 after replacing the sum-rate metrics, $\mathrm{R}_{\mathrm{sum}} \triangleq \sum_{k=1}^{K} \mathrm{R}_k$ with the sum-DL metrics, $\mathrm{DL}_{\mathrm{sum}} \triangleq \sum_{l=1}^{L} \sum_{k=1}^{K} E\{|D_{kl}|^2\}$.

The proposed algorithms are well-constructed to train ML algorithms in the sense that they are consolidated from joint design of analog and digital beam precoders (lines 2-3 for the semilinear search in Algorithm 2 and lines 3-4 for the linear search in Algorithm 3), multiple random initializations of analog precoders (the for-loops at step 2 in Algorithms 1 and 4), multiple iterations of the joint design (the for-loops at steps 7 and 9 in Algorithms 1 and 4, respectively), and finally, the selection of the prioritized search segment to assign the beams first than the other segments (the for-loop at step 6 in Algorithm 4).

As explained earlier, a codebook log is updated and announced at each user for BCC implementation in the cell-free network. As seen at steps 6 and 4 of Algorithms 1 and 4, respectively, the codebook log at each user $\mathcal{U}_k$ is initialized by copying the codebook $\underline{\mathcal{U}}$. On the other hand, for the semilinear search algorithm, a single codebook log, $\underline{\mathbf{U}}^{\mathrm{comb}}$, is tracked for simplicity as seen at step 5 of Algorithm 1. $\underline{\mathbf{U}}^{\mathrm{comb}}$ is the matrix set of beam combinations under BCC

$$\underline{\mathbf{U}}^{\mathrm{comb}} \triangleq \left\{ \underline{\mathbf{U}}_1^{\mathrm{comb}}, \ldots, \underline{\mathbf{U}}_c^{\mathrm{comb}}, \ldots, \underline{\mathbf{U}}_C^{\mathrm{comb}} \right\}, \tag{11}$$

where initially, $C = B(B-1)\ldots(B-K+1)$ and $\underline{\mathbf{U}}_c^{\mathrm{comb}} \in \mathbb{C}^{M \times K}$ is given as

$$\underline{\mathbf{U}}_c^{\mathrm{comb}} = \left[\mathbf{u}_{i_c(1)} \ldots \mathbf{u}_{i_c(K)} \right]. \tag{12}$$

Here, $i_c(k)$ is the c^{th} beam index combination for user k such that $i_c(k) \neq i_c(k'), \forall k, \forall k', k \neq k'$.

In Algorithm 2, the for-loop at step 1 of the semilinear search algorithm tests the combinations from AP l to all users. At step 2, $\underline{\mathbf{U}}_c^{\mathrm{comb}}$ is the c^{th} combination for the precoding matrix with the available codeword combinations at the columns to be tested as given in the equations (11) and (12). At step the 6, maxvec operator returns the index of the maximum vector element. When the index combination that yields the maximum sum-rate is obtained at step 6, i.e., $\underline{\mathbf{U}}_{c^\star}^{\mathrm{comb}}$, $\underline{\mathbf{U}}^{\mathrm{comb}}$ is updated at step 8 as follows

$$\underline{\mathbf{U}}^{\mathrm{comb}} \leftarrow \underline{\mathbf{U}}^{\mathrm{comb}} \setminus \underline{\mathbf{U}}_c^{\mathrm{comb}}, \ \forall c \mid i_{c^\star}(k) = i_c(k'), \forall k, \forall k', k \neq k'. \tag{13}$$

Assume $K = 2$, $B = 3$, and no beam assignment is realized yet. Then the total number of beam combinations to be tested under BCC is 6, i.e., $C = 6$. Hence, the matrix set of beam combinations is initially given as

$$\underline{\mathbf{U}}^{\mathrm{comb}} = \left\{ [\underline{\mathbf{u}}_1 \underline{\mathbf{u}}_2], [\underline{\mathbf{u}}_1 \underline{\mathbf{u}}_3], [\underline{\mathbf{u}}_2 \underline{\mathbf{u}}_1], [\underline{\mathbf{u}}_2 \underline{\mathbf{u}}_3], [\underline{\mathbf{u}}_3 \underline{\mathbf{u}}_1], [\underline{\mathbf{u}}_3 \underline{\mathbf{u}}_2] \right\}.$$

Assume $c^\star = 4$, hence $\underline{\mathbf{U}}_{c^\star}^{\mathrm{comb}} = [\underline{\mathbf{u}}_2 \underline{\mathbf{u}}_3]$. At step 8, the matrix set of beam combinations is updated as follows

$$\underline{\mathbf{U}}^{\mathrm{comb}} = \left\{ [\underline{\mathbf{u}}_1 \underline{\mathbf{u}}_3], [\underline{\mathbf{u}}_2 \underline{\mathbf{u}}_1], [\underline{\mathbf{u}}_2 \underline{\mathbf{u}}_3] \right\}.$$

Finally, at step 9, $C = 3$ update is performed.

For the linear search algorithm, similar steps are executed except that in a search segment, the beam combinations between an AP and a user are tested each time as seen in the for-loops at steps 1 and 2 of Algorithm 3. At step 2 of Algorithm 3, $B_k, \forall k \in \mathcal{K}$, are initially equal to B. However, due to BCC, as codebooks are removed from users' codebook logs, the number of possible beam selections for user k is updated at step 10. At step 9, the selected codeword $\{\mathbf{u}_{kl}^\star\}$ is removed from the codebook logs of all others users k', $\mathcal{U}_{k'}$, for BCC. Hence, when the codeword combinations for the other users k' are being tested next time, the already assigned beam for user k does not exist in $\mathcal{U}_{k'}, \forall k', k' \neq k$ as an option to be tested. Although the BCC implementations are achieved in a single line at steps 8 and 9 of Algorithms 2 and 3, respectively, their coding implementations are significantly challenging [26].

At step 7 of Algorithm 4, the analog precoders are reset back to the random initializations that were determined at step 3. The order of searching process is determined by the $\mathbf{L}_{\mathrm{vec}}$ vector that is initialized at step 5 and updated at step 19 with a circular shift. For instance, during the first iteration of the for-loop at step 6, the search segment-1, i.e., AP 1, does the assignment first since $\mathbf{L}_{\mathrm{vec}} = [1, 2, \ldots, L]$. Whereas, during the second iteration of the for-loop, the search segment-2, i.e., AP 2, does the assignment first, since $\mathbf{L}_{\mathrm{vec}} = [2, \ldots, L, 1]$. When the loop at step 6 is finalized, the priority is given to the search segment l, i.e., AP l, with the highest sum-rate. To improve the sum-rate further at an increased cost, each segment can have a priority order. For simplicity, in this work, only one search segment is prioritized, e.g., if the search segment-2 is prioritized then

Algorithm 5 Pseudocode of the Benchmark Disjoint Linear-DL Algorithm

1: Random initializations of analog precoders $\mathbf{u}_{kl}^{0} \in \underline{\mathcal{U}}$, $\forall k, \forall l$ $\mathbf{u}_{kl}^{0} \neq \mathbf{u}_{k'j}^{0}, \forall l, \forall j, \forall k, \forall k', k \neq k'$ (with BCC)
2: $\mathcal{U}_k = \underline{\mathcal{U}}, \forall k$
3: $\mathbf{u}_{kl} = \mathbf{u}_{kl}^{0}, \forall k, \forall l$
4: **for** $l = 1 : L$ **do**
5:     **for** $k = 1 : K$ **do**
6:         **for** $b = 1 : B_k$ **do**
7:             $\mathbf{u}_{kl} = \underline{\mathbf{u}}_b$
8:             $\mathrm{DL}_{\mathrm{sum}}(b)$
9:         **end for**
10:         $[\sim, b^{\star}] = \mathrm{maxvec}\big[\mathrm{DL}_{\mathrm{sum}}(1) \ldots \mathrm{DL}_{\mathrm{sum}}(b) \ldots \mathrm{DL}_{\mathrm{sum}}(B_k)\big]$
11:         $\mathbf{u}_{kl}^{\star} = \underline{\mathbf{u}}_{b^{\star}}$
12:         $\mathcal{U}_{k'} \leftarrow \mathcal{U}_{k'} \setminus \underline{\mathbf{u}}_{b^{\star}}$, $\forall k', k' \neq k$ (with BCC)
13:         Update $B_{k'}, \forall k', k' \neq k$ (with BCC)
14:     **end for**
15: **end for**
16: $\mathbf{V}_l, \forall l$

the search segments $3, \ldots, L$, and, finally, 1 are executed in the written order.

As supported by the extensive numerical analyses in Section VII, the sum-rates and complexities of the proposed algorithms are given as follows: linear-IIS-rate (Algorithm 4) $>$ semilinear-II-rate (Algorithms 1 and 2) $>$ linear-II-rate (Algorithms 1 and 3) $>$ linear-II-DL (Algorithms 1 and 3 with DL metrics instead).

At an increased cost, a semilinear search option can be added to the step 11 in Algorithm 3 as well, i.e., a semilinear-IIS-rate algorithm. It can be clearly interpolated from the numerical results in Section VII that the semilinear-IIS-rate can achieve the highest sum-rate with an increased simulation duration. The selection feature can be more advantageous in asymmetric networks, i.e., the channels between APs and users are not uniformly distributed, than in symmetric networks. However, the user fairness needs to be carefully addressed [27].

To the best of our knowledge, in the literature and standards including IEEE 802.11ad and IEEE 802.15.3c, the beam training solutions do not consider the joint design of analog beam selection and digital precoder. Hence, our proposed solutions are benchmarked with the disjoint design given in Algorithm 5. As seen at step 16, the digital precoders are designed after the selections of analog beams are finalized. In contrast, in our proposed solutions, the digital precoders are designed at each iteration of the analog beam search process as seen at steps 3 and 4 of Algorithms 2 and 3, respectively.

Finally, the algorithms without BCC implementation can be obtained by inactivating the steps 8 and 9 in Algorithms 2 and 3, respectively, and also by inactivating the condition that the random initializations must satisfy BCC in the initialization steps 3, 3, and 1 of Algorithms 1, 4, and 5, respectively.

VI. SUPERVISED MACHINE LEARNING ALGORITHMS

Among supervised ML algorithms, SVM, MLP, and RFt are prevalent algorithms. These algorithms are powerful, flexible, and swift with regards to their good accuracies, many hyperparameters, and light computational complexities, respectively.

In this section, we discuss multi-output classification and the effect of BCC on it, and provide Scikit implementation details.

A. MULTI-OUTPUT CLASSIFICATION

Single-output classification methods select one class at a time. On the other hand, multi-output classification methods select multiple classes at once. In the literature, the latter method is also referred as multi-label multi-class classification method, or shortly, multi-label classification method.

Between an AP and all users, there are B^K beam options, i.e., the number of classes is B^K. The final decision vector at AP l in base-B notation can be given as

$$\mathbf{b}_l = \big[b_{1l} b_{2l} \ldots b_{kl} \ldots b_{K-1l} b_{Kl}\big], \tag{14}$$

where $b_{il} \in \{0, 1, \ldots, B-1\}$. For instance, consider a network with $B = 3$ and $K = 2$. There are $B^K = 9$ classes. In base-2 notation, these classes are labeled as $00, 01, 02, \ldots, 22$. Hence, beam selection becomes a $K = 2$-label $B = 3$-class classification problem.

The naive approach to solve multi-output classification problems is to build a decision model for each output. Hence, the problem is divided into K single-output classification problems. In the literature, the naive approach of treating each output independently from other outputs and solving a classification problem for each output is known as binary relevance. The disadvantage of this approach is that the correlations between the outputs, i.e., users, are neglected.

To accommodate the correlation between the labels, classifiers can be chained. In this approach, the first classifier is trained by the training data only and the label decisions of the first classifier are noted. Then, for the second classifier, the training data and the label decision of the first classifier are used as inputs. This is iterated until the last classifier, i.e., the last output is trained based on the training data and also, the label decisions of the previous classifiers. Clearly, classifier chains can exploit the correlation between the labels to some extent.

As outlined in [19], [28], binary relevance, classifier chains, and label power-set are known as problem transformation methods. Problem transformation methods transform the multi-label problem into several single-label problems and apply a single-label algorithm. On the other hand, algorithm adaptation methods, such as multi-label k-nearest neighbor (ML-kNN), AdaBoost, and RFt, extend the single-label solutions to multi-label solutions. Random forests can perform multi-label classification [29]. In other words, it can classify all outputs at once. At each leaf, all outputs are stored rather than only a single output. Then, the impurity metric, e.g., Gini, at a split is evaluated by averaging the impurities of the outputs at that split [30, 1.10.3 Multi-output problems]. Since the splits are optimized by considering all outputs, RFt can exploit the correlation between the outputs to some extent as well. In addition, a single model is generated by considering all labels. This significantly reduces the computational complexity compared to the naive approach of generating a model for each of the outputs as mentioned earlier.

Due to BCC, the decision vectors $\mathbf{b}_l$ (14) at the APs become correlated. In other words, the classification outputs of the solution to the problem (7) follows a pattern. For instance, when $B = K$, the decision vectors at APs are equal, $\mathbf{b}_l = \mathbf{b}_{l'}$, $\forall l$, $\forall l'$, $l \neq l'$. In this case, the overall decision output follows a fully repetitive pattern

$$\mathbf{b} = [\mathbf{b}_1 \dots \mathbf{b}_L] = [b_1 \dots b_k \dots b_K \dots b_1 \dots b_k \dots b_K], \quad (15)$$

where the subindex l is omitted since the decisions at user k are equal at all APs.

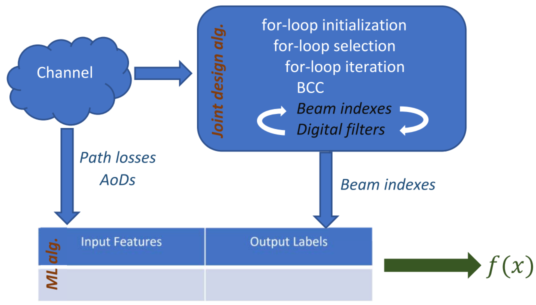
As illustrated by extensive numerical results in Section VII, RFt with the classifier chains can effectively exploit the correlated outputs occurred due to BCC. Since the outputs follow a full pattern when $B = K$, RFt with the classifier chains can achieve the original, i.e., 100%, sum-rate that is achieved by the beam selection algorithms. When $B > K$, it is observed that RFt with the classifier chains can achieve 99% of the original sum-rate in contrast to 63% when the classifier chains are not utilized with RFt.

BCC has a significant affect on the minimum number of beam search options depending on the network architecture. As seen in (15), the minimum number of beams in a cell-free network, $B_{\min}$, is K. In contrast, for an interference network, $B_{\min} = LK$. Hence, BCC can be beneficial in cell-free networks to reduce the beam search time. Moreover, if a DFT codebook is used, where $M = B$, the minimum number of transmit antennas is also significantly low.

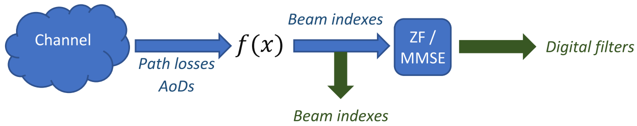
Finally, as mentioned earlier, BCC creates highly correlated outputs in cell-free networks compared to the other networks such as interference networks. Assume a network with $L = 2$, $K = 4$. For cell-free and interference networks, $B_{\min} = 4$ and $B_{\min} = 8$, respectively. Then, the overall decision outputs for cell-free and interference networks are given as

$$\mathbf{b}_{\mathrm{CF}} = [0\ 1\ 2\ 3\ 0\ 1\ 2\ 3] \quad (16a)$$
$$\mathbf{b}_{\mathrm{IN}} = [0\ 1\ 2\ 3\ 4\ 5\ 6\ 7], \quad (16b)$$

respectively. As clearly seen in (16), the output labels of interference network are completely random. Whereas the output labels of cell-free network follow a full repetitive pattern, hence, they are significantly correlated. As demonstrated in Section VII, RFt with classifier chains is the key approach to exploit the correlated outputs to retain 99-100% of the original sum-rate results achieved by the proposed well-constructed designs.



(a) Off-line ML training.



(b) Online ML testing followed by a digital filter design.

FIGURE 3. Off-line training and online testing of ML algorithms.

B. IMPLICIT JOINT DESIGN WITH MULTIPLE FEATURES

Off-line training and online testing of ML algorithms are achieved as follows. For each of the training and testing instances, $2LK$ path loss and AoD values between the APs and users are input as feature vectors. For off-line training, LK beam indexes between APs and users are used as output labels. The beam indexes can be determined by any of the joint design algorithms proposed in Section V. Then, ML algorithms efficiently approximate the input-output mapping functions of the beam selection algorithms by using the feature vectors and output labels. For online ML testing, the trained ML algorithms are executed to determine the output labels, i.e., the beam indexes, by using the approximate mapping functions. At this stage, as detailed in Section V-D, the effective channels are estimated so that a digital precoder, (9) or (10) is designed in the last step. The approach of beam selection by an ML algorithm and then designing a digital precoder is implicitly a well-constructed design since the ML algorithm is trained by a well-constructed design. As illustrated in Figure 3, both off-line training and online testing of ML algorithms are executed one-time, i.e., there are no for-loops. Moreover, the need of our proposed joint designs for frequent effective channel estimations, as detailed in Section V-D, are resolved by ML algorithms. For simplicity, in Figure 3(a), one channel training instance is displayed, hence the input-output table has a single row.

C. SCIKIT IMPLEMENTATION

For the implementations of SVM, MLP, and RFt algorithms, Scikit library [30] is used. These algorithms can be queried in [30] by the following method names: LinearSVC and SVC (both belong to the sklearn.svm module), MLPClassifier (belongs to the sklearn.neural_network module), and RandomForestClassifier (belongs to the sklearn.ensemble module). To access other classification methods under these modules, the module names can be

TABLE 1. Some abbreviations and notations used in the figures.

(Corresponding figures) LxKyMz init = x, iter = y (mm:ss) and (hh:mm)	 A network with $L = x$ APs, $K = y$ users, and $M = z$ antennas Algorithm with number of initializations and iterations are set to x and y, respectively The simulation durations in (minutes:seconds) and (hours:minutes), respectively
(Figs. 4 and 5) A - B - C A: Semilinear / linear / disjoint linear B: II / IIS C: Rate / DL	 Semilinear search algorithm (Section V-A) / linear search algorithm (Section V-B) / Naive disjoint linear design. All other algorithms are joint designs. Multiple initializations and iterations / multiple initializations, iterations, and selection Selection based on the rate metric - equation (6) / based on the direct link metric - equation (8)
(Fig. 6) w/ or w/o BCC w/ BCC init	 Algorithm with or without beam conflict control process presented in Section V-C Algorithm with only beam conflict control initializations, i.e., no beam conflict control process
(Fig. 7) A - B - C A: Linear, B: II C: Naive / Smart	 Explained earlier in the table The number of RF chains are naively fixed to a number / are adaptively set to a number based on a threshold value

queried as well, e.g., sklearn.ensemble can be searched in the top-right box [30].

As detailed earlier, multi-output classification problems can be solved by the naive binary relevance or the efficient classifier chain approach. These approaches can be queried by the wrappers MultiOutputClassifier and ClassifierChain, respectively, in [30].

For hyperparameter optimization, GridSearchCV from Dask-ML library [31] is used for speeding the grid search method. In addition to Dask-ML library, there are many other alternative open source projects available online.

For the interested reader, the complete codes of the proposed beam selection algorithms are shared online [26].

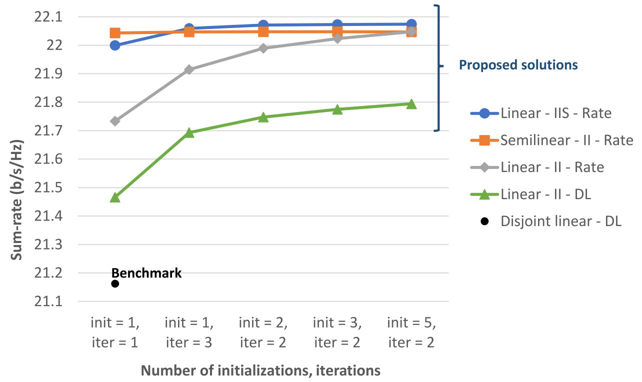
VII. NUMERICAL RESULTS

In this section, we present the sum-rate and simulation run time results for the proposed beam selection and ML algorithms. For all simulations, the channel center frequency and bandwidth are 28 GHz and 850 MHz, respectively, the path loss exponent is 2, the log-normal shadowing is 4 dB, the AP transmit power is 43 dBm, power spectral density of the white Gaussian noise is -174 dBm/Hz, and finally, the distances between APs and users vary uniformly between 95-105 m. The channel gains β_{kl}^{p} are drawn i.i.d. from $\mathcal{CN}(0, 1)$ distribution, and the number of paths, P, is assumed to be 1.

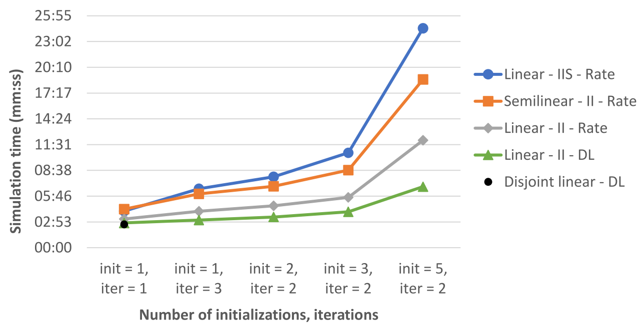
DFT codebook can be adopted as a hardware friendly orthogonal beamforming solution. The element with the row and column indexes (x, y), $x, y = 1, \ldots, M$ of the DFT codebook matrix is given by

$$\mathbf{D}(x, y) = \frac{1}{\sqrt{M}} e^{-\frac{j2\pi(x-1)(y-1)}{M}}, \tag{17}$$

where $j = \sqrt{-1}$. Note that the codeword b, i.e., $\underline{\mathbf{u}}_b$ in the codebook $\underline{\mathcal{U}}$ which is defined in Section III-A, is the b^{th} column of $\mathbf{D}$ in (17).



(a) Sum-rates.



(b) Simulation durations.

FIGURE 4. Sum-rates and simulation durations of the beam selection algorithms with different numbers of initializations and iterations for the L3K2M8 network.

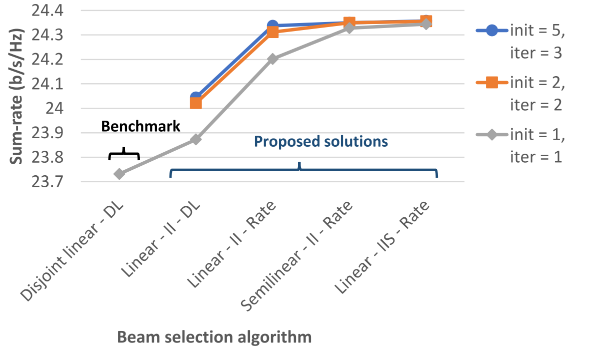
In Section VII-B, the numerical results of ML algorithms with classifier chains are presented. In all cases tested, there is a sharp contrast in performance, as mentioned earlier, when classifier chains are not and are used, i.e., 63% and 99%, respectively. Since including without classifier chain results in the figures can negatively impact the readability of the figures while not bringing beneficial information, they are not presented. The input features for ML algorithms are the sets of path losses and AoD. To avoid bias in the training phase, features are normalized before ML algorithms are executed.

For all simulations, a desktop computer with Intel i7-8700 CPU, 3.20 GHz, 16 GB RAM, with 6 cores and 12 logical processors is used. For the beam selection and ML simulations, MATLAB and Python with Scikit library are used, respectively. For the grid search tasks in ML algorithms, Dask-ML library is used.

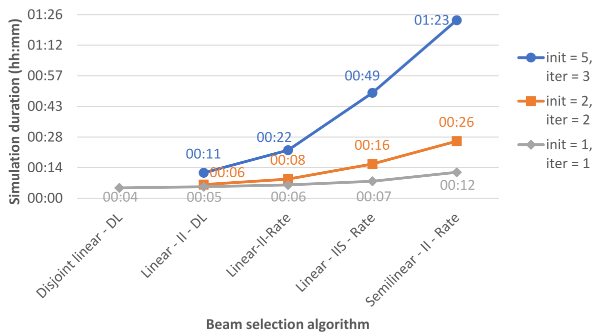
In Table 1, some abbreviations and notations used in the figures presented in the following sections are briefed.

A. BEAM SELECTION RESULTS

In Fig. 4, the sum-rate results and the simulation run times of the beam selection algorithms proposed in Section V are presented for different number of initializations and iterations. In the network, it is assumed there are $L = 3$ APs, $K = 2$ users, and $M = 8$ antennas at each AP which can be denoted by LLKKMM, i.e., the network parameter notations



(a) Sum-rates.



(b) Simulation durations.

FIGURE 5. **Sum-rates and simulation durations of beam selection algorithms with different numbers of initializations and iterations for the L3K2M16 network.**

are followed by their numerical values. The number of Monte Carlo (MC) runs, i.e., the number of channel realizations, to obtain the numerical results is 5×10^3 (5k).

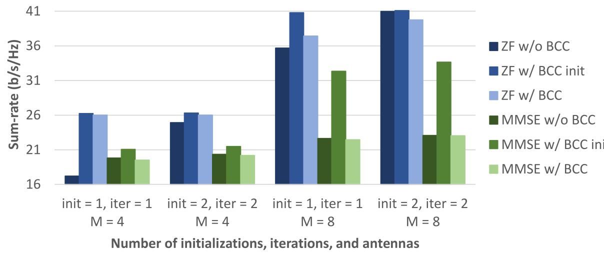
Clearly, as the beam selection algorithm becomes well-constructed from a disjoint design to the complete proposed solution in Algorithm 1, the sum-rate results improve. However, as seen in Fig. 4(b), the improvement is marginal against the cost of increased simulation durations. As the numbers of initializations and iterations increase, the linear-II-rate algorithm achieves almost the same results as the linear-IIS-rate algorithm with a tolerable increase in the simulation duration. Whereas the sum-rate gap to the disjoint linear-DL algorithm increases and the linear-II-DL algorithm still lags behind significantly.

In Fig. 5, the same beam selection algorithms are tested in the L3K2M16 network for again 5k MC runs. As the network size grows, the high complexity of semilinear algorithm starts to take effect and its simulation duration becomes the longest among all selection algorithms. linear-IIS-rate without multiple initializations and iterations achieves a remarkable sum-rate result with a tolerable increase in the simulation duration. This result stresses the influence of the first search segment, i.e., AP, to do the beam selection on the sum-rate.

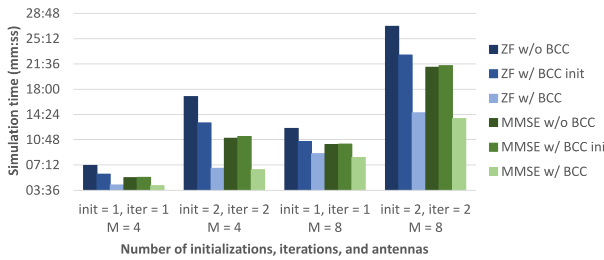
In Fig. 6, the effects of BCC on the sum-rate and simulation duration are demonstrated over varying number of initializations, iterations and antennas. For all results, linear-II-rate beam selection algorithm is used. Algorithms with and without the BCC implementation are denoted by w/ BCC and w/o BCC, respectively. For the algorithms without the BCC implementation, we also propose random initializations that satisfy BCC, i.e., BCC initializations. These algorithms are denoted by w/ BCC Init. In particular, for the w/ BCC Init algorithms, BCC condition for random initialization is active in step 3 of Algorithm 1 but the BCC implementation in step 9 of Algorithm 3 is inactive. As detailed next, our numerical results indicate that the three options, i.e., the algorithms with and without the BCC implementation, and with the BCC initializations, offer trade-offs between the sum-rates and simulation durations.

When there is no BCC implementation, the effective channel in (9b) can be low-rank and can have a high, i.e., poor, channel condition number. As seen in Fig. 6(a), ZF with the BCC implementation achieves a higher sum-rate than ZF without the BCC implementation until a point, i.e., init $= 2$, iter $= 2$, $M = 8$. BCC implementation reduces the number of beam combinations. Thus, BCC implementation is significantly effective in reducing the simulation duration as seen in Fig. 6(b). However, BCC implementation also reduces the search space so that the point that achieves a higher sum-rate can be excluded in the search process. In Fig. 6(a), from left to right along the x-axis, the search space increases since the numbers of random initializations and antennas are increasing. Hence, the advantage of larger search space becomes dominant over the advantages of higher channel rank and lower channel condition number. MMSE without the BCC implementation can achieve a higher sum-rate than MMSE with the BCC implementation due to the increased search space and increased robustness of MMSE against the low-rank and high channel condition number of the effective channel. In all cases, with the BCC implementation option reduces the simulation duration compared to without the BCC implementation option. The reduction is more distinct when multiple random initializations are used.

For the algorithms without the BCC implementation, BCC initialization can increase the sum-rate and decrease the simulation duration in general. In essence, this option benefits from the improved rank and channel condition number due to the BCC initialization, and also from increased search space although BCC implementation is not applied during the search processes. For both ZF and MMSE filters, BCC initialization achieves the highest sum-rates. For init $= 1$, iter $= 1$, $M = 8$, BCC initialization significantly improves the sum-rates of the ZF and MMSE filters, respectively, compared to ZF and MMSE without the BCC implementation. The simulation duration of ZF with BCC initialization is lower compared to ZF without BCC. However, for the MMSE filter, BCC initialization slightly increases the simulation duration compared to MMSE without BCC. For the MMSE filter, it is observed that although BCC initialization increases the rank, it also increases the channel condition number. Increased channel condition



(a) Sum-rates.



(b) Simulation durations.

FIGURE 6. Sum-rates and simulation durations of the linear-II-rate algorithm with BCC, without BCC, and with BCC initialization for the L4K4M4 and L4K4M8 networks.

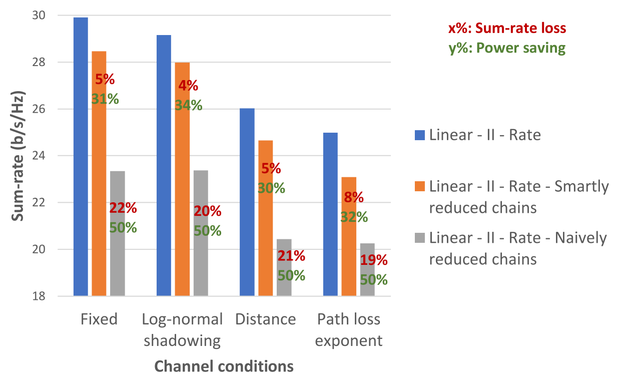


FIGURE 7. Sum-rates of the linear beam search algorithm in different channel conditions for the L4K4M4 network when the numbers of RF chains are full ($M_{rf} = K$), smartly reduced ($M_{rf} < K$ statistically), and naively reduced ($M_{rf} < K$ deterministically).

number causes numerical instability thus the simulation duration is increased.

In Fig. 7, the sum-rates are evaluated for varying channel conditions. For all results, linear-II-rate with a single initialization and iteration beam selection algorithm is used. As the channel conditions get poor, the marginal benefit of serving all users can be low considering the sum-rate gain versus the power cost. In poor channel conditions, shutting off some RF chains can lower the power consumption at a low sum-rate loss. For the naive approach in Fig. 7, denoted by linear-II-rate-naive, the number of RF chains are reduced to a fixed number, $M_{\text{rf}} = 2$. On the other hand, for the smart approach, denoted by linear-II-rate-smart, the number of RF chains are smartly reduced based on a threshold value determined by the path loss value in (4). In particular, the threshold is set to the mean minus $1/4$ power of variance of the path losses between AP l and all users. Users with the path losses below this threshold are served, and the others are not, thus the corresponding RF chains of those users are shut off. The blue bars in Fig. 7 indicate the case when all RF chains are in use, i.e., $M_{\text{rf}} = 4$. The left-most set of results are obtained when channel conditions are fixed to the certain values given in the beginning of this section. The 2$^{\text{nd}}$ and the following set of results from left are obtained when log-normal shadowing, distance, and path loss exponent vary with uniform distribution between 4-6 dB, 100-200 m, and 2-4, respectively. The bold red and green numbers inside the bars are the sum-rate loss and power saving percentages when compared to the blue bars. The power savings for the naive cases are always 50%, i.e., M_{rf} is set to 2 instead of 4. The results clearly indicate that by smartly shutting off the RF chains, significant power savings can be gained at a cost of small sum-rate losses. Similar to the selection feature explained earlier, smartly reducing the chains can be more advantageous in asymmetric networks while raising the fairness issue in the network.

As also demonstrated in Fig. 7, the benchmarks of the considered algorithms in our work are indifferent to varying channel conditions. Over varying channel conditions, the sum-rate losses of smart and naive approaches vary around 5% and 20% in average, respectively, and the power saving of smart approach varies around 30% in average. Only the benchmarks of the selection and smartly reduced chains features are expected to be different in asymmetric versus symmetric channels as mentioned earlier.

Our proposed solutions can substantially benefit from larger network sizes as detailed next. In Fig. 8, the sum-rates are evaluated for large network sizes. For the results, the path loss exponent, log-normal shadowing, and the distances between APs and users vary uniformly between 2-8, 4-10 dB, and 10-200 m, respectively. Due to the large network sizes, MC run is set to 100. For linear-II-rate, single initialization and iteration are used. As the network size grows, same conclusions presented earlier in this section can be drawn while the numerical gaps, i.e., benchmarks, between the results increase. As seen in Fig. 8, the advantage of joint design becomes significant over the naive disjoint design as the network size grows. As the network size grows, more complex algorithms are likely to benefit more than the simpler algorithms by using the rich degrees of freedom in larger networks.

The complexity of the linear search algorithm for the largest network size presented earlier in Fig. 6, i.e., L4K4M8, is only 128. On the other hand, the complexities of the networks L25K20M32 and L30K25M64 in Fig. 8 are 16k and 48k, respectively. Even for small network sizes, the semilinear algorithm presented in Section V-A becomes impractical. For the networks L4K3M16 and L4K4M8, the complexity of the semilinear algorithm is nearly 16k. Interested reader can refer to the spreadsheet in [26] for the complexities of varying network sizes. As detailed in Section V, the complexity evaluations are based on the searching principles without the extra features. If the multiple

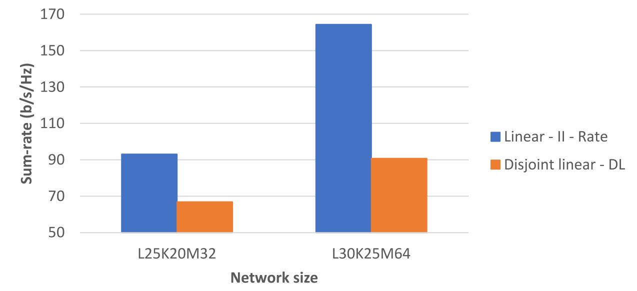


FIGURE 8. Sum-rates of the joint and disjoint linear beam selection algorithms for large network sizes. For the joint design, init = 1 and iter = 1.

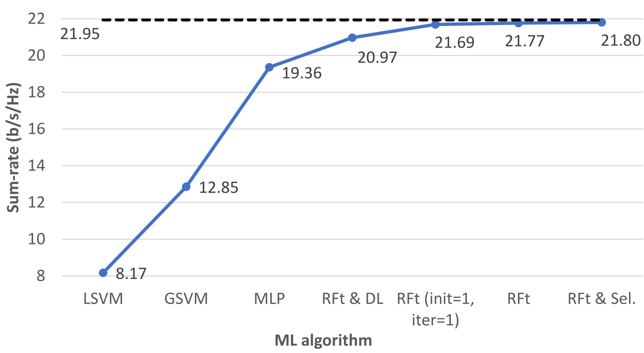


FIGURE 9. Sum-rates of ML algorithms for the L3K2M8 network. RFt & DL, RFt & Sel., and the rest are trained by the linear-II-DL, linear-IIS-rate, and linear-II-rate algorithms. Unless noted, init = 2, iter = 2 parameters are set.

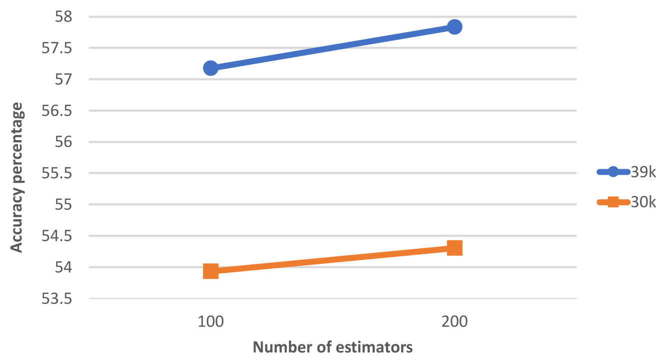


FIGURE 10. The performance plots of accuracy percentage versus number of training data and number of estimators, i.e., trees, for the RFt classification algorithm in the L3K2M8 network. RFt is trained by the linear-II-rate algorithm.

initializations, iterations, and the selection features (IIS) are included in an algorithm, its complexity value is multiplied by init$\times$iter$\times L$.

Next, we benchmark the sum-rates and run times of ML algorithms.

B. MACHINE LEARNING RESULTS

In Fig. 9, the sum-rate results of ML algorithms for the L3K2M8 network are presented. For the training and testing phases, 30k and 10k instances are executed, which give a 75% to 25% ratio of all data, respectively. The training and test data are independently generated. Nevertheless, all ML algorithms are passed through 5-fold cross-validation (CV), and also, through hyperparameter optimization stages.

To train and test the ML algorithms, based on the numerical results presented earlier, the linear-II-rate algorithm with 2 initializations and iterations (i.e., init $= 2$, iter $= 2$) is chosen since it can achieve sum-rates close the best results presented earlier within short simulation durations. This choice is primarily based on saving time for the increased durations of training and testing of ML algorithms and note that the benchmarking of ML algorithms is independent of beam selection algorithms.

For the training of the linear SVM (LSVM), Gaussian (GSVM), MLP, and RFt algorithms, linear-II-rate beam selection algorithm is used. Except RFt with init $= 1$ and iter $= 1$, the numbers of initializations and iterations are set to 2. For the training of RFt&DL and RFt&Sel., linear-II-DL and linear-IIS-rate beam selection algorithms are used, respectively.

In Fig. 9, the dashed line on top is the sum-rate obtained by the proposed linear-IIS-rate beam selection algorithm, i.e., the maximum original sum-rate that can be achieved by the ML algorithms. Clearly, upon with the utilization of the RFt algorithm, sum-rates between 96-99% of the original results can be achieved.

As noted earlier, MATLAB and Python languages are used for the beam selection and ML algorithms. Nevertheless, we can conclude that RFt can achieve remarkable sum-rate results within tolerable durations compared to conventional beam selection algorithms as follows. The total simulation durations of training and testing phases for RFt, MLP, GSVM, and LSVM algorithms are about 01:30, 01:00, 15:00, and 00:05 (mm:ss), respectively. Hence, we can conclude that the RFt algorithm has the best accuracy with a reasonable speed. The simulation duration for the dashed line in Fig. 9 which is produced via the linear-IIS-rate beam selection algorithm is about 12:00. Indeed, even more well-constructed algorithms that have longer durations than the linear-IIS-rate algorithm are admissible to achieve higher sum-rates since the beam selection algorithms are executed off-line to generate the data that train the ML algorithms, and the simulation durations of the ML algorithms are unaffected by the training beam selection algorithms. The results in Figs. 8 and 9 clearly advocate that the more well-constructed algorithm is used for ML training, the more advantage over naive disjoint design can be achieved in large network sizes.

The classification accuracies of all ML algorithms can be improved by increasing the number training data as seen in Fig. 10, where the accuracy percentage of RFt is shown to increase as the training data is increased from 30k to 39k. Increasing the number of estimators, i.e., trees, in the case of RFt can also significantly improve the classification accuracy of the algorithm. Higher classification accuracies of ML algorithms result in higher sum-rates in return.

Among the ML algorithms, the highest accuracy is achieved by RFt&DL which is 96%. Other RFt algorithms that use sum-rate as the beam selection metric achieves about 55% accuracy. The accuracies of MLP, GSVM, and SVM are about 45%, 25%, and 20%, respectively. The reason for high accuracy of RFt&DL is that the input features for the ML algorithms are the sets of path losses and AoD. Thus, the DL power, as the beam selection metric matches better to these input features. Although the accuracy of RFt&DL is high, linear-II-DL beam selection algorithm has a lower sum-rate, as seen in Figs. 4(a) and 5(a), compared to the other selection algorithms that use sum-rate as the beam selection metric. Thus, with 55% accuracy, RFt algorithms with the sum-rate beam selection metric can achieve higher sum-rate results as seen in Fig. 9. Our numerical results indicate that RFt algorithms with the sum-rate beam selection metric can recognize alternative beam combinations that achieve high sum-rate results although the selected beam combination by the RFt algorithm may not match to the beam combination that is originally chosen by the beam selection algorithm. When $M = K$, RFt algorithm can retain the original sum-rate by 100% although its accuracy is still unremarkable because, again, the RFt algorithm can identify alternative beam combinations different than the combinations chosen originally by the beam selection algorithm so that RFt algorithm can still achieve high sum-rate results. In fact, if the exhaustive search algorithm, i.e., centralized algorithm, can be used instead, then the RFt algorithm with a 55% accuracy rate may not retain 99% of the original sum-rate achieved by the exhaustive algorithm since the beam combination selections by the exhaustive search algorithm are optimum. However, as mentioned earlier, the exhaustive search algorithm is impractical due to its high complexity.

As demonstrated in Fig. 11, the proposed joint designs and ML algorithms can be effective in multi-path mm-wave channels which typically have one strong line-of-sight (LoS) path and a few reflected paths. A K-factor of 10 dB implies that the LoS path is nearly 10 dB stronger than the combined power of all reflected paths. Hence, analog beamforming can be effective by aligning the beam with the strong LoS path and ignoring the remaining paths. Similarly, ML algorithms can be effective when the remaining paths are ignored. By ignoring the remaining paths, firstly, the length of input feature vector is reduced by P, e.g., instead of using LKP path losses and AoDs, LK of them are used as input features. Secondly, the data imbalance problem is avoided, for instance, when the absolute channel gains $|\beta_{kl}^{p}|$ and path losses are used as input features. In Fig. 11, the sum-rate and accuracy percentage results of linear-II-rate (init = 1, iter = 1, and MMSE digital precoder) and RFt (200 estimators) algorithms are presented for the L3K2M8 network where $P = 3$ is assumed. For the RFt algorithm, the features regarding the strong LoS paths are considered. For both K-factor values 10 and 6 dBs, the RFt algorithm achieves almost the same sum-rate results as the linear-II-rate algorithm although only the features of strong LoS paths are considered. This implies that the RFt algorithm can effectively identify the alternative solutions as explained earlier in Fig. 9. On the other hand, the accuracy performance of RFt is challenged at low K-factor values. Thus, more MC runs and input features are needed for the training and testing phases of ML algorithms. For the results in Fig. 11, the training and test instances are again set to 30k and 10k, respectively. But, the absolute channel gains $|\beta_{kl}^{p}|$ of the strong LoS paths are included in the feature vectors in addition to the path losses and AoDs for improving the accuracy percentages. Nevertheless, compared to the accuracy percentages presented in Fig. 10 where $P = 1$ is assumed, the achieved accuracy percentages are lower in Fig. 11. This implies that the number MC runs need to be tuned based on the K-factor value to achieve the target accuracy percentages. In short, as demonstrated in this section, for ML algorithms, tuning the number of MC runs and number of input feature types (e.g., absolute channel gains, path losses, and AoDs) depending on the K-factor value while considering only the strong LoS paths can be an effective solution. The proposed linear-II-rate algorithm can be effectively implemented in multi-path channels as demonstrated in Fig. 11. However, there can be situations where the reflected paths can impact the analog beamforming design. These situations can be addressed in a future work.

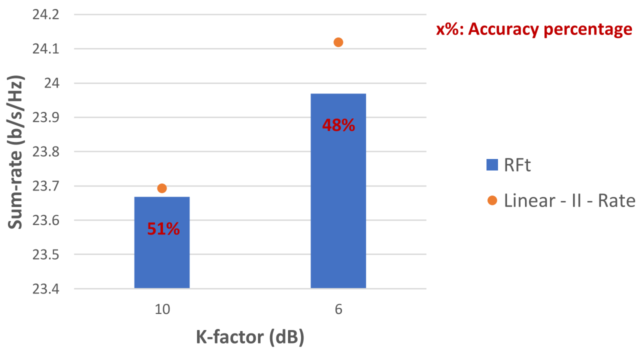


FIGURE 11. Sum-rates and accuracy percentages over varying K-factor values for the RFt and linear-II-rate algorithms in the multi-path ($P = 3$) L3K2M8 network.

VIII. CONCLUSION

In this work, we propose joint design algorithms of analog beam selection (based on the sum-rate metric) and digital precoders. The joint designs are well-equipped with multiple initializations, iterations, and selection features as well as with the BCC implementation. Hence, the network sum-rate gains are much higher compared to the naive disjoint design of analog beam selection (based on the DL power metric) and digital precoders. We show that BCC initialized algorithms without the BCC implementation can achieve the best network sum-rates compared to the algorithms with and without the BCC implementation, and also, BCC initialization can reduce the simulation durations. Finally, in poor channel conditions, RF chains can be selectively shut off to save significant power consumptions at the cost of low network sum-rate losses. Next, we propose supervised ML algorithms that are trained by the beam selection decisions obtained from the well-constructed algorithms. The numerical results obtained via the RFt algorithm are promising since it can retrain 99-100% of the original sum-rate results achieved by the proposed joint design algorithms.

Selectively shutting off the RF chains can only save power consumptions. As a future endeavor, further savings, e.g., chip area, can be achieved by serving more users than the number of RF chains.

REFERENCES

[1] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[2] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 651–663, Sep. 2019.

[3] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.

[4] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson, "Local partial zero-forcing precoding for cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4758–4774, Jul. 2020.

[5] Y. Long, Z. Chen, J. Fang, and C. Tellambura, "Data-driven-based analog beam selection for hybrid beamforming under mm-Wave channels," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 2, pp. 340–352, May 2018.

[6] C. Antón-Haro and X. Mestre, "Learning and data-driven beam selection for mmWave communications: An angle of arrival-based approach," *IEEE Access*, vol. 7, pp. 20404–20415, 2019.

[7] Y. Han, S. Jin, J. Zhang, J. Zhang, and K. K. Wong, "DFT-based hybrid beamforming multiuser systems: Rate analysis and beam selection," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 514–528, Jun. 2018.

[8] Z. Xiao, P. Xia, and X. Xia, "Codebook design for millimeter-wave channel estimation with hybrid precoding structure," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 141–153, Jan. 2017.

[9] C. Liu, M. Li, S. V. Hanly, I. B. Collings, and P. Whiting, "Millimeter wave beam alignment: Large deviations analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1619–1631, Jul. 2017.

[10] J. Choi, "Beam selection in mm-Wave multiuser MIMO systems using compressive sensing," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2936–2947, Aug. 2015.

[11] C. Chen, C. Tsai, Y. Liu, W. Hung, and A. Wu, "Compressive sensing (CS) assisted low-complexity beamspace hybrid precoding for millimeter-wave MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1412–1424, Mar. 2017.

[12] B. Li, Z. Zhou, H. Zhang, and A. Nallanathan, "Efficient beamforming training for 60-GHz millimeter-wave communications: A novel numerical optimization framework," *IEEE Trans. Veh. Technol.*, vol. 63, no. 2, pp. 703–717, Feb. 2014.

[13] K. Ho and S. Tsai, "A novel multiuser beamforming system with reduced complexity and beam optimizations," *IEEE Trans. Wireless Commun.*, vol. 18, no. 9, pp. 4544–4557, Sep. 2019. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9017927

[14] X. Sun, C. Qi, and G. Y. Li, "Beam training and allocation for multiuser millimeter wave massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1041–1053, Feb. 2019.

[15] W. Wu, D. Liu, X. Hou, and M. Liu, "Low-complexity beam training for 5G millimeter-wave massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 361–376, Jan. 2020.

[16] Y. Wang, A. Klautau, M. Ribero, A. C. K. Soong, and R. W. Heath, "MmWave vehicular beam selection with situational awareness using machine learning," *IEEE Access*, vol. 7, pp. 87479–87493, 2019.

[17] X. Song, S. Haghighatshoar, and G. Caire, "A scalable and statistically robust beam alignment technique for millimeter-wave systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4792–4805, Jul. 2018.

[18] Y. Gao, M. Khaliel, F. Zheng, and T. Kaiser, "Rotman lens based hybrid analog-digital beamforming in massive MIMO systems: Array architectures, beam selection algorithms and experiments," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 9134–9148, Oct. 2017.

[19] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," *Mach. Learn.*, vol. 85, pp. 333–359, Jun. 2011.

[20] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5289–5304, Oct. 2017.

[21] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE J. Sel. Areas Commun.*, vol. SAC-5, no. 2, pp. 128–137, Feb. 1987.

[22] A. M. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 1, pp. 156–166, Jun. 2007.

[23] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2212–2223, Jun. 2015.

[24] Y. Ren, Y. Wang, C. Qi, and Y. Liu, "Multiple-beam selection with limited feedback for hybrid beamforming in massive MIMO systems," *IEEE Access*, vol. 5, pp. 13327–13335, 2017.

[25] C. Hu, L. Dai, T. Mir, Z. Gao, and J. Fang, "Super-resolution channel estimation for mmWave massive MIMO with hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8954–8958, Sep. 2018.

[26] C. M. Yetis. *GitHub Repository*. Accessed: Jun. 21, 2021. [Online]. Available: https://github.com/DrCMY/JointDesigninmmWaveCellFreeNetworks

[27] Y. Rong and Y. Hua, "Space–time power scheduling of MIMO links—Fairness and QoS considerations," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 2, pp. 171–180, Apr. 2008.

[28] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *Int. J. Data Warehouse Min.*, vol. 3, no. 3, pp. 1–13, 2007.

[29] F. Liu, X. Zhang, Y. Ye, Y. Zhao, and Y. Li,"MLRF: Multi-label classification through random forest with label-set partition," in*Proc. Int. Conf. Intell. Comput.*,Cham, Switzerland, Aug. 2015, pp. 407–418.

[30] *Scikit-Learn: Machine Learning in Python*. Accessed: May 17, 2020. [Online]. Available: https://scikit-learn.org

[31] *Dask-ML: Scalable Machine Learning*. Accessed: May 17, 2020. [Online]. Available: https://ml.dask.org

CENK M. YETIS (Member, IEEE) received the B.Sc. degree in electronics engineering from Isik University, Istanbul, Turkey, in 2001, and the M.Sc. degree in telecommunications engineering and the Ph.D. degree in satellite communications and remote sensing from Istanbul Technical University in 2004 and 2010, respectively. From 2003 to 2007, he was with Turk Telekom (formerly Avea), one of the top three wireless services providers in Turkey, where he held rotational responsibilities in operation and planning groups. From 2007 to 2010, he was a Visiting Researcher with Ohio State University and the University of California at Irvine, Irvine. From 2010 to 2019, he held academic positions in Universities and Research Centers in Hong Kong, Singapore, Turkey, and Taiwan. In 2019, he joined Lund University, Sweden, as a Researcher to extend his research and teaching activities. In 2017, he received a two year Academia Sinica Fellowship and the Scientific and Technological Research Council of Turkey (TUBITAK) Scholarship from 2005 to 2009. His research interests include signal processing, information theory, communication theory, and optimization theory for wireless communications.

EMIL BJÖRNSON (Senior Member, IEEE) received the M.S. degree in engineering mathematics from Lund University, Sweden, in 2007, and the Ph.D. degree in telecommunications from the KTH Royal Institute of Technology, Sweden, in 2011.

From 2012 to 2014, he held a joint postdoctoral position with the Alcatel-Lucent Chair on Flexible Radio, SUPELEC, France, and the KTH Royal Institute of Technology. He joined Linköping University, Sweden, in 2014, where he is currently an Associate Professor. In September 2020, he became a part-time Visiting Full Professor with the KTH Royal Institute of Technology. He is dedicated to reproducible research and has made a large amount of simulation code publicly available. He performs research on MIMO communications, radio resource allocation, machine learning for communications, and energy efficiency. He has performed MIMO research for over ten years, his papers have received more than 10 000 citations, and he has filed more than 20 patent applications. He is a host of the podcast Wireless Future and he has a popular YouTube channel. He has authored the textbooks *Optimal Resource Allocation in Coordinated Multi-Cell Systems* (2013), *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency* (2017), and *Foundations of User-Centric Cell-Free Massive MIMO* (2021).

Dr. Björnson has received the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA, the 2015 Ingvar Carlsson Award, the 2016 Best Ph.D. Award from EURASIP, the 2018 IEEE Marconi Prize Paper Award in Wireless Communications, the 2019 EURASIP Early Career Award, the 2019 IEEE Communications Society Fred W. Ellersick Prize, and the 2019 IEEE Signal Processing Magazine Best Column Award. He also coauthored papers that received Best Paper Awards at the conferences, including WCSP 2009, the IEEE CAMSAP 2011, the IEEE SAM 2014, the IEEE WCNC 2014, the IEEE ICC 2015, and WCSP 2017. He has also been a guest editor of multiple special issues. He has been on the editorial board of the IEEE TRANSACTIONS ON COMMUNICATIONS since 2017. He has been a member of the Online Editorial Team of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS since 2020.

PONTUS GISELSSON received the M.Sc. and Ph.D. degrees from Lund University, Sweden, in 2006 and 2012, respectively. From 2013 to 2014, he held a postdoctoral position with Stanford University. His research interests include mathematical optimization and its wide range of applications, e.g., in machine learning, control, signal processing, and wireless communication. He is an Associate Professor with the Department of Automatic Control, Lund University. He received the Young Author Prize at the Advanced Control of Chemical Processes IFAC Symposium in 2012, the Young Author Prize at the IFAC World Congress in 2014, and the Ingvar Carlsson Award from the Swedish Foundation for Strategic Research in 2015.