Manuscript ID OJCOMS-00266-2020

Dear Editor and Reviewers,

Thank you very much for coordinating the reviews and your insightful comments. We have thoroughly revised the manuscript based on your suggestions. In the following, we provide point-by-point replies to your comments, and our revised manuscript is attached separately. For your convenience, the corresponding changes in the manuscript are addressed and briefed in this letter as well. To keep the response letter length limited, only some of the changes in the revised manuscript are rewritten in the response letter for the sake of completeness.

We look forward to your reviews and decision on our revised manuscript.

Response to Editor

Highlights

As suggested by the editor and the reviewers, we have significantly revised the manuscript. Some of the major changes are as follows:

- Based on the comment of Reviewer-1, we further analyze the effect of beam conflict control (BCC) on the sum-rate and simulation duration and we make new interesting conclusions which are discussed via the new added Fig. 6.
- Based on the 12$^{\text{th}}$ comment of Reviewer-2, we demonstrate new sum-rate results over varying channel conditions in the new added Fig. 7. In addition, based on the channel conditions of the users, we adaptively shut off some of the radio frequency (RF) chains to reduce the power consumption significantly at a low sum-rate loss as demonstrated in Fig. 7. The new numerical results in Fig. 7 also address the 4$^{\text{th}}$ comment of Reviewer-2.
- Based on the 7$^{\text{th}}$ comment of Reviewer-2, we have added the new Section VI.B and the new Fig. 3 to clearly explain how the off-line training and online testing of the machine learning (ML) algorithms are achieved. In this section, we emphasize the importance of ML algorithms since they can achieve the sum-rate performances of our well-constructed designs at low computational complexities.
- Based on the 8$^{\text{th}}$ comment of Reviewer-2, we have added the new Fig. 1 which can help the notations and problem setup in the manuscript to be rapidly comprehended.

Details

1. The considered scenario seems strange to me. In particular at mm-waves array with multiple elements can be considered also at the users' side, see for example reference [2]. For example, in reference A.1, a beam alignment algorithm is designed considering the beams at the BS and MS sides. I encourage the authors considering this, more general, case.

Reply: We thank the editor for this rightful comment. We definitely agree that assuming multi-antenna users can be a more adequate setting. The main concern we have for having multi-antenna users is the increased computational complexity. As noted in the manuscript, we refer to a joint design with multiple features as a well-constructed design. The multiple features include multiple initializations, iterations, selection features, and finally, BCC. The already significantly high computational complexities of the proposed algorithms with the single antenna users would even increase much more by assuming multi-antenna users. When multi-antenna users are considered, joint designs between the access points (APs) and users can be more adequate solutions. This additional joint design layer on top of our proposed joint designs with multiple features can significantly increase the computational complexities. Such designs can be carefully considered in a future work for achieving improved sum-rate results at the cost of acceptable increased complexities. In such a future work, our current proposed algorithms can be the benchmark as follows. For reduced complexity, a disjoint design between the APs and users can be proposed. In the first stage, all users select beams. Then, our proposed algorithms at the APs can be executed.

Finally, with multi-antenna users, the computational complexities of the ZF and MMSE digital precoders at the APs will also significantly increase due to the increased matrix dimensions and block diagonal matrices since APs jointly serve all users. This would oppose our assumption of low-cost APs with simple components. In that case, base stations rather than APs can be considered in a future work.

In cell-free networks, we have multiple APs and users. Hence, the complexity is significantly higher than the broadcast [A1] and point-to-point [A2,A3] settings. For our proposed linear beam searching algorithm, the complexity of cell-free network with single antenna users (our setup), cell-free network with multi-antenna users (the suggested setup), and broadcast network with multi-antenna APs and multi-antenna users (e.g., the setup in [A1]) are 32000, 128080, and only 1360 when the L100K20M16N4 (cell-free network with multi-antenna users), L100K20M16N1 (cell-free network with single antenna users), and L1K20M16N4 (broadcast network with multi-antenna users) networks are considered, respectively. Finally, the beam selection for single antenna users is also often addressed in the literature such as in [20,23].

2. The assumption of perfect CSI used in section V.D to design very standard digital beamformers (ZF and MMSE) is again suspicious. If you perfectly know the channel, a hybrid (digital+analog) beamforming design can be obtained in different ways, also present in literature, see for example references A.2, A.3. Please comment and clarify the idea behind this assumption.

[A.1] X. Song, S. Haghighatshoar and G. Caire, "A Scalable and Statistically Robust Beam Alignment Technique for Millimeter-Wave Systems," in IEEE Transactions on Wireless Communica-

tions, vol. 17, no. 7, pp. 4792-4805, July 2018.

[A.2] H. Ghauch, T. Kim, M. Bengtsson, and M. Skoglund, Subspace estimation and decomposition for large millimeter-wave MIMO systems, IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 528542, Apr. 2016.

[A.3] O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Spatially sparse precoding in millimeter wave MIMO systems, IEEE Trans. Wireless Commun., vol. 13, no. 3, pp. 14991513, Mar. 2014.

Reply: We appreciate this point being addressed by the editor. We would like to continue the discussion in two parts: CSI and perfect CSI. As we mention in our manuscript, the standard digital precoders can be preferred along with the analog beam selection via beam training for hybrid design in mm-wave channels for their low computational complexities [5-7]. As in [5-7], analog beam selection is achieved via beam training in our work. As in [5-7], we also do not assume CSI during the beam selection process in our work. In works that propose hybrid designs with beam training, such as in [5-7] and our work, effective CSI needs to be estimated at the end of the beam selection process to design only the digital precoders. In contrast, in A.2 and A.3, beam training method is not utilized. In A.2, analog precoders are designed based on the available CSI. Whereas in A.3, analog precoders are designed based on the CSI estimation. To emphasize again, in our work, we utilize beam training to select the beams from a codebook, and perfect effective CSI, i.e., equation (9b) in the revised manuscript, is needed only at the end of beam selection process to design the digital precoders. As explained in the next paragraph, the digital precoders cannot directly see the channels, but the channels multiplied by the analog precoders, i.e., effective CSI (from the perspective of the digital precoders). Since different analog precoders yield different effective CSI, after the analog precoders are selected via training without any CSI, effective CSI are estimated to design the digital precoders.

In Section V.D, we have added a new paragraph starting with the sentence "The digital precoders in (9) and (10) require local baseband CSI ...". As we detail in the new paragraph, our joint designs require more frequent effective CSI than the disjoint designs. For completeness, we rewrite the corresponding paragraph here:

"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."

Finally, in this work, we consider perfect effective CSI is available at the end of the beam selection process to design the digital precoders. The research on joint analog beam selection (via beam training) and digital precoder design with imperfect effective CSI (for only digital precoder design) can be a future work.

Thanks to the editor's comment, we believe the importance of ML algorithms is emphasized more strongly in the revised manuscript. Our proposed well-constructed designs have superior sum-rate results as seen in Figures 4(a) and 4(b) at the cost of major challenges including the multiple for-loops and frequent estimation of effective channels.

We detail these points in the new Section VI.B and in the new Fig. 3. By using off-line training of ML algorithms, the $f(x)$ function that maps the input-output relation of feature vectors (i.e., path losses and angle of departures-AoDs) and beam indexes is obtained via our well-constructed joint designs with multiple features. Hence, by one-time, i.e., no for-loops, execution of an online ML algorithm, the beam indexes are obtained. At this point, as in the other disjoint designs proposed in the literature and standards, estimation of the effective channels is needed for one-time digital precoder design as the last step.

In short, when our proposed ML algorithm is executed online,

- sum-rate performances of our ML algorithm are almost same as well-constructed algorithms. Well-constructed algorithms are computationally too expensive, but have superior sum-rates compared to naive disjoint design as widely proposed in the literature and standards that consider hybrid architectures with beam training,

- at low computational complexities, beams are selected one-time by our ML algorithm. Then based on the one-time effective CSI, the digital precoders are designed one-time also, and finally,

- only path losses and AODs are needed as input features.

3. As suggested by the reviewers, a careful revision of the notation, of the contribution, and of the quality of presentation of

the work should be done. I suggest also adding a Notation subsection in the introduction in order to improve the readability of the paper.

Reply: We carefully read your and the reviewers' comments. We believe the notation, contribution, and the quality of the presentation are significantly improved after following all suggestions. Per your valuable suggestion, we added the notation subsection in Section II.

Response to Reviewer 1

I have only one comment to make and it regards the fig. 4 in numerical results. To be precise, the authors compare the ZF with BCC and MMSE without BCC, but in my opinion it would be better to compare all four cases, ZF with and without BCC and MMSE with and without BCC.

Reply: We thank the reviewer for pointing out this important detail. Per your suggestion, after carefully analyzing the BCC effects on ZF and MMSE further, we have reached to new and interesting conclusions. The corresponding new results are presented in Fig. 6. The new results are discussed in detail in 3 paragraphs that start with the sentence "In Fig. 6, the effects of BCC on the sum-rate and simulation duration ...". Here, we summarize the major changes in Fig. 6 for the brevity of response letter:

- We plot all four cases, ZF& MMSE with and without BCC.
- Moreover, we propose a new approach, and plot the results for both ZF and MMSE. In this approach, BCC during the beam selection process is not implemented. However, the random initializations of the analog beam precoders satisfy BCC. We briefly call this approach as ZF or MMSE w/ BCC Init (as in short form of BCC initialization). The long naming would be ZF or MMSE without BCC implementation but with BCC initialization, e.g., ZF& MMSE - w/o BCC & w/ BCC Init.

Response to Reviewer 2

First of all, thank you for making your code and results publicly available.

Reply: Thank you. We slightly updated the codes and uploaded the new versions of the codes on GitHub repository.

1. Section 2 introduces the solution and the methods used without a proper and clear introduction of the problem and the settings and is thus hard to understand at a first read.

Reply: We thank the reviewer for this point of view. In the earlier manuscript, Section I was significantly shorter than Section II. We use this opportunity in the revised manuscript. We now overview the problem and settings, and also briefly mention the proposed solutions at the end of Section I as a brief introduction to Section II. Moreover, many details in Section II are moved to later sections since the details can be comprehended better after Section III where the system model is introduced in detail. In short, in Section I, now we clearly overview the problem, settings, and solutions as an introduction Section II where the solutions and methods are summarized. The details of solutions and methods are moved to the sections after Section III where we explicitly state the problem formulation in (7). In short, in the revised manuscript, we gradually introduce the details of the problem and settings, thus also the corresponding solutions and methods throughout the sections, i.e., from overviews to fine details as progressed through the sections.

2. It is unclear why classifier chains are used in combination with random forests, an algorithm which is natively able to handle multi-label classification. Could you please elaborate on that?

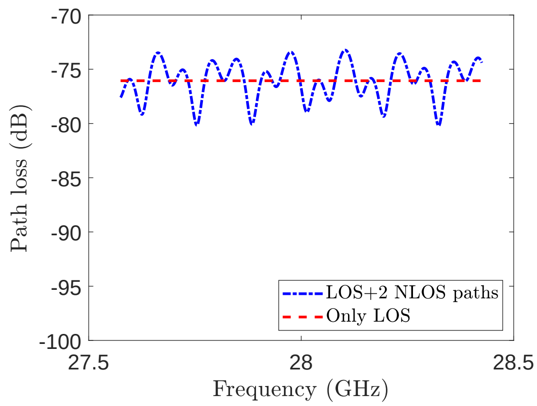
Reply: As detailed in the manuscript, BCC introduces correlations in the output labels. Based on our numerical results, random forest (RFt) can exploit the correlations so that 63% of the original sum-rate results are achieved whereas for the RFt with classifier chains, 99% are achieved. In the previous manuscript, we explained only how classifier chains can exploit the correlations. In the revised manuscript, now we also explain how RFt can exploit the correlations in the paragraph that starts with the sentence "As outlined in [19],[27], binary relevance, classifier chains, and label power-set are known as ...". As far as we see, the RFt algorithm, where the impurities are averaged, is not sufficient alone, and thus, the approach of classifier chains is needed as well. By wrapping the RFt and classifier chains, multi-label decisions are chained to each other by inputting the decision of the previous label as an input feature for the decision of the next label iteratively. Hence, by averaging the impurities via RFt and by chaining the output decisions via classifier chains, the correlated output labels due to BCC are exploited more strongly than single-handedly using RFt. We also consider the possibility that classifier chains with random forests can have the ensembles of classifier chains (ECC) effect which can perform much better than the stand-alone classifier chains as noted in [19].

3. Assuming a frequency flat channel in mmwave communication might be too optimistic given the typical bandwidth used in these systems. What type of system are you considering? In Section 7, a channel with a 28 GHz carrier frequency over 850 MHz is considered. Considering this as a narrow-band channel where only flat fading happens can be dangerous.

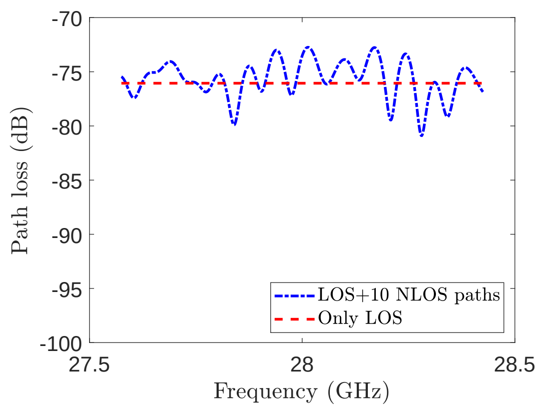
Reply:

It can be shown that the line-of-sight (LOS) path is strongly dominating, thus the narrow-band solution can be expected to work very well over a large bandwidth. In Fig. 1* of this document, we demonstrate that the channel is mainly dominated by the LoS path, and the non-line-of-sight (NLOS) paths add only a few dB of fluctuations on top of the LOS path. To obtain the results in Fig. 1, we use the simulation setup given in the manuscript. Moreover, for the length of a NLOS path, we choose it to be equal to LOS path length plus a random number between 0 and 10% of the LOS path length. In Fig. 2*, we also vary the NLOS path lengths for a channel with a LOS and 10 NLOS paths. Clearly, for all cases, channel fluctuations do occur without deep fades or similar effects. Needless to say, practical channels will never be perfectly frequency flat but it is a common practice in mm-wave communications to design algorithms for this special case, motivated by the fact that the frequency variations are small.

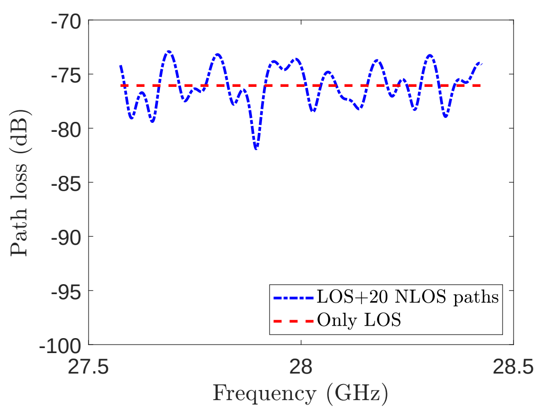
For low complexity, we can assume that our proposed solutions are implemented independently at each frequency of a wideband system which would be a suboptimal solution. We kindly note that joint analog beam selection and digital precoder design with multiple features, i.e., well-constructed design, over multiple frequencies can bring new opportunities to improve the sum-rate performance and can also bring further challenges to be addressed such as increased computational complexity. Our proposed solutions can be definitely extended in this direction with increased complexities. Therefore, these challenges in wideband systems to improve the sum-rates at reduced costs can be addressed in a future work.



(a) 2 NLOS paths.

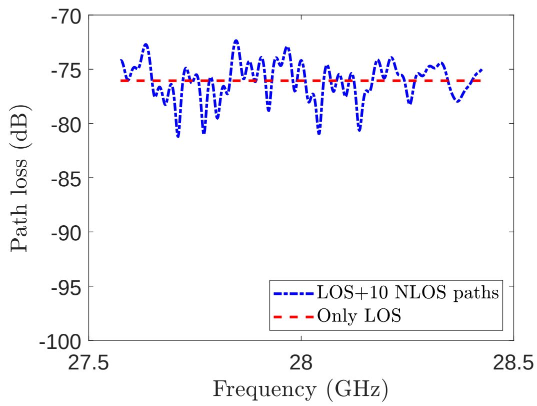


(b) 10 NLOS paths.

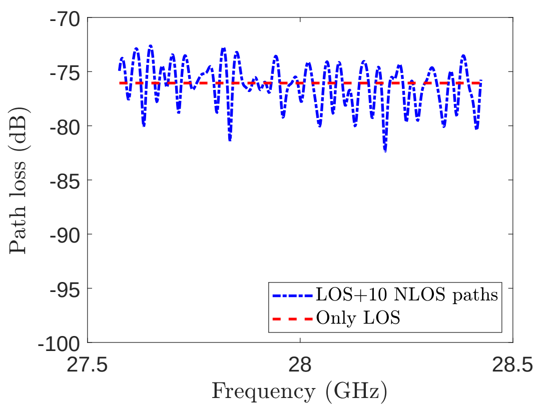


(c) 20 NLOS paths.

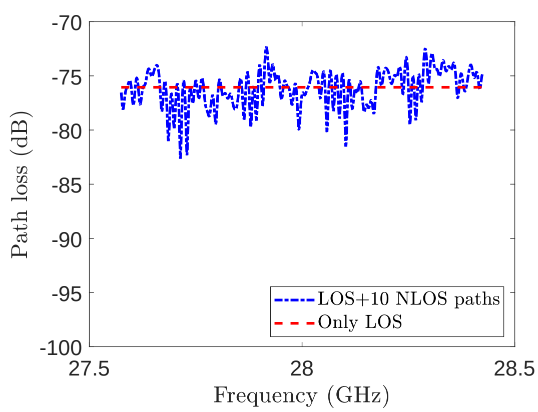
Fig. 1. *Path loss variations over 850 MHz centered at 28 GHz for different number of NLOS paths.



(a) NLOS path length= LOS path length + 20%LOS path length.



(b) NLOS path length= LOS path length + 30%LOS path length.



(c) NLOS path length= LOS path length + 40%LOS path length.

Fig. 2. *Path loss variations over 850 MHz centered at 28 GHz for a LOS and 10 NLOS paths. NLOS path lengths are varied as follows: NLOS path length= LOS path length + x%LOS path length, where $x = 20, 30,$ and 40 in Figures 2(a), 2(b), and 2(c), respectively.

We believe our initial results presented in this manuscript set path to future research in wideband channels. A similar trend in mm-wave channels is often observed in the literature such as in [1]* (Please see below) and [10,11] (Please see the manuscript) where the initial results are obtained in narrow-band channels.

[1]* A. Alkhateeb and R. W. Heath, Frequency Selective Hybrid Precoding for Limited Feedback Millimeter Wave Systems, 2016, doi: 10.1109/TCOMM.2016.2549517.

4. Assuming $M \geq M_{rf} \geq K$, and specifically $M_{rf} = K$ might be unrealistic, could you please elaborate on that?

Reply: In the revised manuscript, we alleviate the $M_{rf} = K$ assumption by utilizing the varying channel conditions in the network. The new numerical results in Fig. 7 demonstrate that significant power savings can be achieved at low sum-rate losses by adaptively shutting off the RF chains depending on the channel conditions. We discuss the details of adaptively shutting off the RF chains in a paragraph starting with the sentence "In Fig. 7, the sum-rates are evaluated for varying channel ...". As also noted in Section VIII of the revised manuscript, further investigation of the challenge $M_{\mathrm{rf}} < K$ can be within our future research goals following the proposed line of joint analog beam selection and digital precoder design work.

5. The notation is inconsistent at times, it should be fixed and unified.

Reply: We have tried our best to identify and correct the inconsistent notations throughout the paper. For instance, we have included a notation subsection in Section II, the new Fig. 1 that highlights some notations in the paper, and finally, we have extended the number of Algorithms from 2 to 5 where we have clearly detailed the proposed solutions. We have also simplified Section VI.A with simpler notations and shorter explanations.

Finally, for instance, we have noticed that the $\mathbf{u}$ and $\mathbf{U}$ symbols have been excessively overloaded. Hence, we have introduced new distinctive notations such as $\underline{\mathbf{U}}_c^{\mathrm{comb}}$ in Algorithm 2 and the DFT codebook $\mathbf{D}$ in equation (13). Similarly, the set notation of the predefined beams has been changed to $\underline{\boldsymbol{\mathcal{U}}} = \{\underline{\mathbf{u}}_1, \ldots, \underline{\mathbf{u}}_B\}$ to have clear distinctions from $\boldsymbol{\mathcal{U}}_k$ and $\mathbf{u}_{kl}$, i.e., the available beam set at user k under BCC and the k^{th} column of the precoding matrix $\mathbf{U}_l$ at AP l, respectively.

We hope the notations have become more unified in this version. Nevertheless, we will appreciate if some of the inconsistent notations are kindly exemplified in case they still exist.

6. It is mentioned that "The sum-rate performances can be improved by another fixed but unequal power allocation solution," but no citation is given.

Reply: Thank you for pointing out this missing point. In the revised version, we have cited a reference: "The sum-rate performances can be improved by another fixed but unequal power allocation solution [20]".

7. It is unclear how ML training can be done offline, could you please explain?

Reply: We appreciate such valuable feedback. In the revised version, we have included the new Fig. 3 and the new Section VI.B to explain the off-line training phase in more detail. We believe the new section and figure emphasize better our major aim, i.e., a joint design with multiple features can significantly improve the sum-rate performance but also, it is nearly an unrealistic application without the help of ML algorithms. As emphasized in Fig. 3 and Section VI.B, by off-line training of ML algorithms through well-constructed designs, significant sum-rate gains can be achieved when ML algorithms are executed online that can suit practical implementations in terms of reduced complexity. In the revised manuscript, we also underline this point by explicitly addressing what well-constructed designs mean, as seen in a paragraph of Section I: "We refer to a joint design with multiple features as a well-constructed design".

8. A figure showing the notation and the problem could really help understanding the work.

Reply: We appreciate the reviewer for addressing this important point. In the revised version, we have included the new Fig. 1 which demonstrates the cell-free network and AP architectures, and also includes the notations in the paper.

9. With the given introduction, it is hard to understand the complexity of the problem. Further explanation would help the readers to better understand it.

Reply: Thank you for addressing this missing point. In Section I, now we briefly describe the problem setting and also the proposed solutions which can indicate the challenges we would like to tackle such as complexity as you mentioned. Along with these new explanations, the existing explanations in Section II on complexity, such as linear versus exponential search complexity, are further improved by detailing the effects of BCC a bit more.

10. The pseudo-code is not always clear. Maybe the syntax from a specific language is assumed as known? It should really be independent from a specific syntax and easily readable anyway.

Reply: We certainly agree with the reviewer that the pseudo-code of the ML algorithm is not self explanatory and also, it is language specific. We also think it can be redundant since we already present the highlights of the algorithms throughout Algorithms 1-5, and moreover, we share all of our codes on GitHub. Thus, we removed the ML pseudo-code. Removing it also allows us to limit the excessive page lengths since the length of the revised manuscript has significantly increased due to the new figures, numerical results, and sections.

11. It is unclear whether chained classifiers can actually exploit the correlation between labels as no results were shown. Could you please show some results on this?

Reply: We agree that we have not shown without classifier chain results in the figures but have only mentioned the performance difference in Sections II and VII, i.e., 63% versus 99%. We noticed such a sharp contrast in all cases we tested. We believe including without classifier chain results in the figures can negatively impact the readability of the figures while not bringing beneficial information. Hence, we opted not to include these results but only mention as a representative case for all the cases we tested. In Section VII of the revised manuscript, we emphasize this important note in the paragraph starting with the sentence "In this section, all numerical results are presented with classifier chains".

12. The parameters used to obtain the results are a optimistic, specifically, the path loss exponent, the shadowing, the noise power, and the AP transmit power

Reply: We appreciate bringing this important point to our attention which has inspired us to present new results that remedy the strong $M_{\text{rf}} = K$ condition. As detailed in our reply to your 4^{th} comment, in Fig. 7, we present new results over varying channel conditions and exploit this opportunity to reduce the power consumptions significantly at low sum-rate losses.

13. It is unclear what the following sentence means: "the distance between each AP and a user is chosen as 100 m". A figure might clarify the simulation setup.

Reply: We certainly agree this assumption is confusing. We simply chose a fixed distance to obtain more consistent results with the less number of Monte Carlo (MC) runs. Along with the randomness introduced via the log-normal shadowing in our simulations, additional randomness in distances can require more MC runs due to the increased randomness, i.e., random distances plus log-normal shadowing.

In the revised version, we varied the distances between 95-105 m that yielded similar numerical results. Moreover, by adding Figs. 1 and 7, we believe the possible confusions might have been reduced to minimum. For instance, in Fig. 7, we present numerical results over varying channel conditions including the distances that are uniformly distributed between 100-200 m.

14. Plots can be improved by adding units of measurement whenever possible, using the right type of plot (line plots are not

always the best choice)

Reply: We certainly agree with the reviewer for this careful observation. In our figures, now we have the sum-rate metrics noted as b/s/Hz and we also have new figures with the bars, i.e., Figs. 6 and 7. As noted by the reviewer, we believe these new figures with the bars can certainly be more proper.

15. In the results section, it is unclear what are the baseline results, i.e., the state of the art that the authors are trying to improve upon.

Reply: We appreciate the reviewer for letting us know about such a major confusion in the paper. Now we stress the benchmark design clearly throughout the revised manuscript, e.g., "... disjoint design benchmark" in the abstract, "Pseudocode of the benchmark ..." in Algorithm 5, and also in Figs. 4 and 5.

16. Overall, the quality of the presentation can be largely improved, as well as the final results

Reply: Following your suggestions, we believe the quality of the presentation and the final results have been significantly improved.

Response to Reviewer 3

In this paper, heuristic search algorithms for beam selection in millimeter-wave cell-free massive MIMO systems are proposed. Subsequently, supervised machine learning algorithms are used to approximate the input-output mapping functions of beam selection algorithms.

Reply: We thank you for your comments. We have tried our best to address your concerns below.

1. The contributions and novelty of this work are marginal. Particularly, in section V. description of the proposed search algorithms are presented which contain neither novel contributions nor analytical discussion.

Reply: To the best of our knowledge, joint design solutions of analog beam selection via beam training and digital filters were proposed for the first time in our work at the initial submission date. To this date, this research direction is still open [1]*. Furthermore, in this work, we present new challenges introduced by BCC that are particular to the cell-free networks, and also provide solutions to them. Although the proposed joint design solutions with multiple features, i.e., well-constructed designs, are too complex for implementations, we show that the complexity challenge is significantly obviated by using the ML algorithms. To the best of our knowledge, the correlated outputs challenge introduced by the BCC are noted for the first time in our work where we also propose classifier chains to tackle this challenge.

[1]* J. Zhang et al., "Joint beam training and data transmission design for covert millimeter-wave communication," in IEEE Transactions on Information Forensics and Security, vol. 16, pp. 2232-2245, 2021, doi: 10.1109/TIFS.2021.3050070.

We kindly note that, we could have expressed the steps of our proposed solutions in fine details via analytical expressions as in [14,20]. Instead, we preferred to present overview explanations (as seen in the paragraphs of Section V.E) with overview mathematical notations (as seen in the Algorithms 1-5) throughout the manuscript. Since there are 5 algorithms presented in the paper, getting into details for each can significantly extend the length of the paper and more importantly, can significantly challenge the readability of the paper.

Thanks to your valuable comment, in the revised manuscript, we have added equations (11)-(13) which we had noticed that the mentioned concept might be challenging to be understood without mathematical expressions. We have also noted an example in the revised manuscript as follows, for instance, for Algorithm 2: "Although BCC is achieved in a single line at step 7 of the given pseudo-code, its coding implementation is significantly challenging [26]".

In summary, we choose to leave out the detailed analytical expressions of the proposed 5 algorithms, and rather use simple explanations and simple mathematical notations. Due to the particularity of BCC in cell-free networks, correlated outputs introduced by BCC and the use of classifier chains to tackle this problem, and most importantly, showing that well-constructed designs can significantly improve the sum-rates at low complexities, thanks to ML algorithms, are identified and demonstrated for the first time in this work. Hence, we strongly believe our work can capture other researchers' interests.

In the following lines, we try to explicitly clarify the motivation to keep off the detailed mathematical expressions of the proposed 5 algorithms.

In Fig. 3* (Please see below), we present a part of the code for Algorithm 4 with BCC. Most of the steps in Fig. 3* can be written in analytical expressions as done in [14,20]. However, in the manuscript, we rather prefer to highlight the main steps and provide simple explanations with simple mathematical expressions, almost like flowchart boxes, as mentioned earlier. For all 5 algorithms presented in the manuscript, we follow this direction to keep the readability of the manuscript high, and also, we believe that the readers who are interested in the fine details will refer to the provided codes on GitHub. In Fig. 4*, we also compare the codes for the semi-linear search algorithm with no BCC and with no features versus with BCC and with two features. Algorithm 2 in the manuscript corresponds to the lines between 7 and 25 in Fig. 4(b)*. Again, most of the steps in Fig. 4(b)* can be written in fine details by using analytical expressions in the manuscript. However, we keep off such expression in the manuscript which also helps us to write the pseudo-code of Algorithm 2 in a flowchart manner. Please note that the pseudo-code of Algorithm 2 is very simple and almost looks like the semi-linear algorithm with no BCC and with no features like in Fig. 4(a)*.

```
for initi=1:Init
    Lvec=0:L-1;
    randindexes=dlmread(fi2,'\t',[loaders(2) 0 loaders(2) L*K-1]);
    loaders(2)=loaders(2)+1;
    for j=1:L
        Usxx=U(:,randindexes);
        rate_Sum_maxx=0;
        for stepi=1:Iter
            mMUsers=repmat(mM,[K,1]);
            kcounter=M*ones(K,1);
            for l=Lvec
                for k=0:K-1
                    kx=kcounter(k+1);
                    rate_O=zeros(kx,1);
                    for m=1:kx
                        Usxx(:,l*K+k+1)=U(:,mMUsers(k+1,m)+1);
                        rate_O(m)=sum(rate_U_MultiwFilter_v02dt2c(sP,L,K,M,H,Usxx,sigma2n));
                    end
                    [~,index]=max(rate_O);
                    indexx=mMUsers(k+1,index);
                    kNot=(0:K-1)';
                    kNot(k+1)=[];
                    for kk=1:K-1
                        indexxx=find(mMUsers(kNot(kk)+1,:)==indexx);
                        if ~isempty(indexxx)
                            mMUsers(kNot(kk)+1,indexxx:end)=[mMUsers(kNot(kk)+1,indexxx+1:end) 0.1];
                            kcounter(kNot(kk)+1)=kcounter(kNot(kk)+1)-1;
                        end
                    end
                    Usxx(:,l*K+k+1)=U(:,indexx+1);
                    UoptsMxx(1,l*K+k+1)=indexx+1;
                end
            end
            rate_Sum=sum(rate_U_MultiwFilter_v02dt2c(sP,L,K,M,H,Usxx,sigma2n));
            if rate_Sum>rate_Sum_maxx
                rate_Sum_maxx=rate_Sum;
                Usx=Usxx;
                UoptsMx=UoptsMxx;
            end
            %%%
        end
        if rate_Sum_maxx>rate_Sum_max
            rate_Sum_max=rate_Sum_maxx;
            Us=Usx;
            UoptsM=UoptsMx;
            stepi_opt=stepi;
        end
        Lvec=circshift(Lvec,-1);
    end
end
```

Fig. 3. *A part of the code for Algorithm 4 with BCC.

```
for l=0:L-1
    lvec=l*K+1:l*K+K;
    for lx=1:lUopts
        [Uopts,~]=permn(mM+1,K,lx);
        Us(:,lvec)=U(:,Uopts);
        rate_O(lx)=sum(rate_U_MultiIa_v03(sP,L,K,M,H,Us,sigma2n));
    end
    [~,lopt]=max(rate_O);
    [index,~]=permn(mM+1,K,lopt);
    Us(:,lvec)=U(:,index);
    UoptsM(1,lvec)=index;
end
```

(a) Algorithms 1 and 2 with no BCC and with no features.

```
for initi=1:Init
    randindexes=dlmread(fi2,'\t',[loaders(2) 0 loaders(2) L*K-1]);
    Usx=U(:,randindexes);
    loaders(2)=loaders(2)+1;
    for stepi=1:Iter
        for l=0:L-1
            lvec=l*K+1:l*K+K;
            rate_O=zeros(1,lindexes);
            for lx=1:lindexes
                Usx(:,lvec)=U(:,indexes(lx,:));
                rate_O(lx)=sum(rate_U_MultiwFilter_v02dt1b(sP,L,K,M,H,Usx,sigma2n));
            end
            [~,lopt]=max(rate_O);
            index=indexes(lopt,:);
            Usx(:,lvec)=U(:,index);
            UoptsMx(1,lvec)=index;
            for k=1:K
                for j=1:K
                    if j~=k
                        temp=find(index(j)==indexes(:,k));
                        indexes(temp,:)=[];
                    end
                end
            end
            [lindexes,~]=size(indexes);
        end
        rate_Sum=sum(rate_U_MultiwFilter_v02dt1b(sP,L,K,M,H,Usx,sigma2n));
        if rate_Sum>rate_Sum_max
            rate_Sum_max=rate_Sum;
            Us=Usx;
            UoptsM=UoptsMx;
            stepi_opt=stepi;
        end
    end
end
```

(b) Algorithms 1 and 2 with BCC and two features: multiple initializations and iterations.

Fig. 4. * Parts of the codes for the Algorithms 1 and 2.

2. The title of the work is joint analog beam selection and digital beamforming, however, the design of the analog and digital beamformers are not discussed. Particularly, in subsection V.D. well-known ZF and MMSE digital beamformers are reviewed.

Reply: We appreciate the reviewer for identifying the insufficient explanation on joint design in the previous manuscript. In the revised manuscript, to explain our proposed joint design and compare it with the disjoint benchmark, we have added a new paragraph in Section V.E that starts with the sentence "To the best of our knowledge, in the literature and standards including ...". To emphasize the difference between our proposed joint design and the disjoint benchmark, we have also added a paragraph in Section V.D, in particular, the following sentences: "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 to 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".

In other words, in the disjoint design, the digital precoders are held fixed while analog beams are tested and selected. Next, analog beams are held fixed and the digital precoders are designed. In contrast, in our proposed joint design, every time a beam is "tested", the digital precoder is designed as well. Therefore, our beam selection decisions are based on this joint design. Hence, our proposed well-constructed designs have superior sum-rate performances compared to disjoint design as shown in Figures 4 and 5.

We choose ZF and MMSE digital beamformers due to their low complexities since we assume APs in the cell-free networks have only simple components. As noted in the manuscript, for instance, in the beginning of the introduction, all users in the cell-free networks are expected to be served by a large number of simple multi-antenna APs. Due to the high costs of such large numbers of AP deployments in the cell-free networks, APs can be preferred to be deployed with simple components as mentioned earlier. In the revised manuscript, we have added one more reference and mentioned that well-known maximum-ratio transmitter/combiner (MRT/MRC) digital filters are also typically used in mm-wave networks. In our future work, we definitely agree that the powerful digital precoder designs can be an interesting direction in the line of our work, assuming for instance base stations are utilized instead of simple APs.

3. Furthermore, Section VI. contains a review of well-known ML techniques which is irrelevant and unnecessary. In the rest of this section application of these techniques for estimating the output

of the beam selection function using Scikit is presented which is trivial and not worthy of being reported in a technical research paper.

Reply: We thank the reviewer for indicating the unnecessity of brief reviews of the ML algorithms in the previous manuscript. In the revised manuscript, we have completely removed these short reviews.

As we also discuss in the comment 7 of reviewer-2, the proposed joint designs with multiple features, i.e., well-constructed designs, are nearly impractical due to their increased computational complexities and also due to the need for frequent estimation of effective channels as we address earlier, i.e., comment 2 of reviewer-3. Thanks to the off-line ML training, via online executions of the proposed ML algorithms, the computational complexities are significantly reduced and the need for frequent effective channel estimation is eliminated. In the revised manuscript, we have added Section VI. B and Fig. 3 to emphasize that by the help of ML algorithms (they can be practical), we can achieve 99-100% of the original sum-rate results of the well-constructed solutions (they cannot be practical) at low computational costs and also, by only one-time effective channel estimation. We kindly note that this is the same amount of one-time effective channel estimation which is needed in all disjoint designs as widely proposed in the literature and standards. Hence, at a low increase in the computational cost of ML algorithms compared to disjoint designs, we can achieve nearly the same superior sum-rate results of the well-constructed designs. Next, we summarize the contributions of our ML algorithms.

- The proposed well-constructed algorithms are superior to naive disjoint design. However, their computational complexities and frequent effective channel estimation requirement are impractical.

- The proposed ML algorithms can alleviate these challenges while achieving nearly the same sum-rate performances, i.e., 99-100% of the well-constructed designs. Well-constructed designs have superior sum-rate performances compared to the disjoint design as shown in Figures 4 and 5.

- The effect of BCC on cell-free networks is quite unique as detailed in Section V.B for the first time in the literature. One of its unique affect is creating highly correlated output labels. In Section VII, we show that RFt ML algorithm 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 for the first time in the literature.

Both Keras and Scikit are well-known and often utilized in the literature. For instance, Scikit and Keras implementations are utilized in [1-2]* (Please see below) and [3-5]*, respectively. As demonstrated in the manuscript, the Scikit library can already provide us fast results with very good approximations. However, when we extend our work to users with multiple antennas and/or wideband mm-wave channels, GPUs might be needed. In this case, Keras with GPU support can be needed. In short, as well-known, both Keras and Scikit are high-level libraries with similar syntaxes that are often used in the literature, and both are very easy to pick up compared to TensorFlow, which is a low-level library. However, currently we do not need TensorFlow. TensorFlow can be needed in future for specific research implementations as its having a speed advantage.

4. Rate the references: Unsatisfactory (explain): -

Reply: Thank you for addressing this shortage of the manuscript. We have cited more papers throughout the revised manuscript including Sections I and II. For instance, we briefed 3 more beam selection algorithms [16-18] in Section II. In particular, the number of references has been increased from 20 to 30. We hope we have covered the literature more extensively in the revised manuscript without reaching to an excessive page length.

In short, we deeply appreciate the constructive comments of the reviewer, and in the revised manuscript, we completely remove the brief reviews of ML algorithms, and also remove the details of Scikit implementation such as the Appendix. We truly thank the reviewer in helping us improving the quality and presentation of the manuscript significantly.

[1]* 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, 2019, doi: 10.1109/ACCESS.2019.2922064.

[2]* C. Antn-Haro and X. Mestre, Learning and Data-Driven Beam Selection for mmWave Communications: An Angle of Arrival-Based Approach, IEEE Access, 2019, doi: 10.1109/ACCESS.2019.2895594.

[3]* H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, Deep-Learning-Based Millimeter-Wave Massive MIMO for Hybrid Precoding, IEEE Trans. Veh. Technol., 2019, doi: 10.1109/TVT.2019.2893928.

[4]* A. Zappone, E. Bjrnson, L. Sanguinetti, and E. Jorswieck, Globally Optimal Energy-Efficient Power Control and Receiver Design in Wireless Networks, IEEE Trans. Signal Process., 2017, doi: 10.1109/TSP.2017.2673813.

[5]* A. Zappone, M. Di Renzo, and M. Debbah, Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both?, IEEE Trans. Commun., 2019, doi: 10.1109/TCOMM.2019.2924010.

Selected References from the Manuscript

[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. Anton-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, Feb. 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.

[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.

[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. 2169-3536, Jun. 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.

[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.

[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.