Manuscript ID OJCOMS-00495-2021.R1
Editor: Prof. Carmen D'Andrea

Dear Respected Editor and Reviewers,

Thank you very much for coordinating the reviews and for your insightful comments. We have carefully revised the manuscript based on your suggestions. We believe significant improvements are achieved once again and the technical and presentation qualities of the manuscript are now fined tuned.

In the following, we provide point-by-point replies to your comments, and the revised manuscript is attached separately.

We highly appreciate your supportive, cooperative, and timely coordination of our manuscript from the beginning which have resulted in a manuscript with a great deal of qualities.

As supplementary files, the following files are uploaded on GitHub [26]. Please kindly see the folder JointDesigninmmWaveCellFreeNetworks/Files/ at https://github.com/DrCMY:

- Complexity.xlsx
- CodewordPossibilities.xlsx
- Manuscript_DC_wHighlights.pdf
- Manuscript_SC.pdf
- Manuscript_SC_wHighlights.pdf

We kindly note that Prof. Carmen D'Andrea and Prof. Rui Dinis are the editor and the area editor, respectively. Based on email exchanges on April 9, 2021, we received a 3-week extension to resubmit our work.

Response to Editor

1. Improve the readability of figures presenting the results.

Reply: We have carefully revised the figures to use the labels neatly and consistently. For clarity, we have removed the data points from and added further explanations on some figures. Finally, we have added a new table in the beginning of Section VII - Numerical Results that summarizes the abbreviations and notations used in the figures. This table can be conveniently referred without scanning the paragraphs while analyzing the figures.

2. The reviewer points out that the simulation scenario seems to be unrealistic for a cell-free massive MIMO system and I agree with him/her, only 3 APs with 2 users is far from a "Massive MIMO" framework. I suggest considering a denser scenario or properly justifying the choice.

Reply: In the revised manuscript, we present new numerical results for large network sizes in Fig. 8. As discussed in the following pages and in the manuscript, the complexity of the semilinear search algorithm becomes impractical even for small network sizes. For instance, the complexities for the largest network size in the previous submission, i.e., L4K4M8, are 128 and 16×10^3 (16k) for the linear and semilinear search algorithms, respectively. The complexities of the linear and semilinear search algorithms reach to 16k and 3.2×10^{31}, respectively, when the network size is L25K20M32 as detailed in the following pages. As discussed later in the response letter and in Fig. 8 of the manuscript, as the network size grows, more complex algorithms, i.e., well-constructed algorithms, achieve more than the simpler algorithms in sum-rate performance by benefiting the rich degrees of freedom in larger networks.

Response to Reviewer 1

The authors amended the paper and they improved it. Thus, in my opinion the paper can be published as is.

Reply: Thank you. We highly appreciate the reviewer's supportive, cooperative, and timely coordination of our manuscript.

Response to Reviewer 2

I thank the authors for carefully and kindly answering the concerns for the first version of the paper. The new manuscript has been largely improved over the first version, but it was hard to track all the changes applied by the authors. It would be desirable to highlight changes somehow, maybe using colors or highlighted text.

Reply: Thank you also for your critical and careful comments from the beginning. The quality of our manuscript has been significantly improved during the review process, and we highly appreciate your end-to-end helps.

The previous manuscript was significantly modified compared to the initial submission. Nearly every paragraph, figure, and page were touched. As suggested by the reviewer and as we usually do in our submissions with modest revisions, we prepared a highlighted manuscript and uploaded on GitHub [26].

1. Fig. 1a: There might be a mistake: User l should probably be User k.

Reply: We appreciate your careful observation. The mentioned typo is corrected in the revised manuscript.

2. Many plots are hard to read: legends, tick labels, axis label, and captions are often not very informative, using complex notation and acronyms.

Reply: We have carefully scrutinized the figures for consistency and clarity. For instance, in Fig. 6, the init-iter ($M = 2$) abbreviations are modified to $\text{init} = 1$, $\text{iter} = 1, M = 2$ which is now consistent with the other figures. Moreover, in Fig. 4, the number of data points is reduced so that instead of the init-iter notation, the $\text{init} = 1$, $\text{iter} = 1$ notation is used for consistency with the other figures. In short, the lengthy, cluttered, and unclear labels and notations in the figures are corrected. Some other changes are briefed in Table 1* of this response letter.

Throughout the figures, the axis titles are chosen to be bold letters to clearly differentiate them from the axis ticks. For instance, the x-axis title "Beam selection algorithm" is bold and the algorithm names, e.g., linear - II - rate, as the x-axis ticks are non-bold in Figs. 4 and 5. Nevertheless, all figures have been carefully scrutinized and unnecessary bold letters have been removed. For instance, in Fig. 1 of the manuscript, AP and user labels are now non-bold. We use bold letters in the figures in the intention of clarity in reading the hard copy manuscript (i.e., if the reader prefers to print out the manuscript) since figures are often squeezed into smaller areas in the manuscripts than needed.

Table 1: *SOME OLD AND NEW LABELS.

Old label	New label
(Figs. 4 and 5) semi-linear-II-Rate metric linear-DL metric-Disjoint	semilinear - II - Rate disjoint linear - DL
(Fig. 4) 1-1 init-iter	init = 1, iter = 1 number of initializations and iterations
(Fig. 6) 1-1(M=4) init-iter (Antenna number)	init = 1, iter = 1, $M = 4$ number of initializations, iterations, and antennas
(Fig. 7) linear-II-SMrf linear-II-Mrf	linear - II - rate - Smartly reduced chains linear - II - rate - Naively reduced chains

Before presenting the figures in Section VII - Numerical Results, all notations and abbreviations used in the figures are clearly explained. Nevertheless, they are now succinctly summarized in Table I of the revised manuscript which can be conveniently referred without scanning the paragraphs while analyzing the figures.

In fact, another challenge we face is that from Fig. 4 to Fig. 10, 9 different numerical results are presented with many beam selection algorithm options, beam conflict control (BCC) options, number of RF chain options, and machine learning (ML) algorithm options. Therefore, the figures differ significantly from one to the next, and each analyzes different aspects. We keep the abbreviations and notations in the figures as informative as possible while keeping them compact so that the figures can fit in. We believe that by including Table I and carefully revising the labels and notations, the numerical results in the figures are now much easier to follow.

Finally, we revised the figure captions and highlighted the changes in the supporting document uploaded on GitHub.

3. In Sec. III-A, when defining the angle θ_{kl}^{p}, shouldn't it be in $[-\pi, \pi)$ rather than $[-1, 1]$? Furthermore, it is not clear how many rays P were simulated for the channel model declared in (3).

Reply: Thank you for noting the typo in our manuscript. The mentioned correction is done and equation (5) is corrected now.

In the first paragraph of Section VII - Numerical results, the number of paths is now provided.

4. Not being a specialist, I would like to ask whether it is correct to use the definition of SINR stated in (6b)? My concern is that D_k and I_k could be highly correlated, and thus taking the ration of the expected values might not coincide with the expected value of the ratio, possibly even by a long shot. This is not intended to be a critic to the authors' work, but I am indeed concerned about the

relevance of this specific assumption.

Reply: The rate expression given in (6) is a typical lower bound from the literature on Massive MIMO, often referred to as the use-and-then-forget bound. It is derived in detail in [1]*.

[1]* E. Björnson, J. Hoydis, and L. Sanguinetti, Massive MIMO networks: Spectral, energy, and hardware efficiency, Foundations and Trends in Signal Processing. 2017.

5. Could you please explain the constraint (7c)? I understand from the text below the equation that it is intended to assert the BCC solution, but it is not clear to me why it is formulated in this way. Specifically, I understand that all AP share the same codebook $\underline{\mathcal{U}}$ for the analog pre-coder, and that the same codeword should not be used for multiple users, but this should only be true for a given AP l, correct? To be clearer: from (7c) I understand that if $u_{kl} = b$ (using codewords indexes), then, following the BCC assumption $u_{jk} \neq b$ for $j \neq k$, while $u_{km} = b$ for $m \neq l$ is possible, but $u_{jm} = b$ for $j \neq k$ and $m \neq l$ is not? Why wouldn't the last case be possible for the BCC assumptions? Is this the interpretation also use in the code?

Reply: We believe the reviewer means $u_{jl} \neq b$ for $j \neq k$ or $u_{jn} \neq b$ for $j \neq k$ and $n \neq l$ rather than $u_{jk} \neq b$ for $j \neq k$.

The question can be fundamentally answered without considering BCC, cell-free, or mm-wave networks. Assume there are 2 transmitter and 2 receiver nodes in the network. If it is a broadcast network, then the transmitters 1 and 2 send $\mathbf{U}_1 s_1$ and $\mathbf{U}_2 s_2$ to all users, respectively. Note that in a broadcast network, the transmitter i transmits the symbol s_i to all users. Clearly, $\mathbf{U}_1 = \mathbf{U}_2$ is not an option. If it is an interference network, then the transmitters 1 and 2 send $[\mathbf{u}_{11}\mathbf{u}_{21}][s_{11}s_{21}]^T$ and $[\mathbf{u}_{12}\mathbf{u}_{22}][s_{12}s_{22}]^T$ to the users. Here, $\mathbf{u}_{kl}$ and s_{kl} are the beamforming vector and the data symbol intended from transmitter l to user k, respectively, and $(.)^T$ is the transpose operator. Again, $\mathbf{u}_{kl} = \mathbf{u}_{jm}$, $\forall j \neq k$ and $\forall m \neq l$, is not an option.

Now consider a cell-free network where all APs $l \in \{1, \ldots, L\}$ jointly serve user k to transmit the same symbol s_k. Then, the transmitters 1 and 2 send $[\mathbf{u}_{11}\mathbf{u}_{21}][s_1 s_2]^T$ and $[\mathbf{u}_{12}\mathbf{u}_{22}][s_1 s_2]^T$ to the users, i.e., the transmitters jointly transmit the symbols s_1 and s_2 to the users 1 and 2, respectively.

Case 1: Again, $\mathbf{u}_{kl} = \mathbf{u}_{jl}$, $\forall j \neq k$ is not an option, i.e., the cases $\mathbf{u}_{11} = \mathbf{u}_{21}$ and $\mathbf{u}_{12} = \mathbf{u}_{22}$ are not possible.

Case 2: Similarly, $\mathbf{u}_{kl} = \mathbf{u}_{jm}$, $\forall j \neq k$ and $\forall m \neq l$ is not an option, i.e., the cases $\mathbf{u}_{11} = \mathbf{u}_{22}$ and $\mathbf{u}_{21} = \mathbf{u}_{12}$ are not possible.

Case 3: However, $\mathbf{u}_{kl} = \mathbf{u}_{km}$, $\forall m \neq l$ is possible, i.e., the cases $\mathbf{u}_{11} = \mathbf{u}_{12}$ and $\mathbf{u}_{21} = \mathbf{u}_{22}$ are possible.

Clearly, the above 3 cases can be collected under a single condition $\mathbf{u}_{kl} \neq \mathbf{u}_{jm}, \forall j \neq k$.

For your convenience, we share the spreadsheet link and also upload the file as a supporting document where we explicitly demonstrate the above mentioned toy example.

Here is another explanation supported by another toy example. Consider Fig. 1* in this response letter, where we assume the cell-free network has 3 access points (APs) and 2 users. As demonstrated in the figure, all APs jointly transmit the symbols s_1 and s_2 to the users 1 and 2, respectively.

Furthermore, let us assume the codebook has 3 codewords, i.e., $B = 3$, $\underline{\boldsymbol{\mathcal{U}}} = \{\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2, \underline{\mathbf{u}}_3\}$. Let $[\mathrm{AP}_1, \mathrm{AP}_2, \mathrm{AP}_3]_{\mathrm{user}_i}$ denote the selected the codewords by APs for user i.

Then, $[\underline{\mathbf{u}}_i, \underline{\mathbf{u}}_i, \underline{\mathbf{u}}_i]_1$ and $[\underline{\mathbf{u}}_j, \underline{\mathbf{u}}_j, \underline{\mathbf{u}}_j]_2$, where $i \neq j \in \{1, 2, 3\}$ (a must), are valid choices regarding beam conflict control (BCC) to avoid low-rank effective channels as explained in the manuscript. For instance, if $[\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_1, \underline{\mathbf{u}}_1]_1$, then $[\underline{\mathbf{u}}_2, \underline{\mathbf{u}}_2, \underline{\mathbf{u}}_2]_2$ or $[\underline{\mathbf{u}}_3, \underline{\mathbf{u}}_3, \underline{\mathbf{u}}_3]_2$ is valid.

However, if $[\underline{\mathbf{u}}_i, \underline{\mathbf{u}}_i, \underline{\mathbf{u}}_j]_1$, where $i \neq j \in \{1, 2, 3\}$ (a choice), then $[\underline{\mathbf{u}}_k, \underline{\mathbf{u}}_k, \underline{\mathbf{u}}_k]_2$, where $k \neq i$ & $k \neq j \in \{1, 2, 3\}$ (a must). For instance, if $[\underline{\mathbf{u}}_1, \underline{\mathbf{u}}_1, \underline{\mathbf{u}}_2]_1$ (a choice), then $[\underline{\mathbf{u}}_3, \underline{\mathbf{u}}_3, \underline{\mathbf{u}}_3]_2$ (a must).

In short, the selected codewords for a user cannot be used for other users. For the last example, since codewords 1 and 2 are chosen for user 1 by the APs, APs cannot choose the codewords 1 and 2 for user 2, but only the remaining codeword, i.e., codeword 3.

Hence, as proposed in the manuscript, a codebook log is kept for each user $\underline{\boldsymbol{\mathcal{U}}}_i$ which is updated and announced in the network after each beam assignment.

In the manuscript, the numerical results with BCC process and with only BCC initialization are obtained following the mentioned principles.

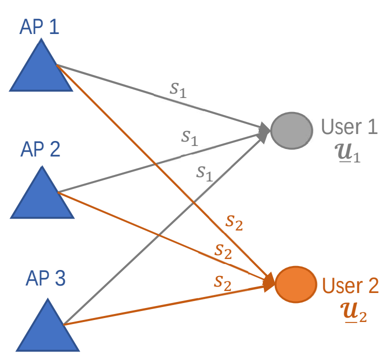

Fig. 1. *An exemplary network with 3 APs and 2 users.

6. The answer regarding the fixed 100 m distance between APs and users is not convincing. Simulation time shouldn't be considered when setting the simulation parameters, and setting a 95-105 m uniform distribution is not convincing either. The authors should clarify this choice, and if no impellent reason can be found, the authors should consider extending the simulation to a more realistic scenario.

Reply: In addition to Fig. 7, which was included in the previous submission based on the same comment of the reviewer, in the current submission, we now provide Fig. 8, where the numerical results are obtained assuming 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.

In Fig. 7, which was submitted as new results in the previous revision based on the same comment, the numerical results are obtained assuming 2-4, 4-6 dB, and 100-200 m. As seen in Fig. 7, all numerical results presented in Section VII benchmark to each other "indifferently" as the channel conditions "vary", and the conclusions presented in the manuscript still hold. Since the proposed algorithms are indifferent to channel conditions, the same benchmark results will be obtained in other channel conditions as well.

In particular, as seen in the orange bars of Fig. 7, the benchmarks (i.e., the comparisons of two algorithms) are nearly fixed to 5% and 30% as the channel conditions vary. Similarly, for the gray bars, the benchmarks are nearly fixed to 20%. In short, for all numerical results in the manuscript, under different channel conditions, the benchmarks will remain same, e.g., the gaps between the beam selection options in Figs. 4 and 5, or BCC options in Fig. 6, and so on, which do not bring new information and all the conclusions reached in the manuscript remain the same.

By including the new results in Figs. 7 and 8, thanks to your valuable comments which made us notice this important lacking aspect in the manuscript, we believe the results in the manuscript are now complete addressing all aspects.

7. Not being a specialist on cell-free mMIMO, I wonder whether L3K2M8, being the default scenario, might be too small for a realistic scenario. Specifically, only 3 APs and 2 users spaced apart about 100 m are considered, which implies an extremely low user density especially in a city-center environment. Furthermore, a larger scenario might make the assumption $M \geq M_{rf} \geq K$, and especially $M_{rf} = K$. not valid anymore.

Reply: In Fig. 8, the numerical results for large network sizes L25K20M32 and L30K25M64 are plotted. Using the linear search approach, the complexity of the largest network L4K4M8 presented earlier in Fig. 6 is 128. Whereas the complexities of the networks L25K20M32 and L30K25M64 in Fig. 8 for the linear search approach are 16k and 48k, respectively. As detailed in the manuscript, these complexity evaluations are based on only the raw searching principles, and the extra features including, the number of initializations and iterations, and the selection method are not included. As seen in Algorithms 1-5 and in Fig. 3, the multiple initializations, iterations, and selection features are iterated in for-loops. This means whether it is the linear or semilinear search approach, the complexity values need to be multiplied by the number of initialization, iterations, and APs (i.e., see step 6 in Algorithm 4). For instance, the complexity values mentioned in the previous and current response letters, and in the manuscript need to be multiplied by 45 if the parameters are set to init $= 5$, iter $= 3, L = 3$ as done in Figs. 4 and 5.

Furthermore, the complexity of the semilinear search approach presented in Figs. 4 and 5 already reach to 16k even for the networks L4K3M16 and L4K4M8. As can be queried in the spreadsheet shared in [26], the complexities of the semilinear search reach to 3.2×10^{31} and 4.2×10^{46} for the networks L25K20M32 and L30K25M64, respectively. Due to the large network sizes in Fig. 8, where the linear and disjoint linear search algorithms are benchmarked, the Monte Carlo run, i.e., the number of channel realizations, is set to 100. Even in this case, the average simulation duration takes 6.5 hours while the simulation durations range from 1.5 to 14 hours.

For your convenience, we share the direct spreadsheet link and also upload the file as a supporting document for raw complexity evaluations of the algorithms.

Thanks to your prudence, by including the new results in Fig. 8, we believe the results in the manuscript are now complete addressing all aspects. Moreover, the advantage of joint design over naive disjoint is significantly emphasized by including the new results in Fig. 8.

Here is an excerpt from the revised manuscript: "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."

[26] C. M. Yetis. GitHub repository. Accessed: May 17, 2020. [Online]. Available: https://github.com/DrCMY.

8. It would be interesting to compare the baseline performance also in Fig. 4b, to understand the difference in complexity between that and more complex but more better performing approaches.

Reply: We thank the reviewer for carefully noticing the missing data in Fig. 4b. The baseline performance is already compared in Fig. 5b where a larger network size than Fig. 4b is tested. In Fig. 4b, the typo is now corrected and the missing data of the baseline is included.

9. Fig. 7 might lead to unclear results. In fact, while the first bars (Fixed) use the parameters given at the beginning of the section, namely, "the path loss exponent is 2, the log-normal shadowing is 4 dB, [...] the distances between APs and users vary uniformly between 95-105 m", the other bars show the results for ranges of parameters where the baseline (fixed) results are at one of the extremes, namely "log-normal shadowing, distance, and path loss exponent vary with uniform distribution between 4-6 dB, 100-200 m, and 2-4, respectively."

Reply: Based on the comment 12 of Reviewer 2 in the previous review process, we presented numerical results over different channel conditions in Fig. 7 in the earlier revision. As seen in Fig. 7, the benchmark percentages are nearly same while the y-axis values differ. For instance, linear-II-rate-smart scales down from 28 to 24 while the benchmark percentages are around 5% and 30%. The same behavior holds for all numerical results presented in the paper because none of the approaches in the manuscript, except the selection and smartly reducing chains features, utilize the channel conditions. Thus, all results scale in the y-axis while the benchmarks between them remain similar over varying channel conditions.

On the other hand, as seen in Fig. 8, the algorithms are not indifferent to network sizes. As the network size grows, more complex algorithms are likely to benefit more than the simpler algorithms in sum-rate performance by using the rich degrees of freedom in larger networks.

Here is an excerpt from the manuscript:

"As also demonstrated in Fig. 7, the benchmarks of the considered algorithms in our work are indifferent to varying channel conditions, i.e., the sum-rate losses and power savings vary around 5%, 20%, and 30% in average, respectively, over varying channel conditions. 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."

10. While the machine learning part of the paper brings small novelty to the community, the base idea is interesting and worth publishing.

Reply: To the best of our knowledge, for wireless networks, classifier chains were proposed for the first time in our work at the initial submission date. To this date, this research direction is still open [1]*.

As noted in the previous revision, to the best of our knowledge, the correlated outputs challenge introduced by the BCC for ML applications in wireless networks are noted for the first time in our work where we propose classifier chains to tackle this challenge.

In our ML codes, we use advanced grid search techniques such as Dask-ML [30] where we introduce a scoring function for grid searches. Similar to the challenge of correlated BCC outputs that required classifier chains, this is one of the challenges since both in Scikit and Dask-ML libraries, there are no scoring functions for multi-output problems.

[1]* C. Ben Issaid, C. Antn-Haro, X. Mestre and M. -S. Alouini, "User Clustering for MIMO NOMA via Classifier Chains and Gradient-Boosting Decision Trees," in IEEE Access, vol. 8, pp. 211411-211421, 2020.

Response to Reviewer 3

Recommendation: Accept. No further comments.

Reply: Thank you. We highly appreciate the reviewer's supportive, cooperative, and timely coordination of our manuscript.