Response for the Reviewer Comments

We would like to thank the associate editor and the reviewers for the valuable comments provided for the manuscript. We have written the response for the reviewer comments in-line in this report. The changes are addressed in the revised manuscript as well by using blue color italic fonts to discriminate the revision from the actual text in the manuscript.

RESPONSE FOR REVIEWER - 1

In this paper, the authors proposed a traffic aware resource allocation scheme for multi-cell MIMO-OFDM systems, where the precoders at all BSs are chosen to minimize the total user queue deviations. The problem is nonconvex and the authors proposed two centralized algorithms based on the successive approximation (SCA) technique to find a stationary point. Moreover, several distributed algorithms are also proposed using primal decomposition, alternating directions method of multipliers (ADMM), and decomposition via KKT conditions, respectively.

Most sections of this paper are well written. The results and algorithms also seem valid. However, the motivation of minimizing the total user queue deviations is not well justified. The convergence results of some algorithms are not clearly presented. The presentation of the distributed solutions needs significant improvement. Analysis and comparison of the signaling overhead and computational complexity between the centralized and distributed algorithms are also necessary to justify the advantages of distributed algorithms.

We thank the reviewer for providing valuable and insightful comments. We have address all the comments in the revised manuscript.

Q-(1) *In Section II.B, please provides more justifications for the problem formulation in (6). For example, the Queue weighted sum rate maximization (Q-WSRM) is throughput optimal, i.e., if there exists a scheme which can make all queues stable, then the Q-WSRM can also do this. How about the proposed formulation in (6)? Is it also throughput optimal?*

Resp : ~~We thank the reviewer for the valid comment.~~ The Q-WSRM scheme and the proposed schemes are all throughput optimal. It can be seen that the proposed extension Q-WSRME and the JSFRA formulations aim~~s~~ at minimizing the number of backlogged packets in addition to avoiding the over-allocation of the available resources. The JSFRA formulation with the ℓ_1 objective minimizes the number of backlogged packets in a greedy manner at each instant. By increasing the exponent $q \to \infty$, we obtain fair allocation at every transmission instant. We have included the discussions on the average number of backlogged packets at each instant for different arrival rates in Section V-C on page 13, col. 1. We have provided justifications for the problem formulation in (6) and (16) in page 5, col. 1, first paragraph. The equivalence between the Q-WSRM scheme and the JSFRA problem with ℓ_2 can be seen easily when the number queue size increases as outlined in the revised manuscript on page 4, col. 2, Section III-B starting paragraph.

Q-(2) *Do the proposed solutions based on (6) achieve better average delay performance than the existing solutions? By the way, in the simulations, you should also add a figure comparing the average delay performance, instead of just comparing the performance metric defined by (6). This will better justify the advantage of the proposed solutions.*

Resp : The proposed algorithm is a function of the number of backlogged packets and the channel conditions of all users in the network. Since the delay is proportional to the average number of backlogged packets in system, the proposed algorithm performs equally good in comparison ~~with~~ the existing Q-WSRM approach. It can be easily verified by the equivalence between the ℓ_2 norm JSFRA and the Q-WSRM scheme on page 4, col. 2, Section III-B starting paragraph. We have also provided the performance of different JSFRA algorithm and Q-WSRM scheme based on the average number of backlogged packets in the system over multiple transmission slots by varying the average number arrivals for all users in the system in Fig. 4 on page 13. We can still control the delay behavior by controlling the priority factor a_k in (6a), which affects the resource allocation. For example, if we consider a VOIP transmission, the delay requirements are stringent and it can be achieved by increasing the priority factor a_k in the current formulation. In addition, we can also control the delay by changing the exponent used in the objective ℓ_q of the JSFRA problem. We considered the residual number of backlogged packets as a performance measure, since we assume only the instantaneous channel state informations and together with the number of backlogged packets, resources can be allocated only for a given instant.

Q-(3) *In Section III.B, the convergence conditions under Algorithm 1 are not clear. First, you should be more specific about what is the SCA subproblem. Do you mean problem (19)? Second, does the uniqueness of the transmit and receive beamformers mean that the solution of the original problem in (16) is unique, or the solutions of the subproblems in (19) and (20) are unique, respectively?*

Resp : ~~We understand the reviewer's point.~~ We have updated the revised manuscript to discuss the convergence of the centralized algorithm in detail in Appendix B on page 14, col 1. The discussions are provided for the convergence of the iterative algorithm. The uniqueness of the convex subproblem (20) can be guaranteed when the objective is non-zero for all coordinating BSs in the precoder design. It is due to the linear constraint on the transmit precoders in (19), which is susceptible to the transmit precoder phase rotations. On the other hand, the constraint in (16b) is ~~not~~ since the precoder variable is inside the absolute value operator. Once the algorithm finds a unique set of transmit precoders, all unitary rotations are also valid for the original problem in (16). The uniqueness of the transmit and the receive beamformers are discussed in detail in Appendix B-C on page 15, col. 2.

Q-(4) *It is better to clearly summarize the convergence conditions and results (i.e., does it converge to a stationary point or the optimal solution) for all algorithms in a theorem/proposition.*

Resp : The discussions on the convergence of the centralized algorithms ~~are~~ provided in Appendix B on page 14, col. 1 and for the distributed algorithms in Section IV-C on page 8, col. 2.

Q-(5) *At the end of Section III, you mentioned that the proposed reduced complexity resource allocation scheme is sensitive to the order in which the subchannels are selected for the optimization problem. Please provide a discussion how to choose this order.*

Resp : ~~We understand the reviewer's concern.~~ Since the algorithm designs the precoders for each sub-channel at a time by using the total number of unserviced packets. For designing the precoders for the sub-channel $j+1$, we assume the transmit precoders and the rates of all users are obtained for the sub-channels $\{1, 2, \ldots, j\}$. Now, the number of backlogged packets for the $j+1^{\text{th}}$ sub-channel is given by

$$Q_k - \sum_{i=1}^{j} \sum_{l=1}^{L} t_{l,k,i}$$

. Since it depends on the rates of the already completed sub-channels $\{1, 2, \ldots, j\}$, it is susceptible to the ordering used to determine the precoders in each sub-channels. It is discussed in detail in Section III-D on page 7, col. 1.

Q-(6) *In the distributed algorithms, it is not clear what exact information is exchanged between the BSs or between the BSs and users. Moreover, the signaling overhead should be analyzed and compared with the centralized solution. The proposed distributed algorithms require exchanging over-the-air signaling or backhaul signaling for many times within each channel coherent time (e.g., from Fig. 2, the distributed algorithm requires 20-30 iterations to converge even when there are only 3 subchannels). I dont think this is acceptable in practice. Is the signaling overhead of the distributed algorithm really smaller than the centralized algorithm which only requires exchange the CSI between the BSs for once within each channel coherent time?*

Resp : The distributed algorithms are derived for the convex subproblem, which leads to the same stationary point asymptotically as that of the centralized solution, but indeed we would require large number of iterations for the convergence. In reality, we have to limit the number of iterations required for each distributed algorithm, thereby leading to a point which may not be the same stationary point as when the algorithm is allowed to converge. The number of ADMM or primal updates are controlled by the parameter $J_{\max}$ in the Algorithm 2. In practice, we can combine the ADMM or primal update, receiver update and the SCA update all at once, *i.e*, $J_{\max} = 1$ to minimize the overhead in the backhaul signaling and to speed up the convergence. To this end, we have ~~provided~~ a practical scheme in Section IV-D on page 9, col. 1. Note that in the time-correlated case, it is often also enough to update the precoders once per radio frame. The decentralized schemes are never allowed to converge until the end, it is only important to follow the fading process when $J_{\max} = 1$. The performance of the distributed algorithm based on dual decomposition scheme is discussed for the time-correlated fading in Section C of [1], which shows that it is enough for the distributed precoder design to follow the fading process to provide desired performance.

Q-(7) *The convergence analysis of the distributed algorithms is not clear. For example, what is the exact condition to ensure the convergence of the distributed algorithms. Does the distributed algorithms also converge to a stationary point?*

Resp : We have updated the manuscript to include the discussions on the convergence of the distributed algorithm on page 8, col. 2 last paragraph in Section IV-C. The condition for the convergence of the distributed algorithm is that when the dual variable updates are not significant or when the decrease in the objective value is below some threshold. The distributed algorithm reaches the same stationary point as of the centralized algorithm if we let the inner loop in Algorithm 2 on page 8, col. 2 to converge.

Q-(8) *Im totally confused with the ADMM approach in Section IV.A. Many notations, such as the local interference vector and consensus interference vector are used without formal definition. What is the difference between the local interference vector and consensus interference vector? What are their relationships with the actual interference vector. It seems that you are using the same notation for all of these interference vectors and I cant tell when a notation refers to a local interference vector, a consensus interference vector, or the actual interference vector. These questions should be clarified and perhaps you should choose the notation system more carefully. For example, in (36), there are 3 similar notations and I dont know which one is local interference vector and which one is the actual interference vector.*

Resp : ~~We understand the reviewer's concern.~~ We have shortened the section on ADMM by citing the relevant earlier work on the distributed implementation of the centralized algorithm for the min power problem [29]. We have rewritten the distributed algorithms in Section IV-A and Section IV-B. The variables are discussed in Section IV on page 8, col. 1. It follows the existing literature on the ADMM scheme, which treats the interference as a variable at each BS and the consensus on the interference is achieved upon convergence of the distributed algorithm as discussed in [11].

Q-(9) *In the distributed algorithms, it is not clear what information is available at each node. For example, what are your assumption on CSIT (CSI knowledge at each BS) and CSIR (CSI knowledge at each user)? How to 2 obtain the information used to perform the required calculation at each node (such as calculating the actual interference, MMSE receiver and the dual variables)?*

Resp : ~~We understand the reviewer's concern.~~ For the distributed precoder design, we assume that each base station (BS) b knows the equivalent downlink channel $\mathbf{w}_{l,k,n}^{\mathrm{H}}\mathbf{H}_{b,k,n}$ of all users in the system by using precoded uplink pilots, where the precoders are the MMSE receiver of all the users. Note that it includes the cross equivalent downlink channels of the neighbor BS users as well. To update the MMSE receiver, the equivalent channel for the k^{th} user $\mathbf{H}_{b,k,n}\mathbf{m}_{l,k',n}, \forall k' \in \mathcal{U}_b, \forall b \in \mathcal{B}$ is obtained from the BSs through user specific downlink precoded pilots. We have included the information on what each network entity knows in page 7, col. 2 last paragraph. We have included the type of duplexing scheme adopted in the model, *i.e*, TDD system, in the system description in Section II, last paragraph. We have included the type of duplexing scheme adopted in the model, *i.e*, TDD system, in the system description in Section II, last paragraph.

Q-(10) *Do you have any convergence result for the proposed distributed solution based on the KKT conditions in Section IV.B? It seems that the iterative method to solve the KKT conditions is totally heuristic.*

Resp : It is true that the KKT conditions are based on the centralized algorithm in (20). It follows the same points as that of the centralized approach if the algorithm is iterated in the same order, *i.e* the dual variables are allowed to converge before the update of the fixed operating point and the receiver. Here instead we perform the update of the transmit precoders, receive beamformers and the dual variables all at once in each iteration. Thus, we sacrifice the formal convergence for the improved speed of convergence. It may not be a stationary point of the original nonconvex problem but it is guaranteed to provide better performance in the sum rate compared to the distributed approaches presented in Section IV-A and Section IV-B for the same number of iterations. We have provided additional details on the convergence of the KKT based approach in page 10, col 1, last paragraph.

Q-(11) *Since queue is a dynamic system evolving according to (3), it doesnt make sense to compare the queue deviations at a given time. You should compare average queue deviations in the simulations. Moreover, you should also compare the average delay performance instead of just comparing the performance metric (queue deviations) defined in this paper. Using the queue deviations as the performance metric also needs more justification.*

Resp : We thank the reviewer for raising the critical comment. Since the manuscript is about the precoder design to minimize the total number of backlogged packets in the system, we have provided the convergence behavior of the proposed precoder design formulations for a given time instant. In accordance with the reviewer comment, we have included the Section V-C on page 13, col. 1 ~~to~~ discuss the queue deviation over multiple transmission slots. We have presented Fig. 4a ~~by~~ comparing the average number of backlogged packets for different algorithms with various arrival rates and Fig. 4b for the number of backlogged packets at each instant. Note that the question on delay arises when we are considering the resource allocation over certain duration. Since the paper is about the precoder design to minimize the number of backlogged packets at each instant, it may not be a valid performance metric for our objective. We still agree that delay can be controlled by reducing the number of packets on an average, as we can see from Fig. 4, on page 13, the proposed algorithms are equally good in minimizing the average number of backlogged packets. Note that we can also prioritize the users by controlling the variable a_k in (6a) to address the QoS constraints for a particular user, in addition to that, we can also change the objective to ℓ_2 and ℓ_∞ norm to address the delay and the fairness implicitly.

Q-(12) *What is SRA in the simulation figures?*

Resp : The spatial resource allocation (SRA) is updated in the revised manuscript on page 7, col. 1, Section III-D.

Q-(13) *In the discussion for Fig. 1, you mentioned that JSFRA converge to the optimal point, and all algorithms are Pareto-optimal. Since the problem is non-convex, why these algorithm can find optimal solution or Paretooptimal point?*

Resp : ~~We thank the reviewer for pointing out the mistake in the text.~~ Since the problem is nonconvex, the JSFRA formulation can find a local optimal point upon convergence. The converged point of the JSFRA problem is in fact the stationary point of the original nonconvex problem, which is discussed in Appendix B-D on page 15, col. 2. We have removed the statement mentioning the pareto-optimal solutions in the discussions on Fig. 1 in the revised manuscript.

REFERENCES

[1] A. Tölli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling," *IEEE Transactions on Wireless Communications*, vol. 10, no. 2, pp. 570–580, 2011.

RESPONSE FOR REVIEWER - 2

We would like to thank the reviewer for providing valuable comments. We have modified the manuscript accordingly by addressing it.

Q-(1) *The logic from (6) to (16) is not clear. The only difference is the two newly introduced NON-CONVEX constraints (16b) and (16c), while the objective function (16a) and the constraint (16d) is the same as (6). The equivalence between (6) and (16) is not straightforward and it is confusing why the reformulation in (16) is beneficial.*

Resp : ~~We understand the reviewer's concern.~~ The reformulation is required, since the SINR expression in (2) cannot be handled directly in the problem defined in (6). Note that the equality constraint imposed by the SINR expression in (2) is handled by the two explicit inequality constraints (16b) and (16c), which are the under estimator to the original problem in (6). Since the inequalities are used to replace the equality constraint, the objective could be improved by the iterative algorithm and the equality will be achieved after the convergence, thereby having one-to-one mapping at the final solution. We have updated the manuscript to include the details for better clarity on page 5, col. 1, first paragraph.

Q-(2) *The authors use the successive convex approximation framework, but the approximate problem proposed by the authors is actually not convex. Inspecting (19), its objective function is the same as in (6), and the non-convexity of (6) comes exactly from the objective function, so (19) is not a convex problem. The same flaw is repeated several times in the approximate problems proposed by the authors.*

Resp : Please note that the objective function is convex, which can be verified by writing the difference function inside the norm operator in epigraph form. Since the problem in (16) is not jointly convex on the variables $\mathbf{m}_{l,k,n}$ and $\mathbf{w}_{l,k,n}$, we use alternating optimization (AO) approach by fixing $\mathbf{w}_{l,k,n}$ and optimize for $\mathbf{m}_{l,k,n}$ as a variable. Even after fixing $\mathbf{w}_{l,k,n}$ as constant, the problem in (16) is nonconvex due to the DC constraint (16b), which is handled by the first order relaxation around some fixed operating point. Once the linear relaxation is performed, the problem in (20) is a convex optimization problem with the variables being $\mathbf{m}_{l,k,n}, \gamma_{l,k,n}, \beta_{l,k,n}$. We have updated the manuscript to illustrate this clearly on page 5, col. 2, first paragraph.

Q-(3) *The authors proposed to use block coordinate descent method to solve (16). But as the authors have already pointed out, to apply block coordinate descent method, the constraint sets for different variables should be disjoint (uncoupled), which is however not the case in (16), because receive and transmit precoders (i.e w and* *m) are coupled in the constraints. It is confusing on its own why the authors made a statement that contradicts the proposed methodology, and the convergence followed is in question.*

Resp : We thank the reviewer for the pointing out the flaw in the text. We have removed the incorrect statement on block coordinate descent method with the AO approach on page 5, col. 1, second paragraph. Indeed, due to the coupling of the transmit and the receive precoder variables, we cannot use the standard block coordinate descent method proof for the convergence of the proposed algorithm as pointed out by the reviewer. We have provided a completely rewritten convergence proof ~~o~~n Appendix B, page no. 14, col. 1.

Q-(4) *Regarding the convergence of the SCA, the authors cited [27] for the convergence conditions, but the reference is wrong, because the conditions after the three bullets on page 6 are not mentioned in [27]. In case the authors disagree, please make the citation more specific, for example, specify the theorem/statement/proposition in [27] where those conditions are specified.*

Resp : We apologize for the inappropriate reference cited in the original manuscript. We have provided a completely rewritten proof on the convergence of the centralized algorithm ~~o~~n Appendix B.

Q-(5) *The authors also cited [28] to establish the convergence of SCA. But the techniques of [27] and [28] are different, and the convergence conditions are different too. It is not clear why the authors need two set of convergence conditions for a single problem, and the resulting convergence analysis itself is not solid enough.*

Resp : We agree with the reviewer's comment. We have provided the updated proof for the convergence of the centralized algorithm on Appendix B in the revised manuscript.

Q-(6) *Another comment on reference: to the reviewer's knowledge, the term SCA is never explicitly used in [2]. So please either correct the reference or be more specific (section, theorem, etc.).*

Resp : We have removed the inappropriate citation of the references.

Q-(7) *The authors propose primal decomposition method, ADMM approach to the non-convex problem (19), while their convergence analysis is based on literature that proved convergence for convex problems only, e.g., [13]. So the convergence analysis is not trustworthy.*

Resp : ~~We understand the issue raised by the reviewer.~~ Please note that the distributed algorithm is performed for the convex subproblem presented in (20) and (28). Since the problem is convex, the distributed approach convergence can be guaranteed by satisfying certain conditions. These are discussed in Section IV-C on page 8, col. 2, last paragraph.

Q-(8) *The length of the paper is too extensive. Some of the reformulations as mentioned in the previous comment can be skipped. Also, Section III.D. is not deeply explained and does not bring additional value to the paper. The implications of ordering the sub-channels for the iterative approach should be carefully studied and extensively explained in a different publication.*

Resp : ~~We understand the reviewer's concern.~~ We have included the sub-channel wise resource allocation or (SRA) scheme in Section III-D for the completeness. It is presented in the manuscript as an alternative suboptimal approach to perform sub-channel wise distributed precoder design by the centralized controller. We have updated the manuscript with more details for better understanding.

Q-(9) *Information regarding the value of q used to obtain the simulation results is missing (with exception of Fig. 3).*

Resp : We have included the information regarding the norm used for the simulation in the figure captions. It is updated for Fig. 1 and Fig. 2.

Q-(10) *In Fig. 1 and Fig. 2, the labels for the system model do not fit with the written description. Additionally, the reference scheme Q-WSRM is not optimal, since it over allocates resources if there are few queued packets. Therefore, it is not interesting for comparison purposes.*

Resp : We understand the reviewer's concern. We have updated the manuscript to include the descriptions in the text to refer the legends used in the figures. Table 1 is provided for the purpose of insight, since it deals with the SISO scenario with 3 sub-channels and 3 users.

Q-(11) *Assuming that Fig. 2 and Fig. 3 where obtained based on the same simulation setup, i.e. user queues, number of transmit and receive antennas and number of base stations, it is not clear why results in Fig. 3 are worse than Fig. 2 when comparing JSFRA. Even more, since the number of sub-channels is larger in Fig. 3, the result seems contradictory.*

Resp : We thank the reviewer for pointing out the issue in the system description for the figures. Please note that the simulation scenario is different for Fig. 2 and Fig. 3 in terms of path loss distribution and the sub-channel numbers. We have provided different scenario with the centralized algorithm as reference for studying the performance under different system model. The user path loss is distributed uniformly between $[0, -6]$ dB for Fig. 2 and between $[0, -3]$ dB for Fig. 3. We have updated the manuscript to include these details. In addition, the number of backlogged packets in the system for Fig. 3 is $[9, 16, 14, 16, 9, 13, 11, 12]$ bits. These

~~informations~~ are updated in the revised manuscript on page 12, col. 1, lines 3-6 in Section V-B and page 12, col. 2, third paragraph in Section V-B.

RESPONSE FOR REVIEWER - 3

This manuscript focuses on the beamforming and scheduling optimization for IBC MIMO-OFDM system, including the centralized and decentralized optimization methods. This is an interesting and important topic.

We thank the reviewer for reading the manuscript and providing valuable comments. The comments are really helpful in improving the manuscript.

Q-(1) *The number of transmitted packets t_k's are optimization variables, which should be explicitly stated in the problem formulation of (6), (16), (19), (20) and (26) to avoid confusing.*

Resp : Please note that the objective function uses $v_k = Q_k - t_k = Q_k - \sum_{n=1}^{N} \sum_{l=1}^{L} \log_2(1 + \gamma_{l,k,n})$ expression instead of including an additional constraint for the transmitted packets using the rate expression and thus t_k is not a variable. However, in the MSE formulation, we have explicitly stated the optimization variable t_k, since it is present in the DC constraint (27).

Q-(2) *The manuscript states that the inequalities (16b) and (16c) achieve equality at optimality(line 23, page 5). This is not obvious. An easy case to check this statement is that assuming the system has two BS and each BS serves one user. When $Q_1 = 0$ and 2nd BS has sufficiently large power, (16b) and (16c) do not hold equality. Rigorous proof is needed if authors stick to this statement.*

Resp : We thank the reviewer for the insightful comment. We have updated the manuscript to include the statement that the proposed approximation in (16b) and (16c) are the under-estimator for the SINR expression in (2). The under-estimator is tight when there is at least one user in each BS that cannot be served by the current transmission. It has multiple solutions, when the objective is zero even for a single BS. In this case, we can use a regularization term with the total power in the objective to obtain an unique solution. We have discussed the uniqueness of the proposed algorithm in detail in Appendix B-C.

Q-(3) *The solution in (21) is obtianed for MMSE, i.e. for 2-norm(q=2). If $q = 1$ or $q = \infty$, it is actually an equivalent linear programming problem. Details for this solution should be provided.*

Resp : We thank the reviewer for the critical comment. Note that the receiver has no explicit relation with the choice of ℓ_q norm used in the objective function. The dependency is implicitly implied by the transmit precoders $\mathbf{m}_{l,k,n}$, which in~~-~~deed depend on the q ~~value~~. We have modified the text to include~~d~~ this information in page 5, col. 2, last paragraph.

Q-(4) *The convergence proof need to be rigorous. The inequality of (23a) is opposite to the reference [28]. Also the* *statement on uniqueness of the transmit and the receive beamformers are not correct. Although we can choose one antenna to be real value, this does not mean the problem has unique solution!*

Resp : We agree with the reviewer. We have updated the manuscript to discuss the convergence proof in a rigorous manner in Appendix B. The uniqueness is also justified implicitly by the linear approximation performed on the DC constraint (16b), which makes the transmit precoders to be susceptible to the phase rotation. The uniqueness discussions are presented in Appendi~~x~~ B-C.

Q-(5) *(25) is generally wrong. (25) only holds when the MSE is minimized(by MMSE receiver) and the snr is the optimized(which is obtained by general eignvalue decompostion). This is clearly stated in the reference [5] and [6]. This can also be easily checked by comparing (25) and (2). Consequently the alternative formulation (26) based on this conclusion is questionable.*

Resp : We agree that the MSE equivalence with the SINR expression is valid only when the receiver is based on MMSE objective. In our MSE reformulation ~~solution~~, we have used the MMSE receiver irrespective of the ℓ_q norm used in the objective. Since the receivers are based on the MMSE objective, the equivalence is valid between the MSE and the SINR expression. Since the transmit precoders are updated in accordance with the objective, the MSE reformulation yields the same solution as the SCA approach based JSFRA scheme presented in Section III-B. It can also be verified from Fig. 1 for ℓ_1 objective. We have updated the manuscript to include this comments on page 6, col. 2, line 10-13.

Q-(6) *For ADMM aproach, the determination of the value of ρ in equation (35a) should be discussed. 1. The numbers of transmitted packets for users t_k's are optimization variables. So they should be explicitly stated in the problem formulation (6), (16), (20) and (26) to avoid confusing.*

Resp : We agree with the reviewer's concern. We have included ~~the~~ reference [11] that discusses on the valid choices of the step size parameter for ADMM. We have included the statement that it depends on the system model under consideration. It is updated in the manuscript on page 8, col. 2, line 4-8.

RESPONSE FOR REVIEWER - 4

First, we thank the reviewer for providing valuable and insightful comments. The comments are really helpful in improving the manuscript.

Q-(1) *First, this reviewer is not convinced by the arguments for showing the convergence of the JSFRA method. "The SCA method" is often referred to, but never really defined or referenced. The three required conditions (as stated on p. 6, col. 1, rows 38-40) do not, as far as I can tell, appear in [27]. Indeed, [27] is concerned with optimization problems where the objective function is non-convex, but the constraint set is convex and separable over the blocks of variables. Perhaps you meant to cite [A], wherein non-convex constraints are handled in a similar way? Numerically, the algorithms do converge, and the argument put forward makes sense, but the treatment must be improved to be more rigorous.*

Resp : We understand the reviewer's concern on the lack of rigorous convergence proof for the JSFRA approach. We have presented a rigorous convergence proof for the proposed algorithm in Appendix B on page 14. We have included the reference suggested by the reviewer in [34].

Q-(2) *Second, the optimization problems formulated only depend on Q_k, the current levels of backlogged packets, and not on the arrival rates. This is due to how the conditional Lyapunov drift is minimized. This approach completely removes the queue dynamics from the optimization problem, essentially leading to greedy one-shot optimization in 7 every time instant. The framework would be more interesting if some sort of optimization (or tracking) is performed over several time instants, rather than the one-shot approach that is currently used for the JSFRA algorithms. Possibly, some expectation over the queues would be optimized then. Even if no analytical treatment of the tracking over several time-steps is added, I would at least highly recommend adding some simulation results where the proposed one-shot algorithms are performed sequentially over several time instants.*

Resp : We thank the reviewer for the useful suggestion. In this paper, we proposed the precoder design to minimize the number of backlogged packets in the system at a given time. Since the greedy approach is performed at each time instant, it minimizes the number of backlogged packets also on an average. We focused on the precoder design based on the current knowledge of the channel state information in the simulation results. In order to justify the gains over multiple time instants, we have now included Section V-C on page 13 to discuss on the performance of the JSFRA scheme for different ℓ_q norm over multiple time slots. Plots comparing the average number of backlogged packets for different schemes at each instant and the number of queued packets remained in the system after each transmission instant are provided in Fig. 4.

Q-(3) *Third, the distributed methods (at least the primal decomposition and ADMM) seem to be fairly straight-forward applications of existing results. This reviewer recommends spending more space on the convergence, than on the description of the distributed techniques. Still, it would be nice with a direct description of what local CSI is required, and how it is acquired, to perform the local computations for the primal decomposition and ADMM methods. For the description of the signaling of the CSI in Sec. IV-B, are you envisioning a TDD system?*

Resp : We agree with the reviewer's comment. We have shortened the discussions on the distributed approaches and discussed the convergence proof of the decentralized algorithm in Section IV-C page 8. For the distributed precoder design, we assume that each BS b knows the equivalent downlink channel $\mathbf{w}_{l,k,n}^{\mathrm{H}}\mathbf{H}_{b,k,n}$ of all users in the system by using precoded uplink pilots, where the precoders are the MMSE receiver of all the users. Note that it includes the cross equivalent downlink channels of the neighbor BS users as well. To update the MMSE receiver, the equivalent channel for the k^{th} user $\mathbf{H}_{b,k,n}\mathbf{m}_{l,k',n}, \forall k' \in \mathcal{U}_b, \forall b \in \mathcal{B}$ is obtained from the BSs through user specific downlink precoded pilots. We have included the information on what each network entity knows in page 7, col. 2 last paragraph. We have included the type of duplexing scheme adopted in the model, *i.e*, TDD system, in the system description in Section II, last paragraph.

Q-(4) *Finally, some readers might be confused by the "joint space-frequency" terminology, believing that the beamforming is performed over a joint space-frequency channel space, where the space-frequency channels are formed by block-diagonal matrices, each block belonging to one subcarrier. This could easily be clarified.*

Resp : We agree with the reviewer's point. We have included the note on the space-frequency terminology in the introduction section on page 2, col. 2, line 2-4.

Q-(5) *Please see the itemized questions*

(a) *- p. 1, col. 1, row 42: "userss"*

Resp : Corrected in the revised manuscript in Section I on page 1, col. 1, line 2.

(b) *- p. 1, col. 2, row 18: the precoders are used _implicitly_ as decision variables. This is the whole point, to avoid explicitly modeling the hard decisions in the optimization, and instead do soft decisions during the iterations, and then finally hard decisions after convergence.*

Resp : We thank the reviewer for highlighting the important point. We have included the statement in the revised manuscript on Section I, page 1, col. 2, line 5-14.

(c) - *p. 1, col. 2, row 33: Which chapter in [2] is referred to? With a quick look-through of the table of contents, I can't find and chapter or section treating the SCA method?*

Resp : We apologize for the wrong reference. We have updated the manuscript to include the proper reference on SCA with [2,3,7,25,26,35].

(d) - *p. 2, col. 2, row 36: Write rank*$(.)$ *and* $\min$ *instead*

Resp : We have updated the revised manuscript with the reviewer suggestions on page 2, col. 2, second paragraph in Section II.

(e) - *p. 3, col. 1, row 26: It would be more clear to explicitly write out the dependence of* $\mathbf{M}$ *and* $\mathbf{W}$ *in* $\tilde{v}$ *here*

Resp : We removed the matrix representation of the transmit precoders and the receive beamformers from the manuscript to avoid confusion as suggested by the reviewer.

(f) - *p. 3, col. 2, row 26: Which general MIMO-OFDM problem are you talking about here, and what is combinatorial about it? Is it the problem of selecting users to be served on orthogonal subcarriers? There is nothing inherently combinatorial over the problem in (6) as far as I can tell, as the beamformers are used as soft decision variables.*

Resp : We thank the reviewer for pointing out the confusion. We have removed the word "combinatorial" from the text to avoid confusion in page 3, col. 2, initial sentence in Section III.

(g) - *p. 4, col. 2, row 40: "In fact, (5) provides similar expression of ..." This sentence is very hard to understand.*

Resp : We have rephrased the sentence to avoid difficulty in the understanding. The manuscript is updated with the revised text on page 4, col. 2, first paragraph in Section III.B.

(h) - *(16d): suggest your write out the power constraints here, in order to be faster be able to interpret the optimization problem. There is hardly any spaced saved by referring back to (6b).*

Resp : We have updated the manuscript with the explicit power constraint in (16d) as

$$\sum_{n=1}^{N}\sum_{k\in\mathcal{U}_b}\sum_{l=1}^{L}\operatorname{tr}\left(\mathbf{m}_{l,k,n}\mathbf{m}_{l,k,n}^{\mathrm{H}}\right) \leq P_{\max}, \forall b$$

in page 5, col. 1.

(i) *- p. 5, col. 1, rows 27-30: Here you might want to quickly mentioned how one could show the NP-hardness of (16).*

Resp : We have included the NP-hardness of the proposed solution by modeling the current problem to solve the weighted sum rate maximization (WSRM) formulation, which is known to be NP-hard in page 5, col. 1, last sentence in Section III-B third paragraph.

(j) *- p. 5, col. 1, row 50: "According to the SCA method...". I am not sure exactly how you define "_the_ SCA method"? Clarify or cite the definition.*

Resp : We have removed the sentence mentioning SCA method. We have updated the manuscript to discuss this as SCA approach in pagt 5, col. 1 last paragraph.

(k) *- p. 5, col. 2, row 31: Here is a case where it makes sense to reference earlier optimization constraints. However, are (19d) and (18) not the same??*

Resp : We thank the reviewer for pointing out the mistake. We have included all the constraints except the linearization constraint in (20) which is (19) in the original manuscript. Expression (19d) and (18) are same. We have removed this double reference in page 5, col. 2, ~~equation~~ (20).

(l) *- p. 5, col. 2, row 51: Slightly confusing with the notation between the iterates in (21b) and the MMSE filter in (22b).*

Resp : We have updated the manuscript to avoid the ambiguity in the representation of the optimal and the MMSE receiver. The optimal receiver is denoted by $\mathbf{w}^o_{l,k,n}$ in (22) and the MMSE receiver by $\mathbf{w}_{l,k,n}$ in (23).

(m) *- p. 6, col. 1, row 9: You might want to add somewhere that (22b) can be used instead of the fixed-point of (21b), since the scaling of the receive filters do not matter in the SINRs. However, does it affect the convergence of the algorithm?*

Resp : We thank the reviewer for providing the valid comment. We have updated the manuscript to include this discussion in page 6, col 1, lines 3-6. The performance remains the same and the convergence rate is not affected by this scaling, which can be seen from Fig. 1 comparing the precoder convergence using the optimal and the MMSE receiver for the JSFRA scheme with ℓ_1 norm.

(n) *- p. 6, col. 2, rows 8-10: I don't fully understand the reasoning on the relation between the constraint sets in the different iterations. Why is this the case?*

Resp : To solve the nonconvex problem (16), we linearize the DC constraint (16b) around a fixed operating point. Since the operating point happens to be the optimal solution from the earlier iteration, the optimal point of the previous iterations is also present in the current feasible set. Therefore, at each iteration, the algorithm finds a better solution or the same compared to the previous solution, which leads to the monotonically decreasing objective. It is explained in detail in Appendix B on the convergence analysis of the centralized algorithm.

(o) - *p. 7, col. 1, row 35: Just because a problem is convex does not mean that it has a unique solution. (Although it seems to me that (26) should have a unique solution.) Is the problem in (26) strictly convex?*

Resp : We thank the reviewer for the critical comment. It is true that the convex problem does not need to have a unique solution in general. By using Minkowski inequality $\|x+y\|_q \leq \|x\|_q + \|y\|_q$, we can show that the JSFRA problem is not strictly convex for all exponents ℓ_q used in the objective function. The uniqueness of the solution is obtained due to the linear constraint (19) in the problem formulation (20). Note that the linear constraint (19) is susceptible to the phase rotations on the transmit precoders, whereas any unitary rotation~~s~~ on the transmit precoders still satisfies the original nonconvex constraint (16b). When the constraints (19),(20b) and (20d) are all active for all BSs, *i.e*, when the objective is nonzero for all BSs, we obtain unique set of transmit precoders. On contrary, it has multiple solutions when the objective is zero even for a single BS. In this case, we can use a regularization term with the total power in the objective to obtain an unique solution. We have discussed the uniqueness of the proposed algorithm in detail in Append~~ix~~ B-C.

(p) - *Table 1: "backpreassure"*

Resp : It is updated in the manuscript on page 11, Table 1.

(q) - *p. 11, col. 1, row 56: "performances". I'm not sure this is a countable noun.*

Resp : We ~~agree with the reviewer. We have modified the manuscript accordingly by replacing it with the singular form~~.