Reviewers Comments & Authors Replies

Manuscript No.	Paper T-SP-18051-2014, submitted to *"IEEE Transactions on Signal Processing"*
Title	"Traffic Aware Resource Allocation Schemes for Multi-Cell MIMO-OFDM Systems"
Authors	Ganesh Venkatraman, Antti Tölli, Markku Juntti, and Le-Nam Tran

The authors would like to thank the associate editor and the reviewers for their valuable comments on the manuscript, which have been greatly helpful to improve the paper quality. Based on the comments, we have made several revisions to the paper, which are summarized below.

1. We have shortened the discussions on the existing *backpressure algorithm* in Section III-A.
2. Initialization of the successive convex approximation (SCA) operating point is included in the revised manuscript
3. Section III-D is updated to include the guidelines involved in selecting the sub-channel order.
4. Strong convexity of the objective function is emphasized in Appendix A-B.
5. Strict monotonicity of the algorithm is discussed in a detailed manner in Appendix A-C.
6. We have provided additional details on the distributed algorithm convergence as suggested by the reviewers

In what follows, the comments are listed, each followed immediately by the corresponding reply from the authors. The reviewers questions in the revised manuscript are highlighted in blue color and the authors responses are presented in black. Unless otherwise stated, all the numbered items (figures, equations, references, citations, etc) in this response letter refer to the revised manuscript. The revisions in the manuscript are highlighted in blue color.

Response to Reviewer - 1's Comments

Comments: The response to the reviewer's concerns are generally satisfying, except the convergence proof.

Reply: We thank the reviewer for providing valuable and insightful comments.

For the resubmitted manuscript, the reviewer still has the following concerns

1. Considering the length of the manuscript, it would be better to shorten some parts that are not new in this manuscript, e.g. III.A. More space can be left for convergence proof, which is very important.

Reply: We have removed a couple of paragraphs from Section III-A and shortened the discussions on Section V (*i.e.*, the simulation results) to provide additional details on the convergence proof. We believe the removed parts do not affect the readability of the paper and the added text regarding the convergence analysis certainly improves the quality.

2. In convergence proof (48), why does the 2nd inequality hold? In fact, to prove the feasibility of $m_{k+1}^{(i)}, w_*^{(i-1)}; m_k^{(i)}$, the part between 2nd and 3rd inequality is not necessary, ≤ 0 directly follows the 2nd inequality since the solution $m_{k+1}^{(i)}, \gamma_{k+1}^{(i)}$ is the optimal solution, and therefore feasible.

Reply: We have removed the additional inequality relating the previous operating point from (49) in the revised manuscript.

3. The solutions SCA iterations $\mathbf{m}_k^{(i)}$ does not necessarily converge. In fact $\mathbf{m}$ has compact feasible region, and thus $\mathbf{m}_k^{(i)}$ has limit points for any specific i. However $m_*^{(i)}$ does not necessarily exist (the whole sequence $\mathbf{m}_k^{(i)}$ may be not convergent). Similar problem happens to $\mathbf{w}_k^{(i)}$.

Reply: We thank the reviewer for the comment. It is true that the sequence of iterates generated by the iterative SCA algorithm, namely, $\mathbf{m}_k^{(i)}$ and $\mathbf{w}_k^{(i)}$ need not converge even if the objective sequence converges. It follows from the fact that there can be multiple minimizers (solutions) for the convex subproblem in each SCA step. However, we have modified the convergence discussion using the strong convexity argument by regularizing the objective of each SCA subproblem with a quadratic term as discussed in [30] and [31]. The additional quadratic term ensures the uniqueness of the minimizer in each of the SCA subproblem. Due to the uniqueness of the minimizer and the strict monotonicity of the objective sequence in each SCA step, the convergence of the iterates is guaranteed. The above discussion is included in Appendix A-B around (46) and in Appendix A-C after (50).

4. Strict monotonicity with respect to the objective function f should be rigorously proved. Note that to guarantee the uniqueness of the beamformer iterates, (52) instead of the objective function is used.

Reply: As suggested by the reviewer, we have included the argument to show the strict monotonicity of the objective sequence in Appendix A-C following (52). Moreover, we have also mentioned that we in fact use the regularized objective in (46b) to analytically prove the convergence of the iterates, although we have numerically observed that the objective sequence of Algorithm 1 (with the original objective) always converges. Please refer to the paragraph before Section III-C.

5. Note that the conclusions [32, Thm 2] and [26, Thm 10] have lots of assumptions. To invoke these reference, explicit exposition should be provided to show that these conclusions can be applied to our problem. The same questions occur to the proof in Appendix B, where conclusions in [11] [36] and [37] are used. Too many details are omitted to make the proof convincing and clear.

Reply: We have updated the manuscript to include the details regarding the stationary point discussion in Appendix A-E. The convergence analysis for the primal and the alternating directions method of multipliers (ADMM) algorithms are presented in Appendix B. We have also mentioned the conditions required for showing the stationarity property of the limit point and included more mathematical steps to make the convergence proof of distributed approaches rigorous.

Response to Reviewer - 2's Comments

The authors have addressed many of my previous comments. However, there are still several major issues that need further clarification.

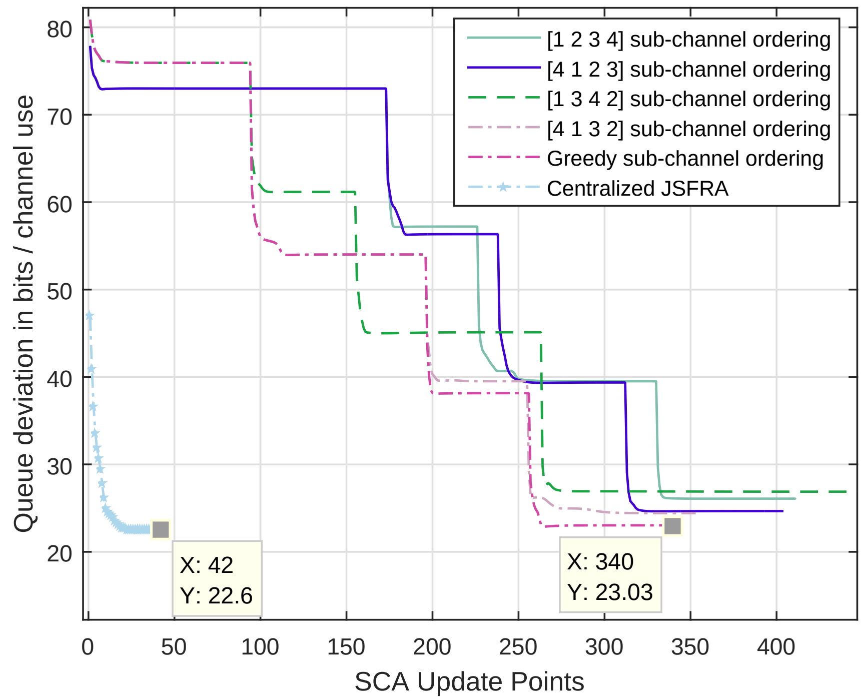
Reply: We thank the reviewer for providing further useful comments which help us to greatly improve the manuscript.

1. The revised paper did not address my previous comment about how to select the sub-channel ordering. I understand that finding the best sub-channel ordering requires exhaustive search which has extremely high complexity. But it is important to provide a guidance on what would be a good choice of sub-channel ordering. For example, can we achieve a good performance by using a low complexity ordering algorithm such as a greedy sub-channel ordering algorithm?

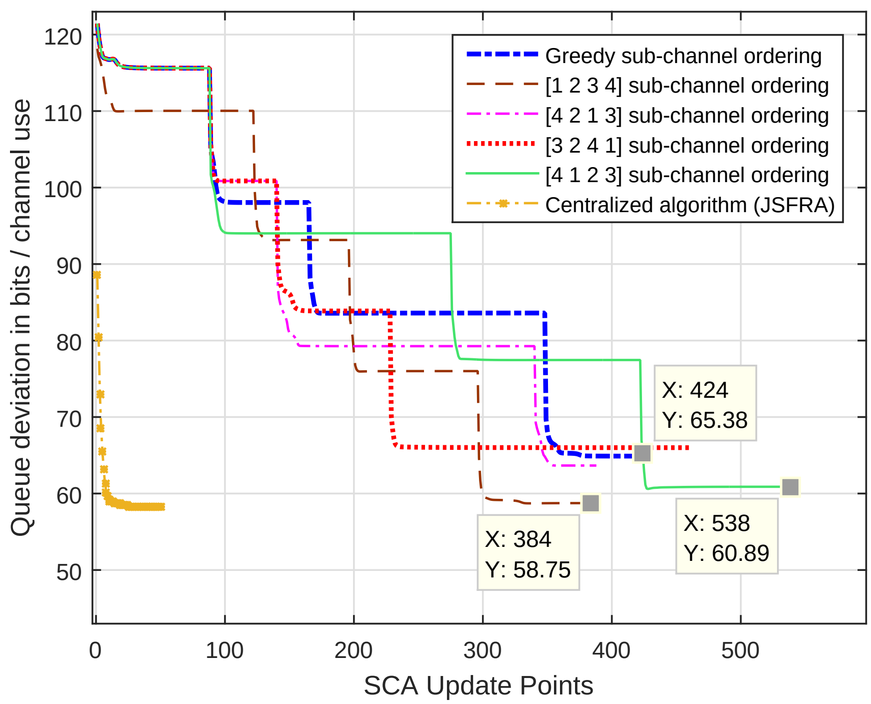
Reply: We apologize for not clarifying the sub-channel ordering in detail in our previous manuscript. For simplicity we use random sub-channel ordering in our paper. That is, after finding the precoders for a current sub-channel, we can choose any previously unselected sub-channels as the next candidate sub-channel for which the precoders are identified using the updated backlogged packets. As suggested by the reviewer, the greedy sub-channel ordering can also be considered while selecting the order. The greedy ordering can be based on sorting the best channel gain from each sub-channel, which is obtained by finding the highest channel norm seen between any user and the corresponding serving base station (BS). However, note that the interference among the neighboring users are not included in this procedure, since the precoders that are required to evaluate the interference are designed after the sub-channel selection step only.

We remark that the choice of a sub-channel ordering scheme should also consider the number of backlogged packets associated with each user. We have emphasized this issue by comparing various ordering schemes for the same system model with two different set of backlogged packets associated with each user. We do not include the following figures in the manuscript due to space limitations, nevertheless, we have included in the response letter to answer the reviewers' question. We considered a system with $N = 4$ sub-channels, $N_B = 2$ BSs with $N_T = 4$ transmit antennas and $K = 12$ single antenna users. The path loss (PL) is distributed uniformly over $[0, -3]$ dB. The number of backlogged packets assumed for each user is provided in the corresponding captions in Figure I.

Figure I compares different sub-channel ordering schemes in terms of the total number of backlogged packets remaining in the system. As can be seen from Figure I(a), the greedy sub-channel ordering provides a favorable way of choosing sub-channels to minimize the total number of backlogged packets. However for another scenario in Figure I(b), it is evident that the greedy sub-channel ordering is no longer better in comparison to the few other random ordering schemes. Note that the difference between the two scenarios is only the queue distribution, which are listed out in the corresponding figure caption. We have included the discussions regarding the greedy sub-channel ordering in the revised manuscript as a heuristic approach under Section III-D final paragraph. However, it is worth noting that as the number of users in the system increases, all sub-channel ordering schemes are similar in minimizing the total number of backlogged packets in the system without any noticeable gain.



(a) Number of backlogged packets for each user in bits $Q_k = [11, 8, 14, 6, 6, 2, 10, 10, 5, 6, 9, 5]$



(b) Number of backlogged packets for each user in bits $Q_k = [8, 9, 12, 8, 12, 5, 4, 10, 8, 5, 7, 9]$

Figure I: Convergence of the algorithms for $\{N, N_B, K, N_T, N_R\} = \{4, 2, 12, 4, 1\}$ using ℓ_1 norm

2. The authors mentioned that the signaling overhead of the distributed algorithm can be reduced by using a smaller number of iterations $J_{\max}$. But still, you didn't answer my question about whether the signaling overhead of the distributed algorithm is smaller than the centralized algorithm. You should first analyze the signaling overhead of the distributed algorithm for fixed $J_{\max}$ and the signaling overhead of the centralized algorithm. Then you should point out under what $J_{\max}$ the distributed algorithm will have less signaling overhead than the centralized algorithm. Is it possible that the distributed algorithm always has more signaling overhead than the centralized algorithm even when $J_{\max} = 1$? Finally, there is a trade-off between performance and signaling overhead ($J_{\max}$) for the distributed algorithm. For the same signaling overhead (we can control $J_{\max}$ to make the signaling overhead of the distributed algorithm approximately equal to that of the centralized algorithm), does the distributed algorithm achieve better performance than the centralized algorithm?

Reply: We apologize for the lack of clarity in explaining this information in our earlier submission. To answer this question in detail, we consider the following scenarios based on the channel correlation between the adjacent transmission instants.

- At first we consider a semi-static scenario, where the channel remains constant for multiple transmission instants. Therefore, the channel state information (CSI) information available at the transmitter or the centralized controller are valid for multiple transmission instants. In order to discuss the performance of the centralized and the distributed schemes, let us consider a system model with $N = 4$ sub-channels, $N_B = 3$ BSs, each equipped with $N_T = 8$ transmit antennas. Let $K = 12$ be the number of users in the system equipped with $N_R = 1$ receive antenna. Let the PL between the BSs and the users are drawn uniformly between $[0, -3]$ dB. Figure II plots the number of backlogged packets present in the system after each SCA update point. Note that for $J_{\max} > 1$, each SCA points includes $J_{\max}$ number of ADMM iterations performed. Therefore, it needs to be considered while analyzing the signaling overhead. Note that we consider only ADMM based distributed approach outlined in Algorithm 2.

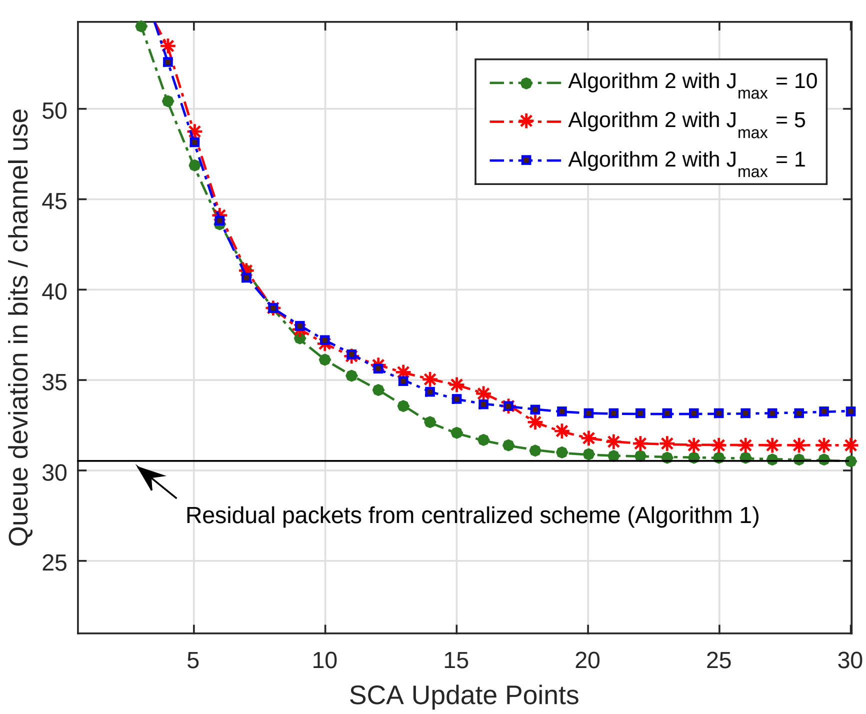


Figure II: Convergence of the centralized and distributed algorithms for $\{N, N_B, K, N_T, N_R\} = \{4, 3, 12, 8, 1\}$ using ℓ_1 norm with the backlogged packets as $Q_k = [17, 7, 14, 14, 12, 14, 8, 13, 8, 11, 9, 12]$ bits

Let us consider the centralized configuration, in which the CSI required to be exchanged between BSs and the centralized controller. In order to quantify the amount of signaling overhead, let us consider that each real entity of the complex channel is represented by $N_C = 10$ bits.[1] Using this information, the amount of signaling required to be exchanged between BSs and the centralized controller is given by $N \times N_B \times N_T \times K \times 2$ real entities, which is equal to $23,040$ bits. In addition, we also need the precoder signaling from the centralized controller to all BSs. It accounts to $N \times K \times N_T \times 2$ real entities, which is equal to $7,680$ bits. On the whole, it requires $30,720$ bits that needs to be exchanged via backhaul over T transmission slots, where T denotes the number of slots for which the channel remains constant. Additionally, we also need to update the queue information to the controller in each transmission instant and the updated precoders are required to be transmitted back. Therefore, even though the CSI is not required to be signaled to the controller, the precoders are fed back to all BSs in each transmission slot based on the current backlogged packets, which are updated by all BSs to the controller. Even though, we are not using quantized CSI in our performance evaluation, it is worth noting that the achievable gains deteriorates significantly if the precoders are designed using the quantized CSIs [R1].

On contrary, the distributed approach based on ADMM requires only the exchange of coupling interference variables that are scalar values. For a fully connected network, each consensus variable binds only two BSs only, therefore, the overall number of entities that needs to be exchanged is $N \times (N_B - 1) \times K \times 2$ real entries, which is equal to 960 bits by considering $N_D = 5$ bits to represent the scalar coupling variables. Note that 5 bits are more than sufficient for representing the interference variables since they are bounded by the transmit power. Therefore, in each transmission instant, the overall signaling required to design the precoders jointly requires $J_{\max} \times I_{\max} \times 960$ bits. Note that there is no need to exchange the backlogged packets to the coordinating BSs in the distributed scheme. Note that $I_{\max}$ denotes the number of SCA iterations and $J_{\max}$ represents the inner ADMM update steps.

Now, let us consider the performance of ADMM based distributed precoder design for different $J_{\max}$ values as in Figure II for analyzing the signaling overhead required and the performance improvement from the exchange of information. The centralized scheme is given as a lower bound in Figure II, since the CSIs available at the controller can be used to design the precoder until convergence. Figure II shows noticeable improvement in the reduction of number of backlogged packets in the system as we increase the number of ADMM iterations. However, note that as we increase the number of iterations, the signaling requirement also grows with the factor of $J_{\max}$.

From Figure II, it is evident that the ADMM with $J_{\max} = 1$ requires $I_{\max} = 15$ iterations to perform close enough to the centralized scheme. In order to reach the performance close enough to that of centralized scheme, using above assumptions, it requires $960 \times 15 = 14,400$ bits to be exchanged among the coordinating BSs. Note that it is ≈ 0.62 times the overhead involved in the centralized scheme to feedback only the CSIs to the controller, assuming single transmission instant. However, since the channel remains constant for certain duration, say T, the centralized approach may be favorable for this type of scenario. It is worth noting that the precoder exchange, which is required at each instant due to the changing traffic, involves significant overhead that also needs to be accounted while considering the selection.

However, if the channel changes once in every three transmission instants, *i.e.*, the fed back CSI is valid only for $T = 3$, then the distributed methods with $J_{\max} = 1$ is preferred to the centralized approaches. The duration of the CSI validity is evaluated using $B_{CSI} + B_P \times T = B_D \times T$, where $B_{CSI} = 23,040$ denotes the number of bits required for the CSI feedback, $N_C \times B_P = 7,680$ corresponds to the number of bits required for the precoder feedback from the controller, and $B_D = 14,400$ refers to the number of bits required to exchange the coupling interference variables in the distributed scheme.

In general, the choice of selecting the centralized scheme over the decentralized approach depends on the factor Γ defined as

$$\Gamma \triangleq \left(\frac{N_C}{N_D}\right) \times \frac{N_T\,(N_B + T)}{T\,(N_B - 1)} \tag{i}$$

[1] Note that $N_C = 10$ bits can be justified for the CSI quantization, since the precoders are designed based on the quantized CSI fed back from the BSs to the centralized controller.

where T represents the duration over which the CSI is valid for designing the precoders by the centralized controller. Note that we considered different quantization scales for the centralized and the distributed exchanges. If the fraction is $\Gamma > 1$, then the distributed scheme is preferred and if $\Gamma < 1$, then the centralized overhead is smaller and therefore, it is preferred. Note that we considered the distributed algorithms are iterated for the same count in each transmission slot. It is worth noting that since the channel is semi-static, the distributed schemes are not required to iterate for the same number of iterations in each transmission instant. Since the operating point can be chosen to be the solution obtained from the previous transmission instant, the convergence speed of the distributed algorithm can be accelerated significantly by doing so.

Note that the above discussions does not include the overhead involved in the queue transfers to the controller and the loss in performance due to the distributed approach. The loss involved in the centralized performance due to the quantized CSI is also not considered. Therefore, by using the above arguments, if the system size grows, *i.e.*, as the number of transmit antennas increase, then the distributed approach is preferred to the centralized scheme even if the channel remains constant for longer duration such that $\Gamma > 1$. The reason is that the overhead involved in feeding back the precoders from the centralized controller will be significantly large.

In practice, it is not required to perform the distributed approaches for higher number of iterations [Chapter 3.2.2, 10]. According to the results in Figure II, it is sufficient to iterate for $I_{\max} = 10, J_{\max} = 1$ iterations to achieve acceptable performance improvement from the coordinated precoder design. The reason is that the performance gained by exchanging additional coupling variables is marginal compared to the overhead involved in doing so.

- Secondly, we consider a time-correlated fading scenario, where the channel changes slowly over each transmission instant based on the user mobility or the surrounding environment. In such cases, the pilots are periodically sent from each user to measure the CSI by each BSs to design the precoders efficiently. In such scenarios, where the channel cannot be assumed constant over multiple transmission instants, the distributed schemes are much preferred to the centralized approaches. One such analysis was studied in Section C of [12] using dual decomposition based distributed precoder design, where they have showed that it is enough to follow the fading process to obtain desired performance instead iterating until convergence. Using this argument, it is beneficial to just follow the fading process by using the precoders designed in the previous transmission as the operating point for the current instant for SCA rather starting at some random feasible point.

  Studying the performance of distributed algorithms for the time correlated case is beyond the scope of our paper and thus is not considered in the current manuscript. However, we take this opportunity to show a plot demonstrating this behavior for the Karush-Kuhn-Tucker (KKT) based algorithm presented in Section IV-C of the manuscript. Note that there is no need for the backhaul exchange to design the precoders in the KKT scheme proposed in Algorithm 3. However, it requires the exchange of information via over-the-air (OTA) transmissions using precoded pilots as discussed in Section IV-C.

  Figure III compares the performance of the distributed KKT approach, *i.e.*, Algorithm 3 for different values of $I_{\max}$. The signaling requirements are outlined in Algorithm 3 and the overhead involved in the signaling is penalized in the achievable rate of the users.[2] We considered that the channel is coherent over $N_S = 100$ symbols and the precoder update is performed by exchanging the required information for $I_{\max} = 3, 5, 10$. The overhead is considered as $\tilde{t}_{l,k,n} = (1 - \frac{I_{\max}}{N_S}) \times t_{l,k,n}$, where $\tilde{t}_{l,k,n}$ is the rate seen by the user and the factor $(1 - \frac{I_{\max}}{N_S})$ is considered as a penalty involved due to the precoded pilot exchange. More details can be found in [R2] and [R3]. The average number of backlogged packets after each transmission slot is evaluated as

$$\chi = \sum_{k=1}^{K} [Q_k - \tilde{t}_k]^+ \tag{ii}$$

  and the PL is distributed uniformly between $[0, -3]$ dB.

[2] Please refer to [25] for further details.

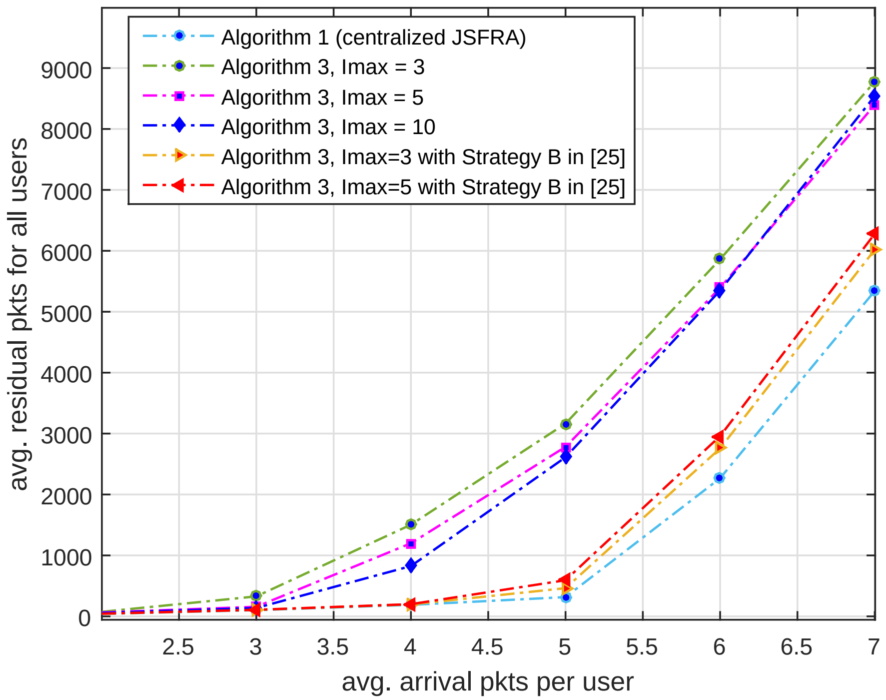


Figure III: Expected number backlogged packets after each transmission instant for a system $\{N, N_B, K, N_T, N_R\} = \{4, 2, 16, 4, 2\}$ evaluated for 500 slots and PL is distributed $\mathbb{U}(0, -3)$

Unlike Algorithm 3, the centralized precoder design presented in Figure III does not include any penalty term and it is used as a benchmark. In order to improve the performance of the distributed scheme, the operating point involved in the SCA algorithm is considered from the earlier frame instead of being generated randomly. Note that the performance of the distributed scheme is significantly improved by performing the internal iterations as discussed in [25] as strategy B. By doing so, we can achieve the performance close enough to the centralized approach (which does not included any penalty term in the rate). Therefore, for the time-correlated scenarios, it would be beneficial to do distributed approaches as discussed above to avail the channel correlation rather than feeding back the CSIs to the centralized controller.

[R1] Muhammad-Fainan Hanif, Le-Nam Tran, Antti Tölli, Markku Juntti, and Savo Glisic, "Efficient Solutions for Weighted Sum Rate Maximization in Multicellular Networks With Channel Uncertainties", *IEEE Trans. Signal Process.*, vol 61, no. 22, pp. 5659-5674, Nov. 2013.

[R2] Changxin Shi, Berry R.A., Honig M.L., "Bi-Directional Training for Adaptive Beamforming and Power Control in Interference Networks," *IEEE Transactions on Signal Processing*, vol.62, no.3, pp.607,618, Feb.1, 2014

[R3] P. Jayasinghe, A. Tölli, J. Kaleva, M. Latva-aho, "Bi-directional Signaling for Dynamic TDD with Decentralized Beamforming", in *Proceedings of IEEE ICC SmallNets Workshop*, London, UK, June, 2015

The discussions are described in a condensed form in the first paragraph of Section IV and in the first and the last paragraph of Section IV-C.

3. If the authors can't prove the convergence of the ADMM algorithm (or the decomposition approach via KKT conditions) in Section IV.B, then at least, you should discuss the property of the fixed point of the algorithm. For example, does there exist a fixed point of the algorithm? If so, is the fixed point of the algorithm unique? Is any fixed point of the algorithm also the optimal solution of the original problem in (20)? Assuming that the ADMM algorithm converges to a fixed point, will the interference vector in (39) converges to the actual interference in the network? These questions must be clarified in the paper. Otherwise, it is not clear how the ADMM algorithm is related to the original problem in (20). Similar questions should also be answered for the decomposition approach via KKT conditions.

Reply: We thank the reviewer for raising the concern regarding the convergence of the distributed algorithms. We have substantially improved the clarity of both primal and the ADMM based approach in the revised manuscript (see Appendix B). There, we have utilized the strong convexity of the objective function to ensure the uniqueness of the minimizer in each SCA step. Additionally, since each subproblem in (20) or (28) has a convex feasible set, the convergence of the ADMM scheme is guaranteed by using [10] and [36, Prop. 4.2] as discussed in Appendix B.

In our paper, the number of ADMM iterations is denoted by $J_{\max}$ in Algorithm 2. In order to obtain the centralized solution, we need to perform the distributed approaches until convergence or for some finite number, say, $J_{\max}$. In practice, it may not be possible to perform the distributed methods until convergence. In such cases, the overall convergence of the objective sequence and beamformer iterates of Algorithm 2 cannot be guaranteed. It follows from the fact that in each distributed step, the global objective need not decrease strictly to ensure the convergence of the beamformer iterates.

However, if we perform the distributed algorithms for significant number of iterations to ensure strict monotonicity of the objective, then the overall of convergence of Algorithm 2 is guaranteed by following the arguments in [27]. Note that in addition, uniqueness of the limit point is also necessary for the sequence convergence. It is achieved by regularizing the original convex objective with a quadratic term as in (46b) in each SCA step to guarantee strong convexity, and thereby ensuring unique minimizer and limit point.

Upon satisfying the strict monotonicity of the objective sequence in each SCA update with some $J_{\max}$, we can also combine the minimum mean squared error (MMSE) receiver update in (23b) after each SCA step for the transmit precoders. It reduces the signaling overhead involved with the information exchange due to multiple loops that are required for alternating optimization (AO) and SCA as mentioned for the centralized scheme. Even though the receivers are updated together with the transmit precoders in each SCA step, strict monotonicity of the objective sequence is still ensured. It follows from the fact that the MMSE receivers are optimal for fixed transmit precoders obtained after each SCA update. Moreover, the newly found solution is also feasible and the objective improves monotonically in strict sense. Therefore, if the above conditions are satisfied, then the convergence of the sequence of iterates is guaranteed.

For our simulation scenario, we have fixed the maximum number of ADMM iterations in each SCA step as $J_{\max} = 20$. Note that it is sufficient to ensure strict monotonic decrease of the objective value in each SCA step. However, the number of iterations to ensure strict monotonicity of the objective depends on the problem under consideration. The practical significance of limited number of ADMM iterations is discussed briefly in [10, Section 3.2.2].

Regarding the specific comments related to the ADMM convergence for subproblem (20), we provide the following remarks by considering that the objective is regularized by a quadratic term as discussed in Appendix A-B to ensure strongly convexity.

(a) The primal objective and the augmented Lagrangian are strongly convex, therefore, the optimal solution exists and is unique in the feasible convex set of subproblem (20). It follows from that the ADMM algorithm converges to a unique limit point, which is a fixed point of the iterative ADMM algorithm.

(b) In general, the fixed point of the subproblem (20) is not unique. It is due to the fact that the original objective is not strongly convex, thereby having possibly multiple solutions in the feasible convex set. It is due to the existence of multiple solutions, the ADMM algorithm can have a set of fixed points in the feasible set. However, if we regularize the objective with a strongly convex quadratic term as discussed in Appendix A-B, then the distributed approaches find a unique fixed point of the original convex subproblem.

(c) By using previous argument, any fixed point of the algorithm is a solution for the problem (20), since all fixed points are in fact the solution for the convex subproblem in (20). The particular choice of the fixed point of the ADMM approach depends on the initial operating point used in the regularization term in the objective.

(d) Upon convergence of the ADMM approach, the interference vector is guaranteed to be equal to the actual interference seen in the network. This is the main principle of ADMM where the local variables (the vectors in (38)) converge to the global variables, which are the actual inference vectors.

However, if the reviewer comments are for the Algorithm 2, then the following response will answer the questions. For that, we consider the following conditions are satisfied in each SCA update step.

- The objective function is regularized by a quadratic term as discussed in Appendix A-B to ensure strict convexity of the objective function, which leads to the uniqueness of the solution.
- The distributed algorithms are performed until convergence or for sufficient number of iterations to ensure strict monotonic decrease of the objective sequence.

Upon satisfying the above conditions, even if we update the receive beamformers using the MMSE receiver in (23b) after each SCA step for solving the transmit precoders, strict monotonicity is still ensured. It follows from the fact that the MMSE receivers are optimal for the fixed transmit precoders, and therefore the objective sequence decreases monotonically after each update in strict sense.

(a) Since the above conditions are satisfied by Algorithm 2, then the existence of a fixed point for Algorithm 2 is guaranteed by following the arguments in [27].

(b) In general, fixed point of Algorithm 2 is not unique. Note that the iterative nature of the SCA approach is based on approximating the nonconvex constraint by a convex one around some fixed operating point. While initializing the iterative Algorithm 2, we need to find a fixed operating point upon which the convex function is approximated. Therefore, depending on the choice of initial feasible point, the algorithm terminates at different fixed point upon convergence. Hence, there exist a set of fixed points for Algorithm 2.

(c) The nonconvex problem (16) can have multiple stationary points and the iterative method proposed in Algorithm 2 finds a stationary point of the nonconvex problem. Therefore, all fixed points found by Algorithm 2 upon convergence are indeed the stationary points of the original nonconvex problem.

(d) Upon convergence of the ADMM approach, the interference vector is guaranteed to be equal to the actual interference seen in the network. This is the main principle of ADMM where the local variables (the vectors in (38)) converge to the global variables, which are the actual inference vectors.

On the other hand, the iterative method described in Algorithm 3 is based on a heuristic iterative approach to find a solution of KKT equations, we cannot prove its convergence. When Algorithm 3 terminates, we will use the transmit and receive beamforming vectors to calculate the original objective. Based on the reviewer comments, we have mentioned that the convergence of the Algorithm 3 cannot be proved theoretically (see Appendix B last paragraph).

We have updated the manuscript to clarify all the points mentioned above. Please refer to Section IV-B last paragraph after (40) in the revised manuscript. Additional information regarding the convergence of the distributed algorithms is provided in Appendix B second paragraph. The discussion on the convergence of the KKT based approach for MSE reformulation scheme is also presented in Appendix B last paragraph. For reference purpose, we have also referred the interested reader to [10], which discusses exclusively about the ADMM approach.

Response to Reviewer - 3's Comments

The authors have introduced changes in the manuscript that improved the paper's quality. Additionally, the authors have taken into account the reviewers' comments giving clarifications and modifying the content when required. More specifically, the following aspects have been treated:

Reply: We thank the reviewer for recognizing the changes made on our earlier manuscript. We also thank the reviewer for providing constructive comments to help us to improve our manuscript.

1. Convexity of problem (16). The paragraphs surrounding (16) allow a better understanding of the usage of the additional variables, i.e. gamma and beta, to remove the equality constraint in (2). For the reviewer remains however unclear, the procedure/criterion to determine the operating point for the parameter $\tilde{\beta}$ required in (19) and used in the convex subproblems (20) and (21).

Reply: We thank the reviewer for the comment. We have included additional material on how to choose the operating point while initializing the iterative algorithm. It is described briefly in the fourth paragraph following (23). We have also provided a detailed discussion on how to update the operating point after each iteration in the paragraphs following (23). Since the objective function is convex, we can update the current operating point with the solution obtained from the previous iteration as

$$\tilde{\mathbf{m}}_j^{(i)} = \mathbf{m}_j^{(i)} \tag{iii}$$

where $\mathbf{m} = \{\mathbf{m}_{l,k,n}\}_{\forall l, \forall k \in \mathcal{U}, \forall n}$ corresponds to the collection of all transmit precoders in the system. Note that $\tilde{\mathbf{m}}_j^{(i)}$ is the operating point for the $(j+1)$th SCA iteration in the ith AO step and $\mathbf{m}_j^{(i)}$ is the solution obtained by solving (20) in the jth SCA iteration.

2. Proof of convergence. The proof of convergence introduced by the authors in the Appendix seems correct and enhances the content of the manuscript.

Reply: We thank the reviewer for the comment. We have included several materials to make the proof of convergence more rigorous.

Additional Comments -

(a) - The reviewer considers that closing statements regarding the applicability of the proposed schemes are missing. Since the results are quite similar (when not identical), which formulation is preferable between the centralized schemes? Which one for the distributed solutions?

Reply: We have updated the manuscript to discuss possible options for selecting an algorithm for a specific implementation. The choice of considering a centralized approach is equally good when the number of receive antennas is larger than one, *i.e.*, $N_R > 1$. However, in case of a single antenna receiver, the joint space-frequency resource allocation (JSFRA) formulation outlined in Section III-B is more efficient than the mean squared error (MSE) reformulation, as there is no need for the receiver update, thereby reducing the complexity significantly.

As far as the distributed approaches are concerned, the primal and the ADMM schemes are equally favorable when $N_R = 1$ since the signaling required to be exchanged between the coordinating BSs involves only the scalar interference values. However, when $N_R > 1$, then the KKT based distributed approach is more efficient than the primal and the ADMM schemes, since it has less signaling overhead. The choice of selecting ℓ_q norm is discussed in Section II-B last paragraph. The above information is included in Section IV last paragraph, *i.e.*, before the simulations section.

(b) - The last discussion in section IV-C could benefit from restructuring. The information on how to obtain a practical distributed precoder design and to avoid backhaul exchange is too condensed and difficult to understand.

Reply: It has been rewritten to improve the readability as suggested by the reviewer. For additional information, the reviewer is requested to refer [R1].

[R1] Changxin Shi, Berry R.A., Honig M.L., "Bi-Directional Training for Adaptive Beamforming and Power Control in Interference Networks," *IEEE Transactions on Signal Processing*, vol.62, no.3, pp.607,618, Feb.1, 2014

(c) - For the simulation results, why not to unify configurations when possible? Having to read a different configuration for each graph is cumbersome and no additional comparisons are possible between figures. E.g. PL uniformly distributed between [0,-6] dB in Fig. 2 and [0,-3] dB in Fig. 3.

Reply: The main purpose of considering several path loss models is simply to show that the proposed algorithms works reasonably well on various scenarios mentioned in Section V (Simulation Section). For benchmarking, we have also included the centralized algorithm to draw the difference between the other schemes. We are conducting more numerical experiments to have a more complete comparison of all the schemes. Unfortunately those results cannot be included in the current manuscript due to space limitation. We aim at making them available online as a companion technical report when they are ready.

(d) - In Fig. 1, the description of the system model does not agree with the statement of N = 3 subchannels.

Reply: The error has been fixed in the revised manuscript.

(e) - In Fig. 4(b), the performance of Q-WSRME seems to be (in average) worse than that of Q-WSRM. However, that should not be the case, since Q-WSRME is taking into account the over allocation. Any reason for this?

Reply: Indeed the performance of the Q-WSRME is better than Q-WSRM in Fig. 4(b). The misleading observation was probably due to the visualization effect where the colors representing the performance of sum arrivals and Q-WSRME are quite similar. We have updated Fig. 4 (and all other figures) in the revised manuscript to avoid such visualization problems.

(f) - p 6, col 2, row 49: it should be $t_{l,k,n}$ instead of $t_{l,n,k}$

Reply: This issue is corrected in the revised manuscript.

(g) - p 10, col 1, row 28: is it λ a dual variable?

Reply: In fact, it should be the dual variable $\sigma_{l,k,n}$ and not $\lambda_{l,k,n}$. We have fixed this problem in the revised manuscript.

(h) - p 10, col 2, row 59: typo wHith

Reply: We have fixed this error.

(i) - In general, a grammar check is recommended, several mistakes with respect to singular and plural nouns have been observed, e.g. -p 6, col 2, row 56 "... for A fixed receiverS" is not correct.

Reply: We thank the reviewer for the comment and we have proofread the manuscript carefully to eliminate all grammar errors.

Response to Reviewer - 4's Comments

This reviewer's concerns have been addressed, and this manuscript is now deemed fit for publication.

Reply: We thank the reviewer for the constructive comments and recommending the revised manuscript for the publication.