

Totally Distributed Energy-Efficient Transmission in MIMO Interference Channels

Cunhua Pan, Wei Xu, *Senior Member, IEEE*, Jiangzhou Wang, *Senior Member, IEEE*, Hong Ren, Wence Zhang, Nuo Huang, and Ming Chen

Abstract

In this paper, we consider the problem of maximizing the energy efficiency (EE) for multi-input multi-output (MIMO) interference channels, subject to the per-link power constraint. To avoid extensive information exchange among all links, the optimization problem is formulated as a noncooperative game, where each link maximizes its own EE. We show that this game always admits a Nash equilibrium (NE) and the sufficient condition for the uniqueness of the NE is derived for the case of large enough maximum transmit power constraint. To reach the NE of this game, we develop a totally distributed EE algorithm, in which each link updates its own transmit covariance matrix in a completely distributed and asynchronous way: Some players may update their solutions more frequently than others or even use the outdated interference information. The sufficient conditions that guarantee the global convergence of the proposed algorithm to the NE of the game have been given as well. We also study the impact of the circuit power consumption on the sum-EE performance of the proposed algorithm in the case when the links are separated sufficiently far away. Moreover, the tradeoff between the sum-EE and the sum-spectral efficiency (SE) is investigated with the proposed algorithm under two special cases: 1) low transmit power constraint regime; 2) high transmit power constraint regime. Finally, extensive simulations are conducted to evaluate the impact of various system parameters on the system performance.

Index Terms

This work was supported by National 863 High Technology Development Project (No. 2014AA01A701), National Nature Science Foundation of China (Nos. 61172077 & 61372106 & 61223001 & 61471114), the Scientific Research Foundation of Graduate School of Southeast University (No. YBJJ1433). C. Pan, W. Xu, H. Ren, W. Zhang, N. Huang and M. Chen are with National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China. (Email: {cunhuapan, wxu, renhong, wencezhang, huangnuo, chenming}@seu.edu.cn). J. Wang is with the School of Engineering and Digital Arts, University of Kent, Canterbury, Kent, CT2 7NZ, U.K. (Email: j.z.wang@kent.ac.uk). Part of this work was presented in IEEE Globecom 2014.

Totally distributed algorithm, MIMO interference channels, energy efficient transmission.

I. INTRODUCTION

Past few years have witnessed tremendous advancement in wireless communications, including the significant improvement of transmission rate [1], [2]. However, the impact of power consumption on the environment is neglected. It is reported that the total energy consumption of the communications takes up more than 3 percent of the worldwide electric energy consumption [3] and the portion is expected to increase due to the explosive growth of high-data-rate applications in the future. Hence, energy efficiency (EE) has gained lots of attention and will be one of key issues in future fifth-generation (5G) mobile networks [4]. On the other hand, the interference channel (IC) has been modeled mathematically for many practical systems where multiple uncoordinated links share the same channel, such as femtocells, ad hoc wireless networks, cognitive radio, etc [4], [5]. Furthermore, due to the development of advanced multi-antenna techniques [6], [7], each transmission node is able to accommodate multiple antennas [8], [9]. It is well known that multi-input multi-output (MIMO) system has the great potential for providing high SE by employing spatial multiplexing techniques [10]. Hence, it is of great importance to study the energy efficient transmission strategy in MIMO ICs.

This paper focuses on the EE maximization problem for MIMO ICs with per-link transmit power constraint. To solve this problem, one may consider centralized solutions, which require a central processing unit (CPU) to collect all complex-valued channel matrices over the network. The CPU will compute all links' transmit covariance matrices and send them to the corresponding links. Hence, for large-scale networks, the centralized approaches suffer from heavy feedback overhead and high computational complexity, which hinders practical implementations. Moreover, there may not exist a CPU for some wireless networks, such as ad hoc or wireless sensor networks.

Recently, distributed algorithms to deal with this problem attract intensive attentions [11]–[14]. Here, “distributed” means that precoders can be computed at the transmitters with only local channel knowledge and limited (or no) information exchange over different links. Generally, distributed processing for MIMO systems has the benefits of low communication exchange overhead, low computational complexity, more scalability, low system costs, etc. The classical distributed algorithm based on dual decomposition technique is designed to decompose the

coupling constraints among links [11], [12]. Specifically, by introducing the dual variables associated with the coupling constraints, the original problem can be divided into several independent subproblems, each of which can be solved in a distributed way. Then, all the links exchange some necessary information to update the dual variables. For more practical networks with individual link power budget constraints, [15] devised a decentralized beamforming EE (DBFEE) algorithm for symmetric MIMO ICs, where the distance from one transmitter to its desired receiver is identical for all links, and each transmitter has the same distance to all its unintended receivers. In each iteration, all the receivers should feed back the equivalent channel matrices to all the transmitters in the network. In [16], the authors designed a two-layer EE (TLEE) algorithm based on the generalized weighted minimum mean square error (WMMSE) approach [14]: The inner layer to update precoders/decoders; the outer layer to update some parameters. Similar to the algorithm in [15], in each inner iteration each receiver should feed back the updated weight matrix and the positive definite covariance matrix to all the transmitters in the network. However, for these distributed algorithms, in each iteration each receiver needs to compute the necessary complex-valued matrix and then feed it back to all the transmitters in the network, which could induce serious implementation challenges such as a large amount of feedback overhead, poor scalability and heavy computational burden at the receivers. Moreover, all the links should be synchronous, which is difficult to be satisfied, especially for large-scale ad hoc networks or wireless sensor networks. One novel distributed algorithm based on the adaptive price was proposed in [17] to deal with the weighted sum EE maximization problem for single-input single-output (SISO) ICs.

Hence, one distributed algorithm with much lower feedback overhead is more desirable. Noncooperative game theoretical approach has been recognized as a powerful tool to devise totally distributed algorithms, in which each link just maximizes its utility without the need of information exchanges among the links. A number of researches have applied game theory to design energy efficient communications for ICs [18]–[21] or multiple access channel (MAC) [22]. The EE optimization problem in flat fading single-input single-output (SISO) ICs was considered in [19], where one distributed algorithm based on non-cooperative game was proposed. Both the existence and uniqueness of Nash equilibrium (NE) were analyzed. This work was extended to frequency-selective channel in [20] and to a relay channel in [21]. Although [20] proved the existence of the NE, regarding the uniqueness of the NE, [20] only showed that the number of

NEs is determined by the cross-channel gains and the direct channel gains, without quantifying how they are related with each other. All these studies considered single-antenna ICs and they apply for multi-antenna case only if the transmit powers are optimized with fixed transmit directions as shown in [23]. In [22], the authors considered the EE maximization problem for the MIMO uplink systems with each user transmitting only one stream. This problem is formulated as a non-cooperative game, where the uniqueness of NE is guaranteed by the fact that the EE function is S-shaped.

In this paper, we apply the non-cooperative game theoretical approach to deal with the EE maximization problem for MIMO ICs, where each link attempts to maximize its own EE by jointly optimizing transmit power and beamformers. It is a nontrivial extension of the SISO case in [20]. There is an explicit relationship between the power allocation among different links and the achievable rates in the SISO case. This property is critical in deriving the conditions of the existence and uniqueness of the NE by using the standard function [20]. However, in MIMO systems, this relationship is implicit, as power allocation is carried out through matrix manipulations. Moreover especially when multiplexing is utilized with the MIMO, different from the beamforming case in [22] for MAC, the EE design is shown more general for the MIMO IC and hence more difficult since there exist both inter-node (mutual) interference and intra-node interference in the MIMO IC using multiplexing.

A. Related Work

Recently, distributed algorithms for MIMO ICs have been extensively studied in the literature, such as linear iterative approximation (LIA) algorithm [13], the WMMSE algorithm [14], noncooperative game theoretic algorithm [24], etc. For the LIA algorithm, it is designed based on the first-order Taylor expansion of the non-convex part of the weighted sum spectrum efficiency (SE) objective function. Sequential convex optimization approaches were then presented for dealing with various scenarios, e.g., the multi-band scenario in [25], the MAC in [26], the broadcast channel in [27], the cognitive radio networks in [12]. However, in the LIA algorithm only one user is allowed to update its covariance matrix at one time, which may lead to significant latency especially in dense networks. By establishing the equivalence between the weighted sum SE problem and weighted sum mean square error minimization problem, [14] proposed the WMMSE algorithm that allows multiple users to update simultaneously. In this algorithm, the local optimal

solution is obtained via alternatively optimizing the linear transceivers and iteratively updating the weight matrices. The authors show that when the utility function satisfies some conditions, the algorithm is guaranteed to converge to the stationary point of the original problem. The WMMSE algorithm has been applied in various setups, please see [28] and references therein.

However, to successfully implement the ILA and WMMSE algorithms in a distributed manner, two assumptions are required: 1) Perfect channel reciprocal between the uplink phase and the downlink phase (in time-division duplexing mode); 2) Synchronization between all the links. In practice, the communication systems usually operate in frequency-division duplexing (FDD) mode. The channel reciprocal is thus hard to achieve. In addition, for wireless sensor or ad hoc networks, asynchronous among the links is more desirable. More importantly, in each iteration each receiver needs to calculate the pricing matrix in the ILA algorithm or the weight matrix in the WMMSE algorithm, and then feeds them back to all the transmitters in the network, which poses serious implementation issues such as a large amount of heavy feedback overhead, poor scalability and heavy computational burden at the receivers.

On the other hand, non-cooperative game theoretical approaches have attracted extensive attentions. Studies in this direction are plentiful in literature, e.g., [29] for the MIMO MAC, [24], [30]–[36] for the MIMO ICs. In [29], the authors formulated the problem as a non-cooperative game and the authors proved that each user's optimal eigenvectors do not depend on the channels of others. Based on this fact, the authors showed that the existence and uniqueness of NE is guaranteed when the numbers of transmit and receive antennas become large. The main technique in [29] is random matrix theory. For the MIMO ICs, the authors in [24] first utilized the noncooperative game framework to deal with the SE maximization problem, where the iterative water-filling algorithm was proposed to find the NE of the game. However, the existence of the NE was only shown by the simulation results without theoretic guarantees, neither the NE uniqueness. Then, in [30], Scutari *et al* proved the existence of NE and provided sufficient conditions for the uniqueness of the NE, which can be checked in practice. However, the results are only valid for square nonsingular channel matrices. Later on, they generalized the results to a more general case with arbitrary channel dimensions in [31]. The cognitive radio network with null shaping constraints on the primary user is considered in [32] and its robust version in [33]. Most recently, these works were extended to the multicell case in [34], [35] with multiple users per cell. In [36], the authors formulated the SE maximization problem as a

cooperative game. Specifically, by fixing the outgoing cooperative set and incoming cooperative set, the authors formulated this problem as a non-cooperative game, where the existence and uniqueness of the NE was analyzed. Then, the coalitional game theory [37] was applied to obtain the stable of the cooperative set. The work that is most closely related to ours is [31], where the noncooperative game was formulated for the SE maximization problem of the MIMO ICs and one asynchronous distributed algorithm was proposed to reach the NE of the game.

In contrast to the most of the above cited papers which focus on the (weighted) sum SE problem, in this work we consider the EE maximization problem. For SE optimization problems, it is known that all the transmitters use full power during transmission in order to maximize its own SE. Based on this fact, the best response strategy at the NE can be written in a closed-form water-filling solution, which can be interpreted as a projector on the convex and closed set. This interpretation enables the authors to derive the uniqueness of the game's NE [31]. However, the study of EE maximization problem cannot be obtained by employing the methodologies since the transmitters in fact use a portion of the power, instead of full, to achieve energy efficient transmission.

B. Contributions

In this paper, we apply the non-cooperative game theoretical approach to deal with the EE maximization problem for MIMO ICs, where each link attempts to maximize its own EE by jointly optimizing transmit power and beamformers.

The main contributions and observations of our work are summarized as follows.

- 1) The EE maximization problem in MIMO interference channels is modeled as a noncooperative game where each MIMO link competes against the others by choosing its transmit covariance matrix to maximize its own EE. We show that the NE of this game always exists and derive sufficient conditions for the uniqueness of the NE for the case of large enough maximum transmit power constraint.
- 2) To reach the NE of the game, we provide a totally distributed EE algorithm named Asynchronous Distributed Energy-Efficient (ADEE) algorithm, which is the extended version of simultaneous updating proposed in [20]. In this algorithm, all users apply the fractional programming to update the transmit covariance matrices and these updates can be performed in a totally asynchronous way, which means some links may update their

transmit covariance matrices more frequently than the others and they may even use the outdated information of the measurement of the interference generated by the other links. In addition, during the updating procedure of the algorithm, there is no need for the links to exchange the signaling overhead mutually. These features make our distributed algorithm more appealing for practical implementations. We provide the sufficient conditions for the global convergence of this algorithm to the unique NE of the game. Interestingly, we find that these conditions coincide with the conditions for the uniqueness of NE.

- 3) We study the impact of the circuit power consumption on the overall SE and EE performance of the system for one special case when the links are separated sufficiently far away. We show that the overall SE increases with the circuit power consumption, but the overall EE decreases with it. Although this trend is derived for this special case, from simulations we find the trend holds for the general case when the interference among the links is sizeable. This observation implies that when the circuit power consumption increases, we should enhance the transmit rate or SE in order to obtain the best EE performance.
- 4) The tradeoff between SE and EE is investigated for the proposed algorithm (denoted as ADEE algorithm) and the SE maximization algorithm (denoted as ADSE algorithm) in [31]. Two special cases are studied: the transmit power constraint approaches zero or infinity. For the case of low transmit power constraint, we show that both algorithms use all power to transmit and thus achieve the same performance in terms of the overall SE and EE performance. However, for the latter case, the ADSE algorithm always uses all available power to transmit, yielding severe interference over the network. In this case, the SE achieved by the ADSE algorithm will not increase. Then, the EE achieved by the ADSE algorithm will approach zero due to the significant power consumption. On the other hand, for the EE metric, the ADEE algorithm is unwilling to consume all power in this case. As a result, the SE and EE achieved by the ADEE algorithm will become constant in the case of the high transmit power constraint.

The rest of the paper is organized as follows. In Section II, we introduce the system model and formulate the optimization problem as a strategic noncooperative game. Then, we show that this game always admits a NE and derive sufficient conditions for the uniqueness of the NE in Section III. To reach the NE, a totally asynchronous and distributed algorithm is given

in Section IV. In Section V, we study the impact of the circuit power consumption on the system performance in terms of the sum-SE and sum-EE, along with the study for the tradeoff between the sum-EE and sum-SE for the proposed algorithm. Section VI provides representative numerical results to study the effects of different system parameters on the proposed algorithm. Finally, some conclusions are drawn in Section VII.

Notations: $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $\text{vec}(\cdot)$, $\text{E}\{\cdot\}$, $\text{tr}(\cdot)$ and \otimes are conjugate, transpose, Hermitian transpose, stacking vectorization operator, expectation operator, trace operators and the Kronecker product operator [38], respectively. Uppercase and lowercase boldface denote matrices and vectors, respectively. For matrix \mathbf{A} , $[\mathbf{A}]_{:,k}$ and $[\mathbf{A}]_{i,j}$ represent the k^{th} column of \mathbf{A} and the (i,j) element of matrix \mathbf{A} , respectively. $\|\mathbf{A}\|_2$ denotes the spectral norm of \mathbf{A} [38]. $\|\mathbf{A}\|_F$ denotes the Frobenius norm of \mathbf{A} . $\lambda_{\min}(\mathbf{A})$ stands for the minimum eigenvalue of \mathbf{A} . $\mathbf{A} \succeq \mathbf{B}$ means $\mathbf{A} - \mathbf{B}$ is positive semidefinite. The spectral radius of \mathbf{A} is denoted by $\rho(\mathbf{A})$ [38]. $\text{rank}(\mathbf{A})$ denotes the rank of \mathbf{A} . For vector $\mathbf{a} \in \mathbb{C}^{n \times 1}$, $\|\mathbf{a}\|_2$ represents the Euclidean norm defined as $\|\mathbf{a}\|_2 = \sqrt{\mathbf{a}^H \mathbf{a}}$. The sets of $m \times n$ complex matrices, $n \times n$ complex positive semidefinite and definite matrices are denoted by $\mathbb{C}^{m \times n}$, $\mathbb{S}_+^{n \times n}$ and $\mathbb{S}_{++}^{n \times n}$, respectively. $\mathbf{D}_{\mathbf{X}} \mathbf{Y}$ denotes the Jacobian matrix of function \mathbf{Y} with respect to (w.r.t.) \mathbf{X} [39]. \mathbf{I} and $\mathbf{0}$ represent the identity and zero matrices with appropriate dimensions, respectively. $[x]^+$ is equivalent to $\max\{0, x\}$.

II. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a K -link MIMO interference channel with K transmitter-receiver pairs. Each link consists of one transmitter with M transmit antennas and one receiver with N receiver antennas. All links are simultaneously communicating over the same channel. At receiver k ¹, the received complex baseband signal vector $\mathbf{y}_k \in \mathbb{C}^{N \times 1}$ is given by

$$\mathbf{y}_k = \mathbf{H}_{k,k} \mathbf{x}_k + \sum_{j \neq k} \mathbf{H}_{jk} \mathbf{x}_j + \mathbf{n}_k, \quad (1)$$

where $\mathbf{x}_k \in \mathbb{C}^{M \times 1}$ denotes the transmit signal vector of link k , $\mathbf{H}_{kk} \in \mathbb{C}^{N \times M}$ is the direct channel matrix of link k , $\mathbf{H}_{jk} \in \mathbb{C}^{N \times M}$ is the cross-channel matrix from transmitter j to receiver k , and $\mathbf{n}_k \in \mathbb{C}^{N \times 1}$ is circularly symmetric, zero-mean, complex Gaussian noise with normalized identity

¹“Receiver k ” represents the receiver of the k^{th} link. In the following, “transmitter k ” means the transmitter of the k^{th} link.

covariance matrix. For each link, the total average transmit power should satisfy the per-link power constraint:

$$P_k = \text{tr}(\mathbf{Q}_k) \leq P_T, \quad (2)$$

where $\mathbf{Q}_k = E \{ \mathbf{x}_k \mathbf{x}_k^H \}$ is the covariance matrix of \mathbf{x}_k and P_T is the maximum transmit power.

To reduce the complexity of decoding at the receivers, it is assumed that joint decoding of the interfering signals is not an option and the interference is treated as noise at the receivers. Thus, the SE of link k is given by (in bit/s/Hz)

$$C_k(\mathbf{Q}_k, \mathbf{Q}_{-k}) = \log_2 |\mathbf{I} + \mathbf{H}_{k,k}^H \mathbf{R}_k^{-1} \mathbf{H}_{k,k} \mathbf{Q}_k|, \quad (3)$$

where $\mathbf{R}_k \triangleq \mathbf{I} + \sum_{j \neq k} \mathbf{H}_{j,k} \mathbf{Q}_j \mathbf{H}_{j,k}^H$ represents the interference-plus-noise (IPN) covariance matrix at receiver k and $\mathbf{Q}_{-k} \triangleq [\mathbf{Q}_1, \dots, \mathbf{Q}_{k-1}, \mathbf{Q}_{k+1}, \dots, \mathbf{Q}_K]$ denotes the set of all links' covariance matrices, except that of link k . In this work, it is assumed that each receiver k can perfectly measure the IPN covariance matrix \mathbf{R}_k and estimate the direct channel $\mathbf{H}_{k,k}$, and then report them back to transmitter k . The channels are assumed to vary sufficiently slowly such that it can be considered fixed during the transmission.

In order to design energy efficient transmissions, the total power consumption should be considered at each link k : power used for reliable data transmission P_k , circuit power during transmission P_C , which is the power consumed by the mixers, filters and digital-to-analog converters, digital signal processing (DSP), etc. The power consumption of the DSP depends on the number of computations of the algorithm and the signaling overhead. It is difficult to accurately model this kind of power consumption. Thus, the circuit power P_C is modeled as a constant in this work for the sake of analysis, as simplified in most of the existing works [15], [19]–[21], [23]. Even though, we will evaluate the system performance under different P_C via simulations. Then, the EE (in bits/Hz/Joule) of link k , defined as the ratio of SE to the total power consumption, is given by

$$\text{EE}_k(\mathbf{Q}_k, \mathbf{Q}_{-k}) = \frac{C_k(\mathbf{Q}_k, \mathbf{Q}_{-k})}{P_k + P_C}. \quad (4)$$

Since our goal is to devise a totally distributed algorithm that requires neither a CPU nor information exchange among the links, we formulate the optimization problem as the following

noncooperative game:

$$(\mathcal{G}) : \begin{aligned} & \max_{\mathbf{Q}_k} \text{EE}_k(\mathbf{Q}_k, \mathbf{Q}_{-k}), \\ & \text{s.t. } \mathbf{Q}_k \in \mathcal{W}_k \end{aligned} \quad \forall k \in \Psi_K, \quad (5)$$

where $\text{EE}_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$ is the payoff function of link k defined in (4), $\Psi_K \triangleq \{1, 2, \dots, K\}$ is the set of all links and \mathcal{W}_k is the set of admissible strategies of link k , defined as

$$\mathcal{W}_k = \{\mathbf{Q}_k \in \mathbb{C}^{M \times M} \mid \mathbf{Q}_k \in \mathbb{S}_+^{M \times M}, \text{tr}(\mathbf{Q}_k) \leq P_T\}. \quad (6)$$

In this game, each user competes against the others by choosing his own covariance matrix that maximizes his own payoff function subject to the strategy set. A solution of the game to reach a NE is when each link, given the strategic profiles of the others, does not get any increase in its objective by unilaterally changing its own strategy and is formally defined as follows.

Definition 1: A strategic profile $\mathbf{Q}^* = (\mathbf{Q}_k^*)_{k \in \Psi_K} \in \mathcal{W}_1 \times \dots \times \mathcal{W}_K$ is a NE of game \mathcal{G} if

$$\text{EE}_k(\mathbf{Q}_k^*, \mathbf{Q}_{-k}^*) \geq \text{EE}_k(\mathbf{Q}_k, \mathbf{Q}_{-k}^*), \forall \mathbf{Q}_k \in \mathcal{W}_k, \forall k \in \Psi_K. \quad (7)$$

In the forthcoming sections, we first show that game \mathcal{G} always admits at least one NE. In general, game \mathcal{G} may admit multiple NEs, depending on the level of the interference from the other links [20]. Then, we study the uniqueness condition of the NE and provide a totally distributed algorithm to reach such a NE.

III. EXISTENCE AND UNIQUENESS OF THE NE

A. Existence of NE

Whether NE exists depends largely on the properties of the payoff function. In the sequel, we first study the property of the EE function and then check the existence of NE.

Lemma 1: Given the other links' strategy \mathbf{Q}_{-k} , the EE function of link k , i.e., $\text{EE}_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$, is quasiconcave in \mathbf{Q}_k . Furthermore, if the channel matrix $\mathbf{H}_{k,k}$ is full column rank, $\text{EE}_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$ is strictly quasiconcave in \mathbf{Q}_k

Proof: Please see Appendix A. □

Based on Lemma 1, the existence of NE is given in the following theorem.

Theorem 1: Game \mathcal{G} always admits at least one NE for any set of channels and transmit power constraints.

Proof: Please see Appendix B. □

B. Uniqueness of NE

In this subsection, we first study the uniqueness condition for the case of full column-rank channel matrices. Then, we extend the results to the more general case without making any restrictive assumptions on the channel structure.

1) Case of full column-rank channel matrices

For the SE maximization game in [31], all the transmitters use full power during transmission in order to maximize its own SE. Based on this fact, the best response strategy at the NE can be written in a closed-form water-filling solution, which can be interpreted as a projector on the convex and closed set. This interpretation enables the authors to derive the uniqueness condition of the game's NE. However, the study of our EE maximization problem cannot be obtained by employing the methodologies developed in [31] since the transmitters do not always, and in fact not in most cases, use the full power to transmit in our EE case. Here, we consider the case when the maximum transmit power is large enough so that each transmitter only uses portion of the maximum power at the NE.

Before providing the uniqueness condition for the EE maximization game, we first introduce some useful intermediate results. Given the other links' strategy \mathbf{Q}_{-k} , the best response of the transmit covariance matrix of link k is denoted as

$$\mathbf{F}_k(\mathbf{Q}_{-k}) = \arg \max_{\mathbf{Q}_k \in \mathcal{W}_k} \text{EE}_k(\mathbf{Q}_k, \mathbf{Q}_{-k}) \quad (8)$$

where function $\mathbf{F}_k(\mathbf{Q}_{-k}): \mathcal{W}_{-k} \mapsto \mathbb{C}^{M \times M}$ is a complex matrix-valued function with $\mathcal{W}_{-k} \triangleq \mathcal{W}_1 \times \cdots \times \mathcal{W}_{k-1} \times \mathcal{W}_{k+1} \times \cdots \times \mathcal{W}_K$. We introduce the following mapping function,

$$\mathbf{F}(\mathbf{Q}) = \mathbf{F}_1(\mathbf{Q}_{-1}) \times \cdots \times \mathbf{F}_K(\mathbf{Q}_{-K}) : \mathcal{W} \mapsto \mathcal{W} \quad (9)$$

where $\mathcal{W} \triangleq \mathcal{W}_1 \times \cdots \times \mathcal{W}_K$ and $\mathbf{Q} \triangleq [\mathbf{Q}_1, \cdots, \mathbf{Q}_K] \in \mathcal{W}$. Using (9) and Definition 1, the NE of the game can now be characterized by the following fixed-point equation:

$$\mathbf{F}(\mathbf{Q}^*) = \mathbf{Q}^*. \quad (10)$$

According to Proposition 1.1 of [40], the contraction mappings have a unique fixed point. It is formally stated as the following lemma.

Lemma 2: Game \mathcal{G} has a unique NE if the mapping $\mathbf{F}(\cdot)$ defined in (9) satisfies:

$$\|\mathbf{F}(\mathbf{Q}^{(1)}) - \mathbf{F}(\mathbf{Q}^{(2)})\|_F < \|\mathbf{Q}^{(1)} - \mathbf{Q}^{(2)}\|_F, \quad (11)$$

for any two different $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$ in the closed set \mathcal{W} . \square

However, directly using this lemma is difficult. Instead, we give a sufficient condition for (12) to be satisfied, which plays a key role in the study of uniqueness of the NE.

Lemma 3: The sufficient condition for (12) to hold is that, for any link k , the mapping defined in (8) satisfies

$$\left\| \mathbf{F}_k(\mathbf{Q}_{-k}^{(1)}) - \mathbf{F}_k(\mathbf{Q}_{-k}^{(2)}) \right\|_F < \sqrt{\frac{1}{K-1}} \left\| \mathbf{Q}_{-k}^{(1)} - \mathbf{Q}_{-k}^{(2)} \right\|_F, \forall \mathbf{Q}_{-k}^{(1)}, \mathbf{Q}_{-k}^{(2)} \in \mathcal{W}_{-k}. \quad (12)$$

Proof: Our main task is to show that, under condition (12), the condition of the mapping $\mathbf{F}(\cdot)$ to be a contraction in (12) is satisfied. We have

$$\left\| \mathbf{F}(\mathbf{Q}^{(1)}) - \mathbf{F}(\mathbf{Q}^{(2)}) \right\|_F^2 = \sum_{k=1}^K \left\| \mathbf{F}_k(\mathbf{Q}_{-k}^{(1)}) - \mathbf{F}_k(\mathbf{Q}_{-k}^{(2)}) \right\|_F^2 \quad (13)$$

$$< \frac{1}{K-1} \sum_{k=1}^K \left\| \mathbf{Q}_{-k}^{(1)} - \mathbf{Q}_{-k}^{(2)} \right\|_F^2 \quad (14)$$

$$= \left\| \mathbf{Q}^{(1)} - \mathbf{Q}^{(2)} \right\|_F^2 \quad (15)$$

where (14) follows from (12), and (15) follows by the definitions of \mathbf{Q}_{-k} and \mathbf{Q} . \square

Based on the above results, we give a sufficient condition to guarantee the uniqueness of the equilibrium for the case of full column-rank matrices in the following theorem:

Theorem 2: Define $\mathbf{T}_k \triangleq \left(\mathbf{I} + P_T \sum_{i=1}^K \mathbf{H}_{i,k} \mathbf{H}_{i,k}^H \right)^{-1}$. Suppose that the channel matrices $\{\mathbf{H}_{k,k}\}_{k \in \Psi_K}$ are full column rank, i.e., $\text{rank}(\mathbf{H}_{k,k}) = M$. Let α_k be

$$\alpha_k = \frac{\rho(\mathbf{H}_{k,k}^H \mathbf{H}_{k,k}) \left\| \mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k \right\|_2}{\left(\lambda_{\min}(\mathbf{H}_{k,k}^H \mathbf{T}_k \mathbf{H}_{k,k}) \right)^2}, \quad (16)$$

where $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k$ is the Jacobian matrix of \mathbf{R}_k w.r.t. \mathbf{Q}_{-k} and is given by

$$\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k = \left[\mathbf{H}_{1,k}^* \otimes \mathbf{H}_{1,k}, \dots, \mathbf{H}_{k-1,k}^* \otimes \mathbf{H}_{k-1,k}, \mathbf{H}_{k+1,k}^* \otimes \mathbf{H}_{k+1,k}, \dots, \mathbf{H}_{K,k}^* \otimes \mathbf{H}_{K,k} \right] \quad (17)$$

Then for sufficient large maximum transmit power P_T , the NE of game \mathcal{G} is unique if

$$\alpha_k < \sqrt{\frac{1}{K-1}}, \forall k \in \Psi_K. \quad (18)$$

Proof: Please see Appendix C. \square

The importance of Condition (18) is that for given direct channel gains, it explicitly qualifies how small the multi-link interference each link can tolerate to guarantee the uniqueness of the NE. Hence, Condition (18) can be checked in practice to facilitate the admission control. In

contrast, [20] only showed that the number of the equilibria is determined by the cross-channel gains and the direct channel gains, without quantifying how they are related with each other.

2) Case of more general channel matrices

In practical systems, the channel matrices may not be full column rank. In this part, we consider the more general case without making any restrictive assumptions on the channel structure.

For each link k , we write the eigendecomposition of $\mathbf{H}_{k,k}$ as $\mathbf{H}_{k,k} = \bar{\mathbf{U}}_k \bar{\mathbf{\Lambda}}_k \bar{\mathbf{V}}_k^H$, where $\bar{\mathbf{U}}_k \in \mathbb{C}^{N \times r_k}$, $\bar{\mathbf{V}}_k \in \mathbb{C}^{M \times \bar{r}_k}$ are semi-unitary matrices, $\bar{\mathbf{\Lambda}}_k \in \mathbb{C}^{\bar{r}_k \times \bar{r}_k}$ is a diagonal matrix with positive eigenvalues, and \bar{r}_k is the rank of matrix $\mathbf{H}_{k,k}$, i.e., $\bar{r}_k = \text{rank}(\mathbf{H}_{k,k}) \leq \min\{M, N\}$. To maximize each link's EE in (4), each link k 's optimal covariance matrix should lie in the subspace orthogonal to the null-space of $\mathbf{H}_{k,k}$ for a given \mathbf{Q}_{-k} . It follows that the best response of the transmit covariance matrix of each link k belongs to the following class of matrices:

$$\mathbf{Q}_k = \bar{\mathbf{V}}_k \bar{\mathbf{Q}}_k \bar{\mathbf{V}}_k^H, \forall k \in \Psi_K, \quad (19)$$

with

$$\bar{\mathbf{Q}}_k \in \bar{\mathcal{W}}_k \triangleq \{ \mathbf{X} \in \mathbb{C}^{r_k \times r_k} \mid \mathbf{X} \in \mathbb{S}_+^{r_k \times r_k}, \text{tr}(\mathbf{X}) \leq P_T \}. \quad (20)$$

By inserting (19) into game \mathcal{G} in (5) and defining $\bar{\mathbf{H}}_{j,k} = \mathbf{H}_{j,k} \bar{\mathbf{V}}_j, \forall j, k$, game \mathcal{G} can be transformed into the following lower-dimensional game $\bar{\mathcal{G}}$, given as

$$(\bar{\mathcal{G}}) : \begin{aligned} \max_{\bar{\mathbf{Q}}_k} \text{EE}_k(\bar{\mathbf{Q}}_k, \bar{\mathbf{Q}}_{-k}) &= \frac{\log_2 |\mathbf{I} + \bar{\mathbf{H}}_{k,k}^H \bar{\mathbf{R}}_k^{-1} \bar{\mathbf{H}}_{k,k} \bar{\mathbf{Q}}_k|}{\text{tr}(\bar{\mathbf{Q}}_k) + P_C} \\ \text{s.t. } \bar{\mathbf{Q}}_k &\in \bar{\mathcal{W}}_k, \end{aligned} \quad \forall k \in \Psi_K, \quad (21)$$

where $\bar{\mathbf{Q}}_{-k} = \bar{\mathbf{Q}}_1 \times \cdots \times \bar{\mathbf{Q}}_{k-1} \times \bar{\mathbf{Q}}_{k+1} \times \cdots \times \bar{\mathbf{Q}}_K$ and $\bar{\mathbf{R}}_k^{-1} = \bar{\mathbf{R}}_k^{-1}(\bar{\mathbf{Q}}_{-k}) = \mathbf{I} + \sum_{j \neq k} \bar{\mathbf{H}}_{j,k} \bar{\mathbf{Q}}_j \bar{\mathbf{H}}_{j,k}^H$. Now the channel matrices $\bar{\mathbf{H}}_{k,k} \in \mathbb{C}^{N \times r_k}, \forall k$ are full column rank. Hence, we can apply the same derivations for Theorem 2 to obtain the uniqueness condition for this more general channel case, which is the same as that in (18) except that the channel matrices $\mathbf{H}_{j,k}, \forall j, k$ are replaced by $\bar{\mathbf{H}}_{j,k}, \forall j, k$.

To give more insights into the physical interpretation of the uniqueness conditions of the NE, Fig. 1 is plotted to show the probability for the uniqueness conditions to be satisfied for different cases. Specifically, full-rank square, fat and tall channel matrices are simulated. If the channel matrices are square or tall, condition (18) in Theorem 2 is applied to check the uniqueness condition of the NE. In the case of fat channel matrices, the above derived condition for general channel matrices is used to check the uniqueness condition of the NE. The probability

of the uniqueness of the NE is defined as the ratio of the number of channel matrices that guarantee the uniqueness condition to the total number of channel matrices. For simplicity, we consider a symmetric system with two links ($K = 2$) where the direct-channel distances for both links are set to be $D_{\text{direct}} = 1$ and both links have the same cross-channel distances (i.e., $D_{1,2} = D_{2,1} = D_{\text{cross}}$). The power constraint and the noise power are set to be $P_T = 10^{-2}\text{W}$, $\sigma^2 = 10^{-3}\text{W}$, respectively. Note that $P_T \gg \sigma^2$. These results are obtained by testing over 10000 random channel matrices whose entries are generated as the circularly symmetric, zero-mean, complex Gaussian random variables with variance equal to the square root of the channel path loss power, where the path loss exponent is assumed to be 3.5. Several interesting observations can be found from Fig. 1: 1) The uniqueness probability of the NE for all cases increase with the cross-channel distance corresponding the decrease in the interference. This is reasonable since in the extreme case when the cross-channel distance approaches infinity, the system becomes two independent point-to-point links. On the other hand, when one of the receivers is too close to one of its unintended transmitters, one of the two links should be shut down, which is instructive for practical use; 2) For the case of square channel matrices, the uniqueness probability decreases with the number of antennas; 3) For the case of fat channel matrices, increasing the number of transmitter antennas while keeping the number of receiver antennas fixed leads to an increase in the uniqueness probability of the NE. For example, the curve associated to the case of 2×5 MIMO channels is higher than that associated to the case of 2×3 MIMO channels for any given D_{cross} . Similar observations hold for the case of tall channel matrices. This observation is of significant importance: equipping unequal number of antennas at transmitters and receivers can dramatically improve the uniqueness condition of the NE.

IV. ASYNCHRONOUS DISTRIBUTED ENERGY EFFICIENT ALGORITHM

To reach the NE of game \mathcal{G} , we employ the totally asynchronous algorithm [40]. The main characteristic of the asynchronous algorithm is that some users are allowed to update their best response more frequently than others. This algorithm has been successfully employed to deal with the rate maximization game in [30], [31].

For the sake of readability, we briefly introduce the asynchronous algorithm and adapt it to our EE maximization problem. To this end, we first introduce some definitions and assume that the set of times at which the links update their solutions is the discrete set $\mathcal{T} = \{0, 1, 2, \dots\}$.

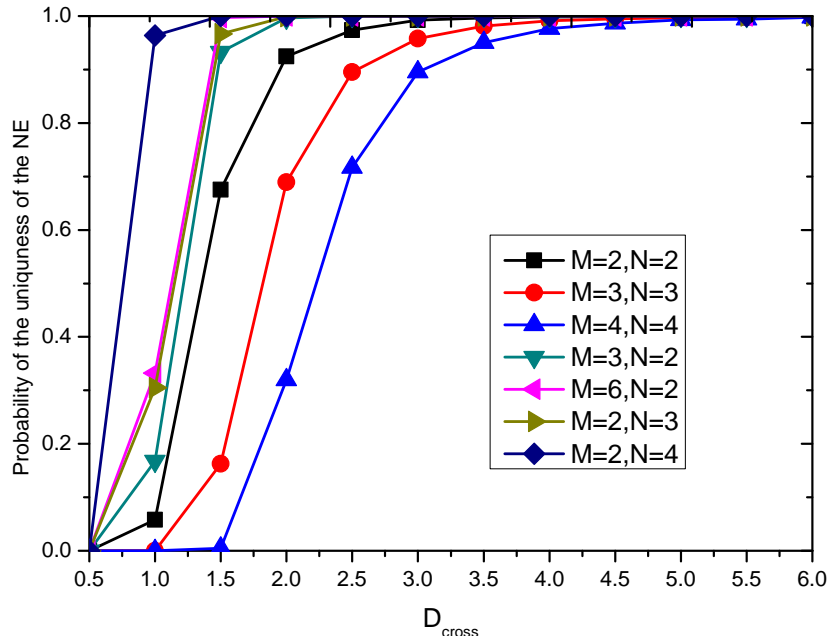


Fig. 1. Probability of the uniqueness of the NE for different numbers of antennas.

Let $\mathcal{T}^k \subseteq \mathcal{T}$ be the subset of times at which transmitter k updates its solution and let $\mathbf{Q}_k(t)$ be the updated transmit covariance matrix of transmitter k at time $t \in \mathcal{T}^k$. Denote $\tau_r^k(t)$ as the most recent time at which the interference from transmitter r is measured at receiver k at time t . Hence, at time t , link k updates its transmit covariance matrix based on the interference from

$$\mathbf{Q}_{-k}(\tau^k(t)) \triangleq (\mathbf{Q}_1(\tau_1^k(t)), \dots, \mathbf{Q}_{k-1}(\tau_{k-1}^k(t)), \mathbf{Q}_{k+1}(\tau_{k+1}^k(t)), \dots, \mathbf{Q}_K(\tau_K^k(t))). \quad (22)$$

To guarantee that the system is totally asynchronous, three conditions should be satisfied by the schedules $\{\tau_r^q(t)\}$ and $\{\mathcal{T}^k\}$ [40]: A1) $0 \leq \tau_r^k(t) \leq t$; A2) $\lim_{k \rightarrow \infty} \tau_r^k(t_k) = +\infty$; and A3) $|\mathcal{T}^k| = \infty$; where $\{t_k\}$ is a sequence of \mathcal{T}^k . Based on the above definitions, the asynchronous distributed EE algorithm is described in Algorithm 1, where T_{\max} is the maximum number of the iterations.

Remark 1 - Two special cases: Note that some well-known algorithms such as sequential and simultaneous algorithms [40], where the links update their solutions sequentially and simultaneously, are special cases of our algorithm. Each one of them can be obtained by appropriately choosing the scheduling parameters $\{\mathcal{T}^k\}$ and $\{\tau_r^q(t)\}$. In the sequential algorithm, all links update their own strategies sequentially, whose scheduling parameters $\{\mathcal{T}^k\}$ and $\{\tau_r^q(t)\}$ are

Algorithm 1 *Asynchronous Distributed Energy-Efficiency (ADEE) Algorithm*

```

1: Set  $t = 0$  and initialize  $\mathbf{Q}_k(0) \in \mathcal{W}_k, \forall k$ ;
2: FOR  $t = 0 : T_{\max}$ 
    FOR  $k = 1 : K$ 
        IF  $t \in \mathcal{T}^k$ 
            Update  $\mathbf{Q}_k(t+1)$  using Algorithm 2 in the following subsection based on  $\mathbf{Q}_{-k}(\tau^k(t))$ .
        ELSE
             $\mathbf{Q}_k(t+1) = \mathbf{Q}_k(t)$ ,
        END
    END
END
END

```

chosen as $\mathcal{T}^k = \{mK + k, m \in \mathcal{N}_+\} = \{k, K + k, 2K + k, \dots\}$, $\tau_r^k(t) = t, \forall k, r$, where \mathcal{N}_+ denotes the discrete set $\mathcal{N}_+ = \{0, 1, 2, \dots\}$. Similarly, in the simultaneous algorithm, all links update their own strategies simultaneously with the scheduling parameters chosen as $\mathcal{T}^k = \mathcal{N}_+, \tau_r^k(t) = t, \forall k, r$. In contrast, for the DBFEE [15] and TLEE algorithms [16], each transmitter should collect the knowledge of the channels to all the receivers over the network at the initial stage of the algorithm. Moreover, in each iteration all the receivers should feed back the complex-valued matrices to all the transmitters in the network, which is not scalable for large scale networks.

A. Per-link Response Problem

The best response problem for link k can be written as ²

$$\max_{\mathbf{Q}_k \in \mathcal{W}_k} \frac{C_k(\mathbf{Q}_k)}{\text{tr}(\mathbf{Q}_k) + P_C}. \quad (23)$$

Since the numerator and denominator in the objective function of Problem (23) are concave and affine in \mathbf{Q}_k respectively, the objective function of (23) is a pseudo-concave function [23]. Moreover, the constraint in (23) is convex. Hence, problem (23) can be solved by the following lemma, the proof of which can be found in Proposition 6 of [41].

²As \mathbf{Q}_{-k} is treated as constant, $C_k(\mathbf{Q}_k)$ is used to represent $C_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$ for simplicity.

Lemma 4: Define function $G(\kappa)$ as

$$G(\kappa) \triangleq \max_{\mathbf{Q}_k \in \mathcal{W}_k} \log_2 |\mathbf{I} + \mathbf{H}_{k,k}^H \mathbf{R}_k^{-1} \mathbf{H}_{k,k} \mathbf{Q}_k| - \kappa (\text{tr}(\mathbf{Q}_k) + P_C), \quad (24)$$

For fixed κ , the solution of (24) is denoted as $\mathbf{Q}_k^*(\kappa)$. Then, solving (23) is equivalent to finding the root of the equation $G(\kappa^*) \triangleq C_k(\mathbf{Q}_k^*(\kappa^*)) - \kappa^* (\text{tr}(\mathbf{Q}_k^*(\kappa^*)) + P_C) = 0$. \square

Lemma 4 gives us insights to solve (23). We can solve Problem (24) with fixed κ firstly, while the optimal κ can be searched via the Dinkelbach method [42].

Now we attempt to solve (24) for fixed κ . To this end, we first write the eigenvalue decomposition (EVD) of $\mathbf{H}_{k,k}^H \mathbf{R}_k^{-1} \mathbf{H}_{k,k}$ for each $k \in \Psi_K$ as

$$\mathbf{H}_{k,k}^H \mathbf{R}_k^{-1} \mathbf{H}_{k,k} = \mathbf{U}_k \mathbf{D}_k \mathbf{U}_k^H, \quad (25)$$

where \mathbf{U}_k is a semi-unitary matrix of the eigenvectors with $r_k \triangleq \text{rank}(\mathbf{H}_{k,k}^H \mathbf{R}_k^{-1} \mathbf{H}_{k,k})$, $\mathbf{D}_k \in \mathbb{R}_{++}^{r_k \times r_k}$ is a diagonal matrix with $\mathbf{D}_k = \text{diag}\{d_{k,1}, \dots, d_{k,r_k}\}$ being the eigenvalues.

Then, given $k \in \Psi_K$ and $\mathbf{Q}_{-k} \in \mathcal{W}_{-k}$, the solution to problem (24) with fixed κ is [13]:

$$\mathbf{Q}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^H, \quad (26)$$

where $\mathbf{\Lambda}_k = \text{diag}\{q_{k,1}, \dots, q_{k,r_k}\}$ represents the power allocations on all subchannels with

$$q_{k,m} = \left[\frac{1}{(\kappa + \lambda_k) \ln 2} - \frac{1}{d_{k,m}} \right]^+, m \in \{1, 2, \dots, r_k\}, \quad (27)$$

where $\lambda_k \geq 0$ is the Lagrange multiplier associated with the power constraint, which should be chosen to satisfy the complementarity slackness condition: $\lambda_k (\text{tr}(\mathbf{Q}_k) - P_T) = 0$.

After solving problem (24), we utilize the Dinkelbach method [42] to update κ as follows

$$\kappa_{n+1} = \frac{C_k(\mathbf{Q}_k^{(n)*}(\kappa_n))}{\text{tr}(\mathbf{Q}_k^{(n)*}(\kappa_n)) + P_C}, n = 1, 2, \dots, \quad (28)$$

where n is the iteration index.

To summarize the above analysis, we give the following algorithm to solve the per-link problem in (23).

Algorithm 2 *The Dinkelbach method to solve (23)*

- 1: Initialization: κ_0 satisfying $G(\kappa_0) \geq 0$, tolerance ε , iteration number $n = 0$;
 - 2: For given κ_n , solve (24) to get the optimal $\mathbf{Q}_k^{(n)*}(\kappa_n)$;
 - 3: If $|G(\kappa_n)| > \varepsilon$, update κ_{n+1} in (28), $n = n + 1$, go back to step 2. Otherwise, terminate.
-

Remark 2 - Distributed nature of the algorithm: For transmitter k to update \mathbf{Q}_k , it requires receiver k to feed back the channel matrix $\mathbf{H}_{k,k}$ and the IPN covariance matrix \mathbf{R}_k according to (24). It is well-known that the channel matrix $\mathbf{H}_{k,k}$ can be estimated at the receiver. For the IPN covariance matrix \mathbf{R}_k , it can be easily computed at the receiver as follows: First, transmitter k sends a sequence of training sequence to receiver k . The statistical information of the training sequence is assumed to be known at the receiver. Then the IPN covariance matrix can be obtained by subtracting the covariance matrix of the received training sequence from the covariance matrix of the total received signal³. Hence, the ADEE algorithm can be performed in a totally distributed and asynchronous way without the need of information exchange among different links. In contrast, for the DBFEE [15] and TLEE algorithms [16], in each iteration all the receivers should feed back the complex-valued matrices to all the transmitters in the network, which is not scalable for large scale networks.

B. Convergence Analysis

In this section, the sufficient condition that guarantees the global convergence of the ADEE algorithm is given. Interestingly, we find that this convergence condition is the same as the uniqueness condition obtained in Theorem 2, as proved in the following.

Theorem 3: In the case of full column-rank channel matrices, suppose that condition (18) in Theorem 2 is satisfied and the maximum transmit power is large enough, then as $T_{\max} \rightarrow \infty$, the sequence generated by the ADEE algorithm converges to the unique NE of game \mathcal{G} , for any given set of feasible initial conditions and updating schedules.

Proof: Please see Appendix D. □

The proof for the more general case can be derived similarly. It is omitted for brevity.

Remark 3 - Robustness of the algorithm: The condition for the convergence of the ADEE algorithm is independent of the update schedule for each link. Hence, all special cases of the ADEE algorithm like the well-known sequential and simultaneous updating are guaranteed to converge under the same condition in (18). Furthermore, there is no restrict constraints on the updating schedule on each link, so that some links are allowed to update their transmit covariance matrices more often than others, without affecting the convergence of the algorithm.

³Note that this has been done in most of the existing distributed algorithms [12], [13], [15], [24], [26], [27], [30]–[36].

V. PERFORMANCE ANALYSIS OF THE PROPOSED ALGORITHM

In this section, we give the performance analysis for the proposed algorithm. Firstly, the impact of the circuit power consumption on the system performance is studied. Then, we investigate the tradeoff between SE and EE.

A. Impact of Circuit Power Consumption on the System Performance

In this part, we investigate the impact of circuit power consumption on the overall SE and EE performance. Due to the coupling interference among different links, it is difficult to provide the analytical results. To simplify the analysis, we only consider the extreme case that all the links are separated far away with each other so that the interference among all the links reduces to almost zero. In this case, the overall system can be regarded as K independent links. Then, we only need to study the impact of circuit power consumption on the each link k 's SE and EE performance, which is given in the following theorem.

Theorem 4: When the links are separated far away and no power constraints are imposed at the transmitters, the maximum achievable EE of each link k for a given circuit power consumption P_C decreases with P_C , but the corresponding SE of each link k increases with P_C .

Proof: Please see Appendix E. □

To validate the analysis in Theorem 4, we plot a figure in Fig. 2 to show the impact of circuit power consumption on system's SE and EE for the same scenario in Fig. 1 with $D_{\text{cross}} = 5$. The sum-EE and the sum-SE are defined as $\text{EE}_{\text{sum}} = \sum_{k=1}^K \text{EE}_k$, $\text{SE}_{\text{sum}} = \sum_{k=1}^K C_k$, respectively. For illustration purpose, the y-axis is shown by $10\log_{10}\text{EE}_{\text{sum}}$. It can be seen from the figure that the sum-EE decreases with the circuit power consumption while the corresponding sum-SE increases with it for all considered numbers of antennas, which validates the correction of the theorem. From the simulation section, the above trend also holds for the general case with sizeable interference among the links, though the above theorem is derived for the extreme case.

B. Tradeoff between Spectral and Energy Efficiency

In this part, we study the tradeoff between SE and EE for the proposed ADEE algorithm and the asynchronous distributed spectral efficient (ADSE) algorithm [31]. Generally, it is difficult to study this problem analytically due to the coupling interference. For this reason, we investigate the tradeoff for two special cases: low transmit power constraint and high transmit power constraint.

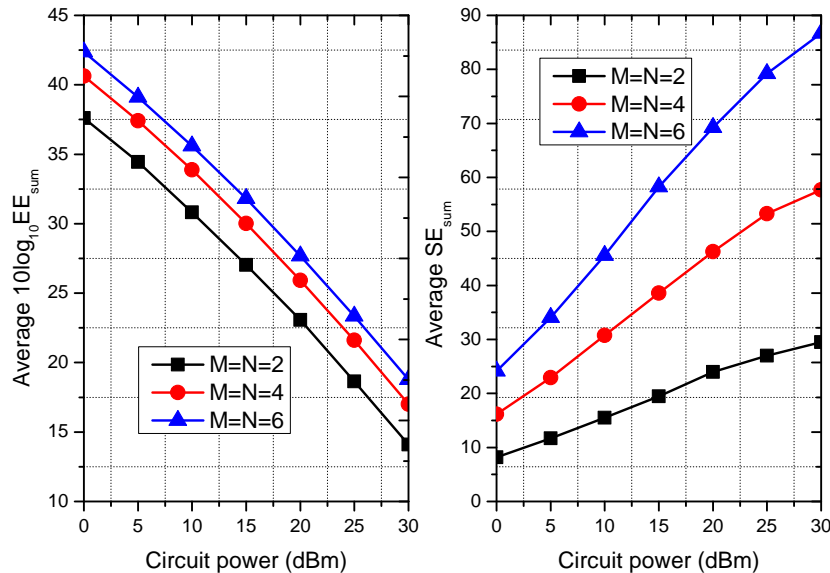


Fig. 2. The impact of circuit power consumption on system's SE and EE when links are separated far away.

1) *Low transmit power constraint: $P_T \rightarrow 0$* : In this case, the interlink interference can be neglected, i.e., $\mathbf{R}_k \approx \mathbf{I}, \forall k$, and the network reduces to K independent point-to-point links. Hence, we only need to analyze one link's performance. Without loss of generality, we only focus on the performance of link k . Denote the transmit power of link k of the ADEE algorithm as $P_k \in [0, P_T]$. Since $P_T \rightarrow 0$, from (E.3) there is only one stream for link k that transmits with positive power. Hence, the SE for link k can be written as $C_k(P_k) = \log_2(1 + d_{k,1}P_k)$, where $d_{k,1}$ is the maximum eigenvalue of $\mathbf{H}_{k,k}^H \mathbf{H}_{k,k}$. The optimal EE for link k can be written as

$$\text{EE}_k(P_k) = \frac{C_k(P_k)}{P_k + P_C}. \quad (29)$$

Then, we have

$$\lim_{P_k \rightarrow 0} \text{EE}'_k(P_k) = \frac{d_{k,1}}{\ln 2 P_C} > 0, \quad (30)$$

where $\text{EE}'_k(P_k)$ is the derivative of $\text{EE}_k(P_k)$ w.r.t. P_k . Hence, $\text{EE}_k(P_k)$ is increasing for $P_k \in [0, P_T]$. Then, the optimal transmit power is $P_k^* = P_T$, i.e., link k transmits with its maximum power.

For the ADSE algorithm, each link maximizes its SE selfishly by always using its maximum power, i.e., $P_k^* = P_T$. Hence, both the ADEE algorithm and ADSE algorithm achieve the same SE and EE.

2) *High transmit power constraint:* $P_T \rightarrow \infty$: In this case, the interlink interference cannot be neglected any more. Due to the interference, the analysis becomes more difficult. To simplify the analysis and get insights, we consider a symmetric system similar to one in [20] for the SISO frequency selective interference channel. We assume that $\mathbf{H} = \mathbf{H}_{k,k}, \forall k$ and $\mathbf{H}_{j,k} = \alpha\mathbf{H}, \forall j \neq k$, where α is a constant. In this symmetric system, all links transmit with the same covariance matrices, i.e., $\mathbf{Q} = \mathbf{Q}_k, \forall k$. Denote the transmit power as $\text{tr}(\mathbf{Q}) = P$.

The overall network EE is

$$\text{EE} = \frac{K \log_2 |\mathbf{I} + \mathbf{H}^H \mathbf{R}^{-1} \mathbf{H} \mathbf{Q}|}{P + P_C}, \quad (31)$$

and the overall network SE is

$$\text{SE} = K \log_2 |\mathbf{I} + \mathbf{H}^H \mathbf{R}^{-1} \mathbf{H} \mathbf{Q}|, \quad (32)$$

where $\mathbf{R} = \mathbf{I} + (K - 1)\alpha^2 \mathbf{H} \mathbf{Q} \mathbf{H}^H$.

For the ADSE algorithm, each link selfishly maximizes its own SE by using its maximum transmit power, i.e., $\text{tr}(\mathbf{Q}) = P = P_T$. When $P_T \rightarrow \infty$, the interlink interference becomes very large. According to Theorem 3 of [24], one of the optimal solutions employs beamforming (1-stream signaling) for all links. Hence, the optimal covariance matrix can be written as $\mathbf{Q} = P_T \mathbf{w} \mathbf{w}^H$, where \mathbf{w} is the beam direction with unit norm. Then, the network SE of the ADSE algorithm in the high transmit power is upper bounded by

$$\text{SE}_{\text{ADSE}} = \lim_{P_T \rightarrow \infty} K \log_2 \left(1 + \frac{g P_T}{1 + \alpha^2 (K - 1) g P_T} \right) = K \log_2 \left(1 + \frac{1}{\alpha^2 (K - 1)} \right), \quad (33)$$

where $g = \mathbf{w}^H \mathbf{H}^H \mathbf{H} \mathbf{w}$ is a constant. The corresponding EE is $\text{EE}_{\text{ADSE}} = \lim_{P_T \rightarrow \infty} \text{SE}_{\text{ADSE}} / (P_T + P_C) = 0$, which is not desirable from the EE point of view.

On the other hand, for the ADEE algorithm, the system will not use the maximum power to transmit due to the power value in the denominator of (31). Hence, the interlink interference may not be so large. Thus, the above derivations for the high interference scenario are not applicable and it is difficult to analyze the performance of the ADEE algorithm in this case. However, from the above discussion, we can conclude that its overall EE will increase when the transmit power constraint is low, and keeps constant when the transmit power constraint is high, which is significantly larger than that of the ADSE algorithm. The overall SE has the same trend as the overall EE.

VI. SIMULATION RESULTS

We consider an ad hoc network contained in a $250\text{m} \times 250\text{m}$ square area, in which all links are randomly distributed. The distance from one transmitter to its unintended receiver is at least 35m. The channel is modeled by path-loss [43] and independent Rayleigh fading with the complex normal distribution, $\mathcal{CN}(0, 1)$. The channel path-loss is modeled as $38.46 + 35\log_{10}(d)$ [43]. Each channel realization is obtained by generating a random set of links' positions as well as fading channel realizations. It is assumed that the transmitters and the receivers have the same number of antennas. Unless otherwise specified, the other main system parameters are given in Table I. For comparisons, the metrics used are the sum-EE and the sum-SE, which are defined as $\text{EE}_{\text{sum}} = \sum_{k=1}^K \text{EE}_k$, $\text{SE}_{\text{sum}} = \sum_{k=1}^K C_k$, respectively.

TABLE I
MAIN SIMULATION PARAMETERS

Parameters	Value
Number of links K	4
Number of transmit antennas M	4
Number of receiver antennas N	4
Direct-channel distance D_{direct}	80 m
Noise power σ^2	-106 dBm
Circuit power consumption P_C	23 dBm [44]
Tolerance ε	$10^{(-5)}$
Maximum transmit power P_T	30 dBm
Maximum number of iterations T_{max}	20
Channel path loss model	$38.46 + 35\log_{10}(d)$ [43]

A. Convergence behavior of the ADEE algorithm

Fig.3 illustrates the convergence behavior of the ADEE algorithm for different updating schedules for one randomly generated channel realization. For comparison, the performance of another updating scheme, named ‘unbalanced ADEE’, is also shown, where the parameters are chosen as $\mathcal{T}^k = \{k, 2k, 3k, \dots\}$, $\tau_r^k(t) = t, \forall k, r$. Hence, in the unbalanced ADEE, Link 1 is set to update its strategy faster than the other three links, and Link 4 is set to be the slowest to update its strategy. Also, in each time, the number of links updating their strategies is no

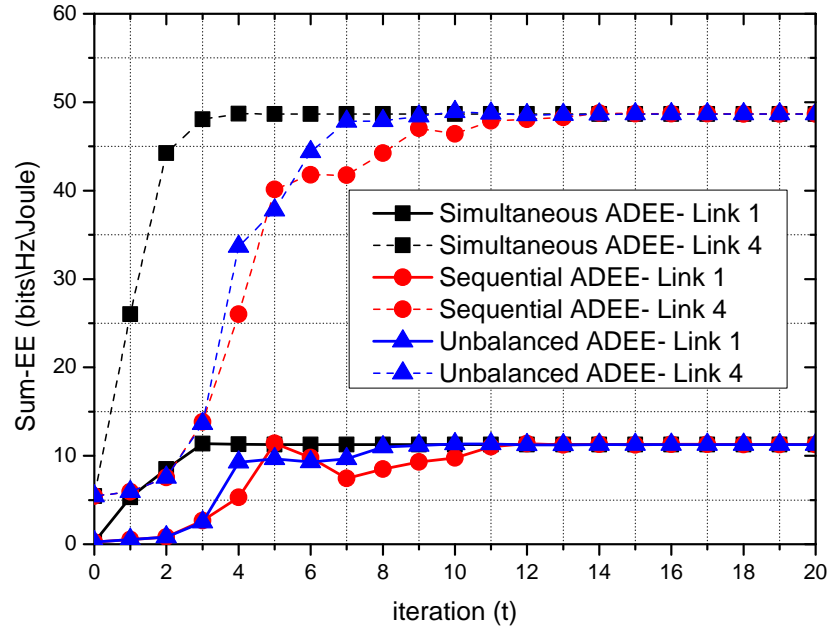


Fig. 3. Convergence behaviors for different updating schedules.

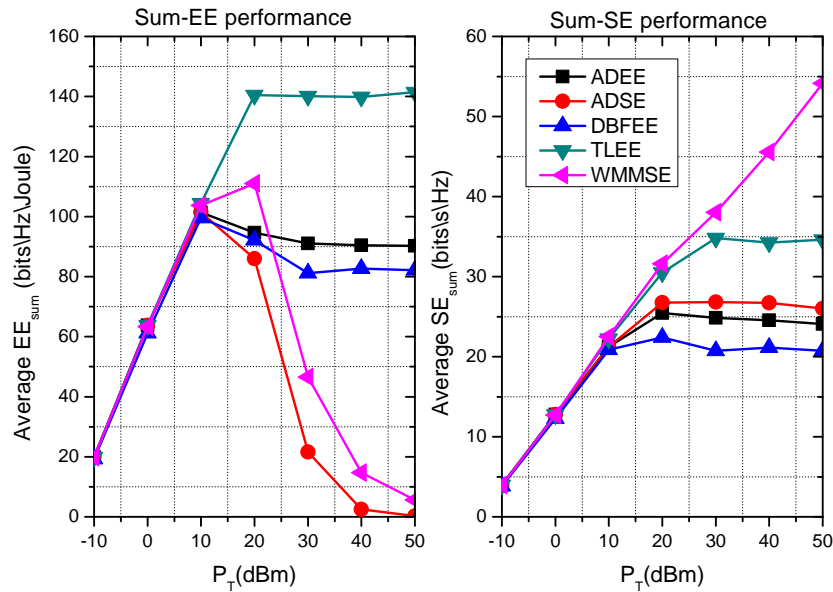


Fig. 4. Average sum-EE (left) and sum-SE (right) versus P_T for various algorithms.

less than that of the sequential ADEE (only one link), no more than that of the simultaneous ADEE (all four links). To make the figure not too crowded, we report only the curves of two links (Link 1 and Link 4). It can be seen from this figure that the simultaneous ADEE converges faster than the other two schemes, and can converge within a few iterations. However, it takes about 12 iterations for the sequential ADEE to converge. The reason is that each user in the sequential ADEE is forced to wait for all the users scheduled in advance, before updating its own strategy. Moreover, the unbalanced ADEE converges a little faster than the sequential ADEE due to more links involved in updating their strategies in each iteration. From this figure, we find that different scheduling methods yield almost the same performance. For this reason, we only report the performance of the simultaneous ADEE due to its rapid convergence speed in the following simulations.

B. Performance Comparison with Existing Algorithms

We next compare the performance of the ADEE algorithm with some existing algorithms, including the ADSE algorithm [31] where each link always uses its maximum transmit power, and DBFEE algorithm [15]. Since in the proposed ADEE algorithm each link only attempts to maximize its own EE selfishly, its achieved overall EE is generally suboptimal. Hence, it is interesting to study the performance gap between the proposed ADEE algorithm and the (near-)optimal sum-EE maximization algorithm. For this reason, we simulate the TLEE algorithm in [16] that aims at the sum-EE maximization. Moreover, the WMMSE algorithm in [14] that focuses on the sum-SE maximization is also simulated.

1) *Impact of the Maximum Transmit Power on the System Performance:* Fig. 4 shows the average sum-EE (left) and the average sum-SE (right) versus the maximum transmit power for various algorithms. It can be seen that in the low transmit power regime, i.e., $P_T \leq 10$ dBm, both the sum-EE and the sum-SE of all algorithms monotonically increase with the maximum transmit power. Interestingly, we find that in this regime the proposed ADEE algorithm almost achieves the optimal sum-EE and sum-SE. That is, the performance gap with the optimal sum-EE and sum-SE performance approaches zero. Fortunately, many short-distance wireless networks operate in this regime, such as ad hoc networks [45], [46], femtocells, and wireless sensor networks [47]. However, in the high transmit power regime, the sum-EE of the ADRM algorithm decreases dramatically, while its sum-SE performance becomes stable. This is because in the

high transmit power regime, the system becomes interference limited and increasing the transmit power may slightly help increase the sum-SE, which leads to a significant performance loss in terms of the sum-EE due to the increased transmit power. Note that the sum-SE achieved by the ADEE algorithm is comparable with that achieved by the ADRM algorithm, which corroborates with the analysis in Section V. Fig. 4 also shows that the ADEE algorithm outperforms the DBF algorithm in terms of both the sum-EE and the sum-SE. This is because the DBFEE algorithm is primarily designed for the symmetry system and may not be suitable for the asymmetric ad hoc network considered here. Moreover, this algorithm requires a substantial feedback overhead on the network and further the power consumption due to these information exchanges is not considered. As expected, both the sum-EE and the sum-SE performance of the TLEE algorithm is superior to that of the ADEE algorithm in the high transmit power regime due to the selfish nature of the ADEE algorithm. However, this benefit comes at the cost of the heavy information exchange overhead, high computational complexity, synchronization requirements of the networks, and the need for coordination among the users. In contrast, the proposed ADEE algorithm does not need information exchange among different links, and can be implemented in a totally distributed and asynchronous manner, which is appealing for some practical applications.

2) *Impact of the Number of Antennas on the System Performance:* Fig. 5 illustrates the average sum-EE (left) and the average sum-SE (right) versus the number of antennas for various algorithms. As expected, increasing the number of antennas at both the transmitters and the receivers leads to a significant increase in both the sum-EE and the sum-SE due to the fact that a larger number of degrees of freedom (DoF) in the spatial domain can be exploited to strengthen the signal power received by the intended receiver while avoiding the interference imposed on the unintended ones. It can be seen from this figure that the ADEE algorithm outperforms the DBF algorithm in terms of the sum-EE and the sum-SE. The performance gain monotonically increases with the number of antennas, meaning the ADEE algorithm uses the spatial resources more effectively. Again, the ADEE algorithm is observed to achieve a significantly higher sum-EE than that achieved by the ADRM algorithm and the performance gain becomes more significant with more antennas. The sum-SE of the ADRM algorithm is slightly higher than that of the ADEE algorithm. As expected, the TLEE algorithm has superior performance over the ADEE algorithm in terms of the sum-EE and sum-SE.

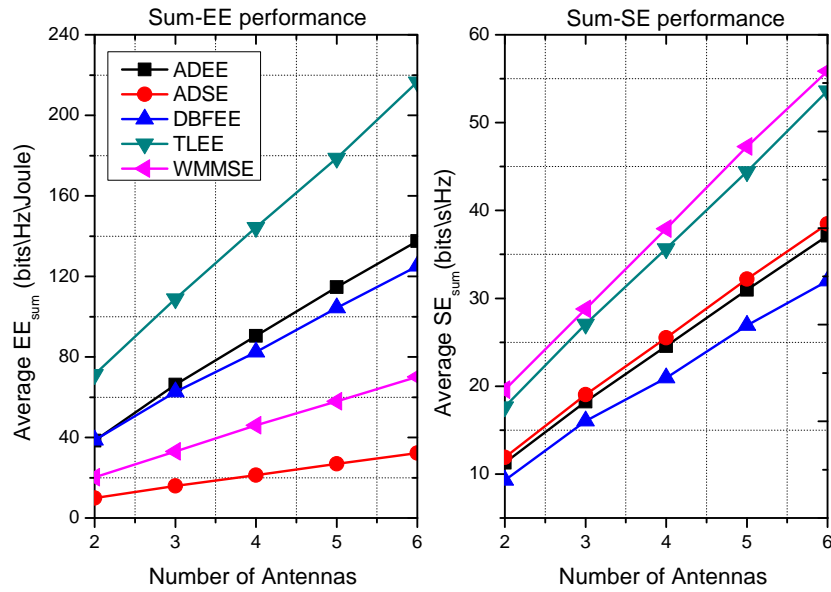


Fig. 5. Average sum-EE (left) and sum-SE (right) versus the number of antennas for various algorithms.

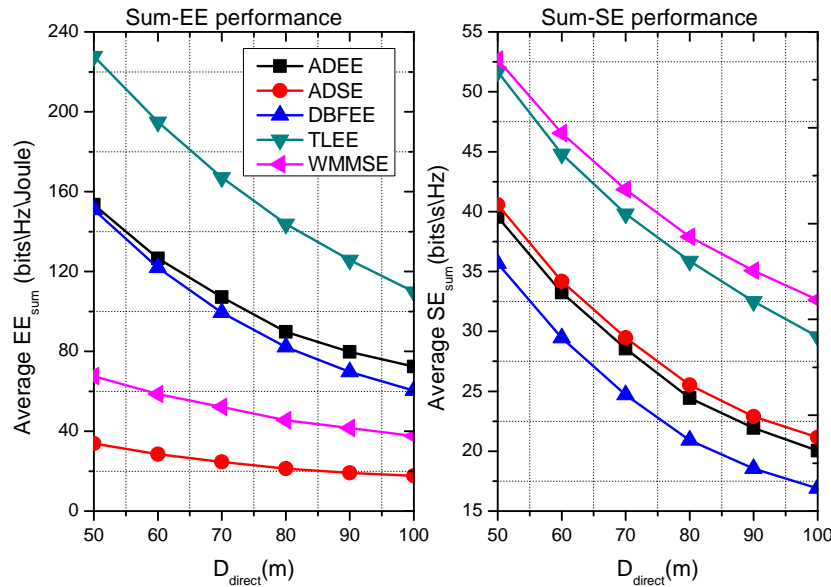


Fig. 6. Average sum-EE (left) and sum-SE (right) versus the direct-channel distance D_{direct} for various algorithms.

3) *Impact of the Direct-Channel Distance on the System Performance:* Fig. 6 illustrates the impact of increasing the direct-channel distance D_{direct} on the sum-EE and sum-SE. It can be seen from this figure that both the sum-EE and sum-SE monotonically decreases upon increasing D_{direct} . The reason is that increasing D_{direct} will degrade the direct-channel gains, and thus reducing the attainable sum-EE and sum-SE. This figure shows much of the same properties as Fig. 5. For example, the ADEE algorithm outperforms the DBFEE and ADRM algorithms in terms of the sum-EE, and the sum-SE loss is negligible compared with the ADRM algorithm. With much more information exchange overhead and higher computational complexity, the TLEE and WMMSE algorithms have the best sum-EE and sum-SE performance, respectively.

4) *Impact of the Circuit Power on the System Performance:* The impact of the circuit power on the attainable sum-EE and sum-SE is characterized in Fig. 7. For illustration purpose, the y-axis is shown by $10\log_{10}EE_{\text{sum}}$. As expected, the sum-EE achieved by all the algorithms decreases with the increase of the circuit power due to the increased power consumption and finally converges to almost the same value. On the other hand, the corresponding sum-SE of the ADEE algorithm and the DBFEE algorithm monotonically increases with the increase of circuit power and finally converges to the sum-rate achieved by the ADRM algorithm. This indicates that more circuit power will encourage each link to use more power to achieve a higher SE, which shows the trend in Theorem 4 also holds in the more general case. Note that the similar trend has been observed for the MIMO interference channel in [15]. It is interesting to find that the sum-EE gain provided by the TLEE algorithm over the ADEE algorithm is small and fixed during the overall circuit power consumption regime.

5) *Impact of number of links on the System Performance:* Finally in Fig. 8, we investigate the impact of increasing the number of links on the sum-EE and sum-SE performance of different algorithms. It can be seen from this figure that the sum-EE of the ADEE algorithm decreases with the increase of the number of the links. The reason is that when the number of links increases, the interference power received at the receivers increases. To compensate for this negative effects, each transmitter will increase its transmit power. However, this slightly improves the sum-SE but significantly reduces the sum-EE. To improve the performance of our algorithm for this interference-limited scenario, we may combine our algorithm with some scheduling methods, which is beyond the scope of this paper. On the other hand, by increasing the number of links, the sum-EE performance of the DBFEE algorithm gradually increases. This can be explained as

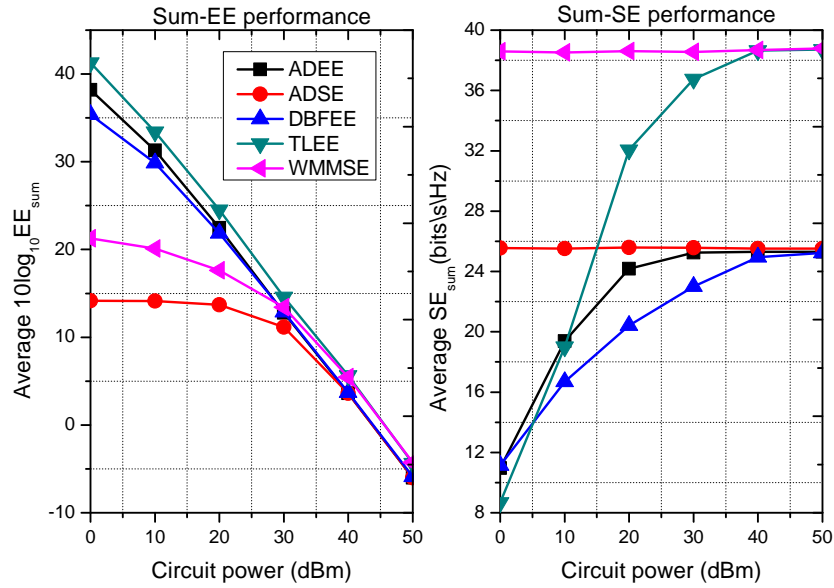


Fig. 7. Average sum-EE (left) and sum-SE (right) versus the circuit power for various algorithms.

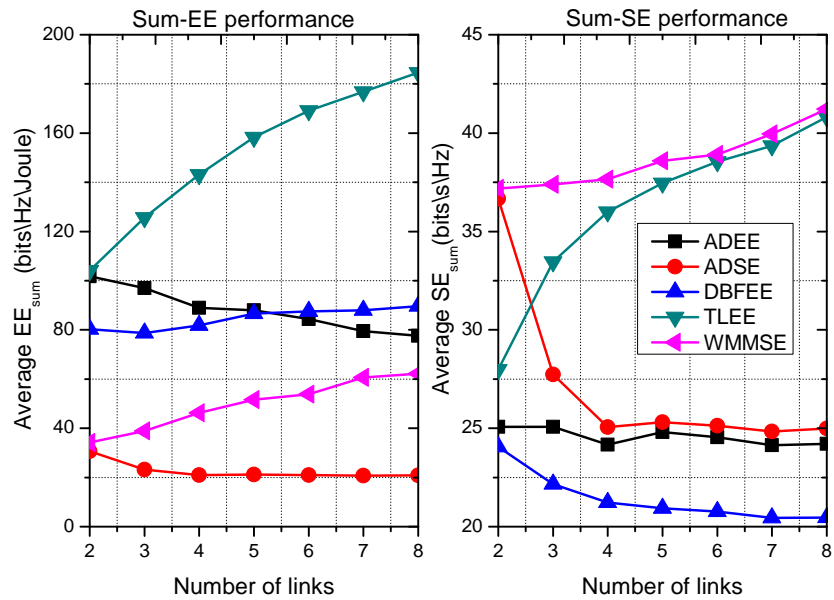


Fig. 8. Average sum-EE (left) and sum-SE (right) versus the number of links for various algorithms.

follows: As the number of links increases, the interference power perceived at different receivers become comparable with each other due to the law of large numbers. As a result, the system will become symmetry, for which the DBFEE algorithm will be suitable. It can also be observed from this figure that there is a point beyond which the DBFEE algorithm outperforms the ADEE algorithm. However, this superiority comes at the cost of more energy consumption incurred by the information exchange, which is not accounted for in the simulations. Finally, it is observed that both the sum-EE and sum-SE increases with the number of links due to the multiuser diversity.

VII. CONCLUSION

In this paper, we have considered a game theoretical formulation of the maximization of the EE on each link, subject to the power constraints, in the MIMO interference channel. We have provided a complete characterization of the game, by showing the existence of the NE and deriving the sufficient conditions for the uniqueness of the NE for the case of large enough maximum transmit power constraint. We have provided a totally asynchronous iterative distributed algorithm, named the ADEE algorithm, to reach the NE of this game. The ADEE algorithm has three advantages: no rigid scheduling in the updates of the players is required, the synchronization requirement of both sequential and simultaneous algorithms can be removed and information exchange among the players is not necessary, which make the proposed algorithm appealing to implement in practice. Furthermore, the sufficient conditions guaranteeing the convergence of the algorithm have been provided. Interestingly, our sufficient condition does not depend on the scheduling of the links. Extensive simulation results have shown that the ADEE algorithm performs better than the DBFEE and ADRM algorithms in terms of the sum-EE performance, and is comparable to the ADRM algorithm in terms of the sum-SE performance.

APPENDIX A

PROOF OF LEMMA 1

Before proving Lemma 1, the following definition is given.

Definition 2 [48]: A real-valued differentiable function $f(\mathbf{X}) : \mathbf{X} \in \mathbb{S}_+^{M \times M}$, is (strictly) quasiconcave on $\mathbb{S}_+^{M \times M}$ if its sublevel sets $S_\alpha = \{\mathbf{X} \in \mathbb{S}_+^{M \times M} | f(\mathbf{X}) \geq \alpha\}$, for $\alpha \in \mathbf{R}$, are (strictly) convex. □

For simplicity, we omit the dependency of $EE_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$ and $C_k(\mathbf{Q}_k, \mathbf{Q}_{-k})$ on \mathbf{Q}_{-k} . Define two functions $f_1(\mathbf{Q}_k) = C_k(\mathbf{Q}_k)$ and $f_2(\mathbf{Q}_k) = \text{tr}(\mathbf{Q}_k) + P_C$, and define $f(\mathbf{Q}_k) \triangleq EE_k(\mathbf{Q}_k) = \frac{f_1(\mathbf{Q}_k)}{f_2(\mathbf{Q}_k)}$.

Now, we give the proof of the lemma. Since $f_2(\mathbf{Q}_k) > 0$, S_α is equivalent to

$$S_\alpha = \{\mathbf{Q}_k \in \mathbb{S}_+^{M \times M} | f_1(\mathbf{Q}_k) - \alpha f_2(\mathbf{Q}_k) \geq 0\}. \quad (\text{A.1})$$

Since $f_2(\mathbf{Q}_k)$ is affine in \mathbf{Q}_k , S_α is (strictly) convex for any given α if $f_1(\mathbf{Q}_k)$ is (strictly) concave in \mathbf{Q}_k . In the following, we will prove $f_1(\mathbf{Q}_k)$ is concave w.r.t. \mathbf{Q}_k for any channel matrix $\mathbf{H}_{k,k}$ and is strictly concave w.r.t. \mathbf{Q}_k if $\mathbf{H}_{k,k}$ is full column rank.

To this end, we adopt the technique in [24]. We consider the convex combination of two different feasible $\mathbf{X}_k \in \mathcal{W}_k$ and $\mathbf{Z}_k \in \mathcal{W}_k$, which is

$$\mathbf{Q}_k = t\mathbf{Z}_k + (1-t)\mathbf{X}_k = \mathbf{X}_k + t\mathbf{Y}_k, \quad (\text{A.2})$$

where $0 \leq t \leq 1$, and $\mathbf{Y}_k = \mathbf{Z}_k - \mathbf{X}_k$, which is a non-zero Hermitian matrix. Obviously, $\mathbf{Q}_k \in \mathcal{W}_k$. Then $f_1(\mathbf{Q}_k)$ is (strictly) concave w.r.t. \mathbf{Q}_k if and only if $(d^2/dt^2)f_1(\mathbf{Q}_k) \leq 0$ ($(d^2/dt^2)f_1(\mathbf{Q}_k) < 0$) for any feasible \mathbf{X}_k and \mathbf{Z}_k . Denote $\mathbf{A}_k = \mathbf{H}_{k,k}^H \mathbf{R}_k^{-1} \mathbf{H}_{k,k}$, the second derivative of $f_1(\mathbf{Q}_k)$ w.r.t. t can be calculated as

$$\frac{d^2 f_1(\mathbf{Q}_k)}{dt^2} = -\frac{1}{\ln 2} \text{tr} \left(\mathbf{A}_k \mathbf{Y}_k (\mathbf{I} + \mathbf{A}_k \mathbf{Q}_k)^{-1} \mathbf{A}_k \mathbf{Y}_k (\mathbf{I} + \mathbf{A}_k \mathbf{Q}_k)^{-1} \right). \quad (\text{A.3})$$

Since $(\mathbf{I} + \mathbf{A}_k \mathbf{Q}_k)^{-1} \in \mathbb{S}_{++}^{M \times M}$, it is possible to write $(\mathbf{I} + \mathbf{A}_k \mathbf{Q}_k)^{-1} = \mathbf{D}_k \mathbf{D}_k^H$ with $\mathbf{D}_k^H \in \mathbb{S}_{++}^{M \times M}$.

Then, it is seen that

$$\frac{d^2 f_1(\mathbf{Q}_k)}{dt^2} = -\frac{1}{\ln 2} \text{tr} \left((\mathbf{D}_k^H \mathbf{A}_k \mathbf{Y}_k \mathbf{D}_k) (\mathbf{D}_k^H \mathbf{A}_k \mathbf{Y}_k \mathbf{D}_k) \right) \leq 0. \quad (\text{A.4})$$

Furthermore, when $\mathbf{H}_{k,k}$ is a full column-rank matrix, $\mathbf{A}_k \mathbf{Y}_k$ is always non-zero since \mathbf{Y}_k is a non-zero matrix. Then $(d^2/dt^2)f_1(\mathbf{Q}_k) < 0$ always holds. Hence, the proof is completed.

APPENDIX B

PROOF OF THEOREM 1

We prove the existence of NE for game \mathcal{G} using the following well-known game theory result.

Lemma 5 [49]: The strategic noncooperative game $\mathcal{G} = \{\Psi_K, \{\mathcal{W}_k\}_{k \in \Psi_K}, \{U_k\}_{k \in \Psi_K}\}$ admits at least one NE if, for all $k \in \Psi_K$: 1) the set \mathcal{W}_k is a nonempty compact convex subset of a Euclidean space, and 2) the payoff function U_k is quasiconcave on \mathcal{W}_k . \square

It can be easily verified that game \mathcal{G} satisfies the two conditions required by Lemma 5 according to Lemma 1 and the convexity of the admissible power set.

APPENDIX C

PROOF OF THEOREM 2

Before proving Theorem 2, we provide three lemmas [38] that will be used in our derivations.

Lemma 6: Given matrices $\mathbf{A} \in \mathbb{C}^{m \times n}$, $\mathbf{B} \in \mathbb{C}^{n \times l}$, the relation $\|\mathbf{AB}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$ holds. \square

Lemma 7: If $\mathbf{A}, \mathbf{B} \in \mathbb{S}_{++}^{n \times n}$, then we have $\rho(\mathbf{AB}) \leq \rho(\mathbf{A})\rho(\mathbf{B})$. \square

Lemma 8: Given matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ with eigenvalues $\{\lambda_i, i = 1, \dots, n\}$ and eigenvectors $\{\mathbf{x}_i, i = 1, \dots, n\}$, and matrix $\mathbf{B} \in \mathbb{C}^{m \times m}$ with eigenvalues $\{\mu_i, i = 1, \dots, m\}$ and eigenvectors $\{\mathbf{y}_i, i = 1, \dots, m\}$, the eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ are given by $\{\lambda_i \mu_j, i = 1, \dots, n; j = 1, \dots, m\}$ and the corresponding eigenvectors are given by $\mathbf{x}_i \otimes \mathbf{y}_j, i = 1, \dots, n; j = 1, \dots, m$. \square

Based on the above results, we now proceed to prove Theorem 2. Our main idea lies in showing that under condition (18), condition (12) in Lemma 3 will hold. To this end, we will employ the mean-value theorem for complex matrix-valued functions stated in [31]. That is, for any two different points $\mathbf{Q}_{-k}^{(1)}$ and $\mathbf{Q}_{-k}^{(2)}$, there exists some $t \in (0, 1)$ such that

$$\left\| \mathbf{F}_k(\mathbf{Q}_{-k}^{(1)}) - \mathbf{F}_k(\mathbf{Q}_{-k}^{(2)}) \right\|_F \leq \left\| \mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{F}_k(\Delta) \right\|_2 \left\| \mathbf{Q}_{-k}^{(1)} - \mathbf{Q}_{-k}^{(2)} \right\|_F, \quad (\text{C.1})$$

where $\Delta = t\mathbf{Q}_{-k}^{(1)} + (1-t)\mathbf{Q}_{-k}^{(2)}$ and $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{F}_k(\Delta)$ represents the Jacobian matrix of function $\mathbf{F}_k(\mathbf{Q}_{-k})$ w.r.t. \mathbf{Q}_{-k} , evaluated at $\mathbf{Q}_{-k} = \Delta$ [39].

Let $\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta)$ be the Jacobian matrix of function $\mathbf{F}_k(\mathbf{Q}_{-k})$ w.r.t. \mathbf{R}_k , evaluated at $\mathbf{R}_k = \mathbf{R}_k(\Delta)$. Let $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta)$ be the Jacobian matrix of function $\mathbf{R}_k(\mathbf{Q}_{-k})$ w.r.t. \mathbf{Q}_{-k} , evaluated at $\mathbf{Q}_{-k} = \Delta$. Then, we have:

$$\left\| \mathbf{F}_k(\mathbf{Q}_{-k}^{(1)}) - \mathbf{F}_k(\mathbf{Q}_{-k}^{(2)}) \right\|_F \leq \left\| \mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta) \cdot \mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta) \right\|_2 \left\| \mathbf{Q}_{-k}^{(1)} - \mathbf{Q}_{-k}^{(2)} \right\|_F \quad (\text{C.2})$$

$$\leq \left\| \mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta) \right\|_2 \left\| \mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta) \right\|_2 \left\| \mathbf{Q}_{-k}^{(1)} - \mathbf{Q}_{-k}^{(2)} \right\|_F, \quad (\text{C.3})$$

where (C.2) follows from the chain rule and (C.3) follows from Lemma 6.

Our goal is to obtain the upper bound of $\left\| \mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta) \right\|_2 \left\| \mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta) \right\|_2$. If this upper bound is less than $\sqrt{1/(K-1)}$, then Condition (12) is satisfied. In the following, we derive the upper bounds of $\left\| \mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta) \right\|_2$ and $\left\| \mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta) \right\|_2$, respectively.

A. The upper bound of $\left\| \mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta) \right\|_2$

To derive the upper bound of $\left\| \mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta) \right\|_2$, we should obtain the expression of $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta)$. To this end, the expression of the Jacobian matrix $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k$, which is a function of \mathbf{Q}_{-k} , should be obtained firstly. Then $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta)$ can be obtained by inserting $\mathbf{Q}_{-k} = \Delta$ into $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k$.

The Jacobian matrix $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k$ can be computed using the three-step procedure in [39]. Specifically, we compute firstly the differential of \mathbf{R}_k and then the Jacobian matrix. Function \mathbf{R}_k is differentiable at \mathbf{Q}_{-k} , with differential given by $d\mathbf{R}_k = \sum_{j \neq k} \mathbf{H}_{j,k} d\mathbf{Q}_j \mathbf{H}_{j,k}^H$. By vectorizing $d\mathbf{R}_k$ and using the equality $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B})$ [38], we obtain

$$d\text{vec} \mathbf{R}_k = [\mathbf{H}_{1,k}^* \otimes \mathbf{H}_{1,k}, \dots, \mathbf{H}_{k-1,k}^* \otimes \mathbf{H}_{k-1,k}, \mathbf{H}_{k+1,k}^* \otimes \mathbf{H}_{k+1,k}, \dots, \mathbf{H}_{K,k}^* \otimes \mathbf{H}_{K,k}] d\text{vec}(\mathbf{Q}_{-k}). \quad (\text{C.4})$$

Using the identification rule in [39], we can obtain $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k$ given in (17) in Theorem 2. Note that $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k$ does not depend on \mathbf{Q}_{-k} . Hence, we have $\mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k(\Delta) = \mathbf{D}_{\mathbf{Q}_{-k}} \mathbf{R}_k$.

B. The upper bound of $\|\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta)\|_2$

Similarly, to obtain the upper bound of $\|\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta)\|_2$, we should obtain the expression of $\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k$, which is a function of \mathbf{Q}_{-k} . Then the expression of $\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta)$ can be obtained by inserting $\mathbf{Q}_{-k} = \Delta$ into it.

According to Lemma 1, the EE function $\text{EE}_k(\mathbf{Q}_k)$ is strictly quasiconcave on $\mathbb{S}_+^{M \times M}$ since $\mathbf{H}_{k,k}$ is assumed to be full column rank. Also, the maximum transmit power is assumed to be very large. Then, for given \mathbf{Q}_{-k} (and thus \mathbf{R}_k), the gradient of $\text{EE}_k(\mathbf{Q}_k)$ w.r.t. \mathbf{Q}_k , evaluated at $\mathbf{Q}_k = \mathbf{F}_k$, must be zero [50]. Thus, using $\partial |\mathbf{BXC}| / \partial \mathbf{X} = |\mathbf{BXC}| [\mathbf{C}(\mathbf{BXC})^{-1} \mathbf{B}]^T$ [51], we have

$$\mathbf{G}(\mathbf{R}_k, \mathbf{F}_k) \triangleq \mathbf{H}_{k,k}^H (\mathbf{R}_k + \mathbf{H}_{k,k} \mathbf{F}_k \mathbf{H}_{k,k}^H)^{-1} \mathbf{H}_{k,k} - \alpha_k \mathbf{I} = \mathbf{0}. \quad (\text{C.5})$$

where $\alpha_k = C_k(\mathbf{Q}_k, \mathbf{Q}_{-k}) / (P_k + P_C)$. Since function $\text{EE}_k(\mathbf{Q}_k)$ is strictly quasiconcave in \mathbf{Q}_k for given \mathbf{R}_k , there exists a unique globally optimal solution \mathbf{F}_k that satisfies (C.5) [50]. Hence, equation (C.5) defines an implicit function [52]. Taking the derivative of (C.5) w.r.t. \mathbf{R}_k and using the chain rule, we have ⁴

$$\mathbf{D}_{\mathbf{R}_k} \mathbf{G} + \mathbf{D}_{\mathbf{F}_k} \mathbf{G} \cdot \mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k = \mathbf{0}. \quad (\text{C.6})$$

Now we first obtain the Jacobian matrices $\mathbf{D}_{\mathbf{R}_k} \mathbf{G}$ and $\mathbf{D}_{\mathbf{F}_k} \mathbf{G}$, then $\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k$ can be solved from (C.6). We again use the three-step procedure in [39] to compute $\mathbf{D}_{\mathbf{R}_k} \mathbf{G}$ and $\mathbf{D}_{\mathbf{F}_k} \mathbf{G}$. Function \mathbf{G} is differentiable w.r.t. both \mathbf{R}_k and \mathbf{F}_k , with the differential given by

$$d\mathbf{G} = -\mathbf{H}_{k,k}^H \mathbf{J}_k d\mathbf{R}_k \mathbf{J}_k \mathbf{H}_{k,k} - \mathbf{H}_{k,k}^H \mathbf{J}_k \mathbf{H}_{k,k} d\mathbf{F}_k \mathbf{H}_{k,k}^H \mathbf{J}_k \mathbf{H}_{k,k}. \quad (\text{C.7})$$

⁴For simplicity, the dependency of function $\mathbf{G}(\mathbf{R}_k, \mathbf{F}_k)$ on both \mathbf{R}_k and \mathbf{F}_k is omitted.

where $\mathbf{J}_k = (\mathbf{R}_k + \mathbf{H}_{k,k} \mathbf{F}_k \mathbf{H}_{k,k}^H)^{-1}$ and we used $d\mathbf{X}^{-1} = -\mathbf{X}^{-1}d\mathbf{X}\mathbf{X}^{-1}$ [39]. Then by vectorizing both sides of (C.7) and using the identification rule in [39], we can obtain the Jacobian matrices $\mathbf{D}_{\mathbf{R}_k} \mathbf{G} = -\mathbf{U}_k$ and $\mathbf{D}_{\mathbf{F}_k} \mathbf{G} = -\mathbf{V}_k$, with $\mathbf{U}_k = \mathbf{H}_{k,k}^T \mathbf{J}_k^T \otimes \mathbf{H}_{k,k}^H \mathbf{J}_k$ and $\mathbf{V}_k = \mathbf{H}_{k,k}^T \mathbf{J}_k^T \mathbf{H}_{k,k}^* \otimes \mathbf{H}_{k,k}^H \mathbf{J}_k \mathbf{H}_{k,k}$. Note that \mathbf{V}_k is a Hermitian matrix. Since the channel matrix $\mathbf{H}_{k,k}$ is assumed to be full column rank and \mathbf{J}_k is nonsingular, $\mathbf{H}_{k,k}^H \mathbf{J}_k \mathbf{H}_{k,k}$ is nonsingular. Then \mathbf{V}_k is also nonsingular. This is because [51]

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}, \text{ for all nonsingular } \mathbf{A}, \mathbf{B}. \quad (\text{C.8})$$

Hence, $\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k$ can be solved from (C.6):

$$\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k = -\mathbf{V}_k^{-1} \mathbf{U}_k. \quad (\text{C.9})$$

Thus, $\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta)$ can be obtained by inserting $\mathbf{Q}_{-k} = \Delta$ into (C.9):

$$\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta) = -\mathbf{V}_k^{-1}(\Delta) \mathbf{U}_k(\Delta). \quad (\text{C.10})$$

Then, we have

$$\|\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta)\|_2^2 = \rho(\mathbf{U}_k^H(\Delta) \mathbf{V}_k^{-2}(\Delta) \mathbf{U}_k(\Delta)) \leq \rho(\mathbf{V}_k^{-2}(\Delta)) \rho(\mathbf{U}_k(\Delta) \mathbf{U}_k^H(\Delta)), \quad (\text{C.11})$$

where the inequality follows from Lemma 7. We now aim to obtain the upper bound of $\rho(\mathbf{V}_k^{-2}(\Delta))$ and $\rho(\mathbf{U}_k(\Delta) \mathbf{U}_k^H(\Delta))$, respectively.

1) *The upper bound of $\rho(\mathbf{V}_k^{-2}(\Delta))$* : The upper bound of $\rho(\mathbf{V}_k^{-2}(\Delta))$ can be obtained as follows:

$$\rho(\mathbf{V}_k^{-2}(\Delta)) \quad (\text{C.12})$$

$$= \rho\left(\left(\mathbf{H}_{k,k}^T \mathbf{J}_k^T(\Delta) \mathbf{H}_{k,k}^* \otimes \mathbf{H}_{k,k}^H \mathbf{J}_k(\Delta) \mathbf{H}_{k,k}\right)^{-2}\right) \quad (\text{C.13})$$

$$= \rho\left(\left(\mathbf{H}_{k,k}^T \mathbf{J}_k^T(\Delta) \mathbf{H}_{k,k}^*\right)^{-2} \otimes \left(\mathbf{H}_{k,k}^H \mathbf{J}_k(\Delta) \mathbf{H}_{k,k}\right)^{-2}\right) \quad (\text{C.14})$$

$$= \left(\rho\left(\left(\mathbf{H}_{k,k}^H \mathbf{J}_k(\Delta) \mathbf{H}_{k,k}\right)^{-2}\right)\right)^2 \quad (\text{C.15})$$

$$= \left(\lambda_{\min}\left(\mathbf{H}_{k,k}^H \mathbf{J}_k(\Delta) \mathbf{H}_{k,k}\right)\right)^{-4} \quad (\text{C.16})$$

$$\leq \left(\lambda_{\min}\left(\mathbf{H}_{k,k}^H \left(\mathbf{I} + P_T \sum_{i=1}^K \mathbf{H}_{i,k} \mathbf{H}_{i,k}^H\right)^{-1} \mathbf{H}_{k,k}\right)\right)^{-4} \quad (\text{C.17})$$

where $\mathbf{J}_k(\Delta)$ is obtained by inserting $\mathbf{Q}_{-k} = \Delta$ into \mathbf{J}_k , (C.14) follows from (C.8) and the equality $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ [51], (C.15) follows from Lemma 8, (C.16) results

from the fact that $\mathbf{H}_k^H \mathbf{J}_k(\Delta) \mathbf{H}_k \in \mathbb{S}_{++}^{M \times M}$, the last inequality results from $\lambda_{\min}(\mathbf{A}^H \mathbf{B} \mathbf{A}) \geq \lambda_{\min}(\mathbf{A}^H \mathbf{C} \mathbf{A})$ for all $\mathbf{B} \succeq \mathbf{C} \succeq \mathbf{0}$ and the following relations

$$\mathbf{J}_k = \left(\mathbf{I} + \sum_{i \neq k} \mathbf{H}_{i,k} \Delta \mathbf{H}_{i,k}^H + \mathbf{H}_{k,k} \mathbf{F}_k(\Delta) \mathbf{H}_{k,k}^H \right)^{-1} \quad (\text{C.18})$$

$$\succeq \left(\mathbf{I} + \text{tr}(\mathbf{F}_k(\Delta)) \mathbf{H}_{k,k} \mathbf{H}_{k,k}^H + \sum_{i \neq k} \text{tr}(\Delta) \mathbf{H}_{i,k} \mathbf{H}_{i,k}^H \right)^{-1} \quad (\text{C.19})$$

$$\succeq \left(\mathbf{I} + P_T \sum_{i=1}^K \mathbf{H}_{i,k} \mathbf{H}_{i,k}^H \right)^{-1}. \quad (\text{C.20})$$

2) *The upper bound of $\rho(\mathbf{U}_k(\Delta) \mathbf{U}_k^H(\Delta))$:* The upper bound of $\rho(\mathbf{U}_k(\Delta) \mathbf{U}_k^H(\Delta))$ can be obtained as follows:

$$\rho(\mathbf{U}_k(\Delta) \mathbf{U}_k^H(\Delta)) = \rho((\mathbf{H}_{k,k}^T \mathbf{J}_k^T(\Delta)) \otimes \mathbf{H}_{k,k}^H \mathbf{J}_k(\Delta)) (\mathbf{J}_k^*(\Delta)) \mathbf{H}_{k,k}^* \otimes \mathbf{J}_k^H(\Delta) \mathbf{H}_{k,k}) \quad (\text{C.21})$$

$$= \rho(\mathbf{H}_{k,k}^T \mathbf{J}_k^T(\Delta) \mathbf{J}_k^*(\Delta) \mathbf{H}_{k,k}^* \otimes \mathbf{H}_{k,k}^H \mathbf{J}_k(\Delta) \mathbf{J}_k^H(\Delta) \mathbf{H}_{k,k}) \quad (\text{C.22})$$

$$= (\rho(\mathbf{H}_{k,k}^H \mathbf{J}_k^2(\Delta) \mathbf{H}_{k,k}))^2 \quad (\text{C.23})$$

$$\leq (\rho(\mathbf{H}_{k,k}^H \mathbf{H}_{k,k}))^2 \quad (\text{C.24})$$

where in (C.22) we use $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$ [51], (C.23) follows from Lemma 8, and (C.24) follows from $\mathbf{J}_k(\Delta) \preceq \mathbf{I}$.

Combining (C.11) with (C.17) and (C.24), we obtain

$$\|\mathbf{D}_{\mathbf{R}_k} \mathbf{F}_k(\Delta)\|_2 \leq \frac{\rho(\mathbf{H}_{k,k}^H \mathbf{H}_{k,k})}{(\lambda_{\min}(\mathbf{H}_{k,k}^H \mathbf{T}_k \mathbf{H}_{k,k}))^2} \quad (\text{C.25})$$

Finally, by combining (C.3) with (17) and (C.25), we have

$$\left\| \mathbf{F}_k(\mathbf{Q}_{-k}^{(1)}) - \mathbf{F}_k(\mathbf{Q}_{-k}^{(2)}) \right\|_F \leq \alpha_k \left\| \mathbf{Q}_{-k}^{(1)} - \mathbf{Q}_{-k}^{(2)} \right\|_F \quad (\text{C.26})$$

where α_k is given in (16). Hence, if condition (18) is satisfied, condition (12) in Lemma 3 holds, which completes the proof.

APPENDIX D

PROOF OF THEOREM 3

Before proving the theorem, we introduce some basic definitions that will be used in our derivations. Given the multiuser mapping $\mathbf{F}(\mathbf{Q})$ defined in (9), we introduce the following block-

maximum norm on $\mathbb{C}^{KM \times KM}$, defined as [53]

$$\|\mathbf{F}(\mathbf{Q})\|_{F,\text{block}} \triangleq \max_{k \in \Psi_K} \|\mathbf{F}_k(\mathbf{Q}_{-k})\|_F. \quad (\text{D.1})$$

Let $\|\cdot\|_{\infty,\text{vec}}$ be the vector maximum norm, defined as [38]

$$\|\mathbf{x}\|_{\infty,\text{vec}} \triangleq \max_{k \in \Psi_K} |x_k|, \mathbf{x} \in \mathbb{R}^K, \quad (\text{D.2})$$

and let $\|\cdot\|_{\infty,\text{mat}}$ denote the matrix norm induced by $\|\cdot\|_{\infty,\text{vec}}$, given by [38]:

$$\|\mathbf{A}\|_{\infty,\text{mat}} \triangleq \max_k \sum_{r=1}^K |[\mathbf{A}]_{k,r}|, \mathbf{A} \in \mathbb{R}^{K \times K}. \quad (\text{D.3})$$

Finally, we introduce the nonnegative matrix $\mathbf{B} \in \mathbb{R}_+^{K \times K}$ defined as

$$[\mathbf{B}]_{k,r} \triangleq \begin{cases} \alpha_k^2, & \text{if } k \neq r, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{D.4})$$

Define $\alpha_{\max} \triangleq \max_{k \in \Psi_K} \alpha_k^2$ and $\beta \triangleq (K-1)\alpha_{\max} < 1$, then $\|\mathbf{B}\|_{\infty,\text{mat}}$ can be easily computed as

$$\|\mathbf{B}\|_{\infty,\text{mat}} = \beta. \quad (\text{D.5})$$

Based on the above results, we then give the contraction property of the multiuser mapping function in the following lemma, which will be used in the proof of the theorem.

Lemma 9: If the uniqueness condition in (18) in Theorem 2 is satisfied, then the multiuser mapping function $\mathbf{F}(\mathbf{Q})$ defined in (9) is Lipschitz continuous on \mathcal{W} :

$$\|\mathbf{F}(\mathbf{Q}^{(1)}) - \mathbf{F}(\mathbf{Q}^{(2)})\|_{F,\text{block}} \leq \sqrt{\beta} \|\mathbf{Q}^{(1)} - \mathbf{Q}^{(2)}\|_{F,\text{block}} \quad (\text{D.6})$$

$\forall \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)} \in \mathcal{W}$, where $\|\cdot\|_{F,\text{block}}$ is defined in (D.1). Furthermore, the mapping is a block-contraction with modulus $\sqrt{\beta} < 1$.

Proof: Given $\mathbf{Q}^{(1)} = (\mathbf{Q}_1^{(1)}, \dots, \mathbf{Q}_K^{(1)}) \in \mathcal{W}$ and $\mathbf{Q}^{(2)} = (\mathbf{Q}_1^{(2)}, \dots, \mathbf{Q}_K^{(2)}) \in \mathcal{W}$, define, for each $k \in \Psi_K$, $e_{F_k} \triangleq \left\| \mathbf{F}_k(\mathbf{Q}_{-k}^{(1)}) - \mathbf{F}_k(\mathbf{Q}_{-k}^{(2)}) \right\|_F^2$ and $e_k \triangleq \left\| \mathbf{Q}_k^{(1)} - \mathbf{Q}_k^{(2)} \right\|_F^2$. Then we have

$$e_{F_k} = \left\| \mathbf{F}_k(\mathbf{Q}_{-k}^{(1)}) - \mathbf{F}_k(\mathbf{Q}_{-k}^{(2)}) \right\|_F^2 \leq \alpha_k^2 \left\| \mathbf{Q}_{-k}^{(1)} - \mathbf{Q}_{-k}^{(2)} \right\|_F^2 = \sum_{j \neq k} [\mathbf{B}]_{k,j} \left\| \mathbf{Q}_j^{(1)} - \mathbf{Q}_j^{(2)} \right\|_F^2 = \sum_{j \neq k} [\mathbf{B}]_{k,j} e_j \quad (\text{D.7})$$

$\forall \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)} \in \mathcal{W}$ and $\forall k \in \Psi_K$.

Introducing the vectors $\mathbf{e}_F \triangleq [e_{F_1}, \dots, e_{F_K}]^T$, and $\mathbf{e} \triangleq [e_1, \dots, e_K]^T$, the set of inequalities in (D.7) can be rewritten as:

$$\mathbf{0} \leq \mathbf{e}_F \leq \mathbf{B}\mathbf{e}, \forall \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)} \in \mathcal{W}. \quad (\text{D.8})$$

Then we have [53]

$$\|\mathbf{e}_F\|_{\infty, \text{vec}} \leq \|\mathbf{B}\mathbf{e}\|_{\infty, \text{vec}} \leq \|\mathbf{B}\|_{\infty, \text{mat}} \|\mathbf{e}\|_{\infty, \text{vec}} = \beta \|\mathbf{e}\|_{\infty, \text{vec}} \quad (\text{D.9})$$

where $\|\cdot\|_{\infty, \text{vec}}$ and $\|\cdot\|_{\infty, \text{mat}}$ are defined in (D.2) and (D.3), respectively, and the last equality follows from (D.5).

Finally, using (D.9) and (D.1), one obtains

$$\|\mathbf{F}(\mathbf{Q}^{(1)}) - \mathbf{F}(\mathbf{Q}^{(2)})\|_{F, \text{block}}^2 = \|\mathbf{e}_F\|_{\infty, \text{vec}} \leq \beta \|\mathbf{Q}^{(1)} - \mathbf{Q}^{(2)}\|_{F, \text{block}}^2 \quad (\text{D.10})$$

$\forall \mathbf{Q}^{(1)}, \mathbf{Q}^{(2)} \in \mathcal{W}$. Hence, the lemma follows. \square

Interestingly, we find that if condition (18) in Theorem 2 is satisfied, the multiuser mapping function $\mathbf{F}(\mathbf{Q})$ is not only a block-contraction for the Frobenius norm as shown in Lemma 2 but also block-maximum norm as shown in Lemma 9.

The remaining task is to show that, under condition (18) in Theorem 2, conditions of the asynchronous convergence theorem in Prop.2.1 of [40] on page 431 are satisfied. Based on the above lemma, the proof can be done in a similar fashion to the proof for the rate maximization game in [31]. For this reason, the details are omitted for simplicity.

APPENDIX E

PROOF OF THEOREM 4

Since the links are located far away, the interference among the links is negligible so that the covariance matrix of each link k (i.e., \mathbf{R}_k) can be regarded as an identity matrix \mathbf{I} . Hence, the SE function of links k reduces to

$$C_k(P_k) = \max_{\text{tr}(\mathbf{Q}_k) = P_k} \log_2 |\mathbf{I} + \mathbf{H}_{k,k}^H \mathbf{H}_{k,k} \mathbf{Q}_k|. \quad (\text{E.1})$$

Obviously, the optimal solution of \mathbf{Q}_k is $\mathbf{Q}_k^* = \tilde{\mathbf{U}}_k \mathbf{P}_k \tilde{\mathbf{U}}_k^H$, where $\tilde{\mathbf{U}}_k$ is the eigenvector matrix of $\mathbf{H}_{k,k}^H \mathbf{H}_{k,k}$, and \mathbf{P}_k is nonnegative matrix with $\mathbf{P}_k = \text{diag}\{p_{k,1}, \dots, p_{k,r_k}\}$, which is the power allocation over the streams. By inserting \mathbf{Q}_k^* into (E.1), the SE function can be equivalently obtained as

$$\begin{aligned} C_k(P_k) &= \max_{\{p_{k,m}\}} \sum_{m=1}^{r_k} \log_2(1 + d_{k,m} p_{k,m}) \\ &\text{s.t. } \sum_{m=1}^{r_k} p_{k,m} \leq P_k, p_{k,m} \geq 0. \end{aligned} \quad (\text{E.2})$$

The solution to the above problem can be easily solved as $p_{k,m} = (\mu_k - d_{k,m}^{-1})^+$, $m \in \{1, 2, \dots, r_k\}$, where μ_k is the water-level chosen to satisfy $\sum_{m=1}^{r_k} p_{k,m} = P_k$, and $d_{k,m}$'s are the eigenvalues of $\mathbf{H}_{k,k}^H \mathbf{H}_{k,k}$, which are arranged in decreasing order.

Define $g_l = l d_{k,l}^{-1} - \sum_{m=1}^l d_{k,m}^{-1}$ and $g_{r_k+1} = \infty$. Obviously, $g_1 = 0$. Then, the SE function can be rewritten as

$$C_k(P_k) = \sum_{m=1}^l \log_2 \left(\frac{d_{k,m}}{l} \left(P_k + \sum_{m=1}^l d_{k,m}^{-1} \right) \right), \quad P_k \in [g_l, g_{l+1}], l = 1, \dots, r_k. \quad (\text{E.3})$$

Note that l represents the number of streams that are allocated with positive power allocation.

Based on the above analysis, we now provide the properties of the SE function in the following lemma.

Lemma 10: The SE function $C_k(P_k)$ is a continuous, strictly increasing, differential and strictly concave function of P_k . Moreover, the first derivative of function $C_k(P_k)$ (denoted as $C_k'(P_k)$) is a continuous function of P_k . Also, the SE function $C_k(P_k)$ is twice differentiable in each interval of the subregions $[g_l, g_{l+1}]$, $l = 1, \dots, r_k$, which is given by

$$C_k''(P_k) = -\frac{l}{\ln 2} \left(P_k + \sum_{m=1}^l d_{k,m}^{-1} \right)^{-2} < 0, \quad P_k \in (g_l, g_{l+1}), l = 1, \dots, r_k. \quad (\text{E.4})$$

Proof: Obviously, in each interval of the subregions, the SE function $C_k(P_k)$ is a continuous, strictly increasing, differential and concave function of P_k . We only need to verify the boundary points. For the boundary points, we have

$$\lim_{P_k \rightarrow g_{l+1}^-} C_k(P_k) = \sum_{m=1}^l \log_2 \left(\frac{d_{k,m}}{d_{k,l+1}} \right) = \lim_{P_k \rightarrow g_{l+1}^+} C_k(P_k), \quad (\text{E.5})$$

$$\lim_{P_k \rightarrow g_{l+1}^-} C_k'(P_k) = \frac{d_{k,l+1}}{\ln 2} = \lim_{P_k \rightarrow g_{l+1}^+} C_k'(P_k) > 0. \quad (\text{E.6})$$

Equation (E.5) means that the SE function $C_k(P_k)$ is a continuous function and equation (E.6) means that the SE function is a strictly increasing and differential function.

Unfortunately, the SE function is not twice differentiable since

$$\lim_{P_k \rightarrow g_{l+1}^-} C_k''(P_k) = -\frac{1}{\ln 2} \frac{d_{k,l+1}^2}{l} \neq -\frac{1}{\ln 2} \frac{d_{k,l+1}^2}{l+1} = \lim_{P_k \rightarrow g_{l+1}^+} C_k''(P_k). \quad (\text{E.7})$$

Hence, the concavity of the SE function cannot be proved by using the above method. To overcome this issue, we use the duality theory to prove the concavity of the SE function. Specifically, for a given P_k , the SE function $C_k(P_k)$ can be obtained by solving problem (E.2),

which is a strictly convex optimization problem. It can also be verified that the constraints in problem (E.2) satisfy the Slater's condition. Hence, the duality gap for this problem is zero. Thus, the SE function can also be obtained by solving the following min-max problem

$$C_k(P_k) = \min_{\mu \geq 0} \max_{p_{k,m} \geq 0, \forall m} \sum_{m=1}^{r_k} \log_2(1 + d_{k,m} p_{k,m}) - \mu \left(\sum_{m=1}^{r_k} p_{k,m} - P_k \right) \quad (\text{E.8})$$

$$= \min_{\mu \geq 0} \sum_{m=1}^{r_k} \left[\log_2 \left(\frac{d_{k,m}}{\mu \ln 2} \right) \right]^+ - \sum_{m=1}^{r_k} \left[\frac{1}{\ln 2} - \frac{\mu}{d_{k,m}} \right]^+ + \mu P_k \quad (\text{E.9})$$

$$= \sum_{m=1}^{r_k} \left[\log_2 \left(\frac{d_{k,m}}{\mu(P_k) \ln 2} \right) \right]^+ - \sum_{m=1}^{r_k} \left[\frac{1}{\ln 2} - \frac{\mu(P_k)}{d_{k,m}} \right]^+ + \mu(P_k) P_k, \quad (\text{E.10})$$

where $\mu(P_k) \geq 0$ denotes the optimal value for given P_k , which is unique due to the fact that problem (E.3) is a strictly convex problem.

For any two points $P_k^{(1)}, P_k^{(2)}$, define $P_k^{(3)} = \theta P_k^{(1)} + (1 - \theta) P_k^{(2)}$, where $\theta \in (0, 1)$. Let $\mu(P_k^{(1)})$, $\mu(P_k^{(2)})$ and $\mu(P_k^{(3)})$ be the optimal μ for $C_k(P_k^{(1)})$, $C_k(P_k^{(2)})$ and $C_k(P_k^{(3)})$, respectively.

Then, for $i = 1, 2$, we have

$$C_k(P_k^i) = \sum_{m=1}^{r_k} \left[\log_2 \left(\frac{d_{k,m}}{\mu(P_k^i) \ln 2} \right) \right]^+ - \sum_{m=1}^{r_k} \left[\frac{1}{\ln 2} - \frac{\mu(P_k^i)}{d_{k,m}} \right]^+ + \mu(P_k^i) P_k^i \quad (\text{E.11})$$

$$< \sum_{m=1}^{r_k} \left[\log_2 \left(\frac{d_{k,m}}{\mu(P_k^3) \ln 2} \right) \right]^+ - \sum_{m=1}^{r_k} \left[\frac{1}{\ln 2} - \frac{\mu(P_k^3)}{d_{k,m}} \right]^+ + \mu(P_k^3) P_k^i, \quad (\text{E.12})$$

where the strict inequality follows due to the facts that $\mu(P_k^3)$ is not the optimal solution to problem (E.9) for given P_k^i , $i = 1, 2$, and problem (E.2) has a unique global solution since it is a strictly convex optimization problem [48]. By using the above inequality, we have

$$\theta C_k(P_k^1) + (1 - \theta) C_k(P_k^2) < C_k(P_k^3) = C_k(\theta P_k^1 + (1 - \theta) P_k^2). \quad (\text{E.13})$$

Hence, the SE function is also a strictly concave function of P_k . Obviously, the SE function $C_k(P_k)$ is twice differentiable in each interval of the subregions $[g_l, g_{l+1}]$, $l = 1, \dots, r_k$ and the second-order derivative of $C_k(P_k)$ w.r.t. P_k can be easily calculated in (E.4). \square

We now proceed to prove the first part of Theorem 4: the corresponding SE increases with the circuit power consumption. The EE function can be rewritten as

$$\text{EE}_k = \frac{C_k(P_k)}{P_k + P_C}. \quad (\text{E.14})$$

According to Lemma 10, $C_k(P_k)$ is a strictly concave function of P_k . Moreover, the denominator of the EE function is affine in P_k . Hence, EE_k is strictly quasiconcave in P_k [48]. Then, the

optimal solution of P_k (denoted as P_k^*) to maximize the EE function is unique [54] and should satisfy the first order condition [48]: $C'_k(P_k^*)(P_k^* + P_C) - C_k(P_k^*) = 0$, which is equivalent to

$$\frac{C_k(P_k^*) - P_k^* C'_k(P_k^*)}{C'_k(P_k^*)} = P_C. \quad (\text{E.15})$$

Our task is to analyze the monotonicity of P_k^* w.r.t. P_C . If P_k^* is strictly monotonically increasing w.r.t. P_C , the corresponding SE is monotonically increasing w.r.t. P_C as well since the SE is increasing in P_k based on Lemma 10. However, directly proving the monotonicity of P_k^* w.r.t. P_C is difficult. Instead, if the following two conditions hold: 1) P_C is a continuous function of P_k^* ; 2) there is a one-to-one mapping relation between P_k^* and P_C , this proof can be equivalently transformed into the opposite side [55]: P_C is strictly monotonically increasing w.r.t. P_k^* . Obviously, P_C is a continuous function of P_k^* since both $C_k(P_k^*)$ and $C'_k(P_k^*)$ are continuous function of P_k^* according to Lemma 10. The one-to-one mapping relation between P_k^* and P_C is also obvious: For given P_C , there is a unique P_k^* satisfying (E.15) since EE_k is strictly quasiconcave in P_k [54]; for given P_k^* , the unique P_C can be calculated from the left hand side of (E.15).

The remaining task is to prove that P_C is strictly increasing with P_k^* . Define the left hand side of (E.15) as $f(P_k^*)$. Function $f(P_k^*)$ is obviously differentiable in each interval of the subregions, i.e., $[g_l, g_{l+1}]$, $l = 1, \dots, r_k$, which can be calculated as

$$f'(P_k^*) = -C''_k(P_k^*)C_k(P_k^*)(C'_k(P_k^*))^{-2} > 0, P_k \in (g_l, g_{l+1}), l = 1, \dots, r_k. \quad (\text{E.16})$$

where the inequality follows from the facts that $C_k(P_k^*)$ is positive and $C''_k(P_k^*)$ is negative according to (E.4) in Lemma 10. Combining with the fact that $f(P_k^*)$ is a continuous function of P_k^* , we conclude that function $f(P_k^*)$ is a strictly increasing function w.r.t. P_k^* and the proof for the first part of Theorem 4 completes.

Finally, we prove the last part of Theorem 4: the EE decreases with the circuit power consumption. Supposing $P_C^{(1)} > P_C^{(2)}$, define $P_k^{(1)}$ and $P_k^{(2)}$ respectively as the optimal solutions for given $P_C^{(1)}$ and $P_C^{(2)}$. Then, we have

$$\text{EE}_k(P_C^{(1)}) = \frac{C_k(P_k^{(1)})}{P_k^{(1)} + P_C^{(1)}} < \frac{C_k(P_k^{(1)})}{P_k^{(1)} + P_C^{(2)}} < \frac{C_k(P_k^{(2)})}{P_k^{(2)} + P_C^{(2)}} = \text{EE}_k(P_C^{(2)}), \quad (\text{E.17})$$

where the second inequality follows from the assumption that $P_k^{(2)}$ is the optimal solution for given $P_C^{(2)}$. From the above inequalities, we know that the EE decreases with P_C .

REFERENCES

- [1] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - part I: Chunk allocation," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2734–2744, Sep. 2009.
- [2] —, "Chunk-based resource allocation in OFDMA systems - part II: Joint chunk, power and bit allocation," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 499–509, Sep. 2012.
- [3] G. Fettweis and E. Zimmermann, "ICT energy consumption-trends and challenges," in *Proceedings of the 11th International Symposium on Wireless Personal Multimedia Communications*, vol. 2, 2008, p. 6.
- [4] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.
- [5] C. Liang and K. Dandekar, "Power management in MIMO ad hoc networks: A game-theoretic approach," *IEEE Trans. Wireless Commun.*, vol. 6, no. 4, pp. 1164–1170, Apr. 2007.
- [6] K.-S. Chin, H.-T. Chang, J.-A. Liu, H.-C. Chiu, J. Fu, and S.-H. Chao, "28-GHz patch antenna arrays with PCB and LTCC substrates," in *Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC), 2011*, vol. 1, Jul. 2011, pp. 355–358.
- [7] S. H. Yeung, A. Garcia-Lamperez, T. Kumar Sarkar, and M. Salazar-Palma, "A thin and compact high gain planar antenna integrated with a CMRC compact filter," in *Wireless Symposium (IWS), 2014 IEEE International*, Mar. 2014, pp. 1–4.
- [8] R. Vaze and R. Heath, "Transmission capacity of ad-hoc networks with multiple antennas using transmit stream adaptation and interference cancellation," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 780–792, Feb. 2012.
- [9] S. Chu, X. Wang, and Y. Yang, "Adaptive scheduling in MIMO-based heterogeneous ad hoc networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 5, pp. 964–978, May 2014.
- [10] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.
- [11] D. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.
- [12] Y. Liu and L. Dong, "Spectrum sharing in MIMO cognitive radio networks based on cooperative game theory," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4807–4820, Sep 2014.
- [13] S.-J. Kim and G. B. Giannakis, "Optimal resource allocation for MIMO ad hoc cognitive radio networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3117–3131, May 2011.
- [14] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.
- [15] C. Jiang and L. Cimini, "Energy-efficient transmission for MIMO interference channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2988–2999, June 2013.
- [16] S. He, Y. Huang, L. Yang, and B. Ottersten, "Coordinated multicell multiuser precoding for maximizing weighted sum energy efficiency," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 741–751, Feb. 2014.
- [17] C. Pan, W. Xu, W. Zhang, J. Wang, H. Ren, and M. Chen, "Weighted sum energy efficiency maximization in ad hoc networks," *IEEE Wireless Communications Letters*, vol. 4, no. 3, pp. 233–236, June 2015.
- [18] F. Meshkati, H. V. Poor, and S. C. Schwartz, "Energy-efficient resource allocation in wireless networks," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 58–68, May 2007.

- [19] S. Betz and H. Poor, "Energy efficient communications in CDMA networks: A game theoretic analysis considering operating costs," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5181–5190, Oct. 2008.
- [20] G. Miao, N. Himayat, G. Y. Li, and S. Talwar, "Distributed interference-aware energy-efficient power optimization," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1323–1333, Apr. 2011.
- [21] A. Zappone, Z. Chong, E. Jorswieck, and S. Buzzi, "Energy-aware competitive power control in relay-assisted interference wireless networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1860–1871, Apr. 2013.
- [22] S. Buzzi, H. V. Poor, and D. Saturnino, "Energy-efficient resource allocation in multiuser MIMO systems: A game-theoretic framework," *arXiv preprint arXiv:0807.3094*, 2008.
- [23] A. Zappone, P. Cao, and E. Jorswieck, "Energy efficiency optimization in relay-assisted MIMO systems with perfect and statistical CSI," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 443–457, Jan. 2014.
- [24] S. Ye and R. Blum, "Optimized signaling for MIMO interference systems with feedback," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2839–2848, Nov. 2003.
- [25] D. N. Nguyen and M. Krunz, "Price-based joint beamforming and spectrum management in multi-antenna cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2295–2305, Dec. 2012.
- [26] D. Nguyen and T. Le-Ngoc, "Sum-Rate Maximization in the multicell MIMO multiple-access channel with interference coordination," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 36–48, Jan. 2014.
- [27] —, "Sum-rate maximization in the multicell MIMO broadcast channel with interference coordination," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1501–1513, Mar. 2014.
- [28] H. Baligh, M. Hong, W.-C. Liao, Z.-Q. Luo, M. Razaviyayn, M. Sanjabi, and R. Sun, "Cross-layer provision of future cellular networks: A WMMSE-based approach," *IEEE Signal Process. Mag.*, vol. 31, no. 6, pp. 56–68, Nov. 2014.
- [29] S. Lasaulce, A. Suárez, M. Debbah, and L. Cottarelli, "Power allocation game for fading MIMO multiple access channels with antenna correlation," in *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007, p. 1.
- [30] G. Scutari, D. P. Palomar, and S. Barbarossa, "Competitive design of multiuser MIMO systems based on game theory: A unified view," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1089–1103, Sep. 2008.
- [31] —, "The MIMO iterative waterfilling algorithm," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1917–1935, May 2009.
- [32] G. Scutari and D. P. Palomar, "MIMO cognitive radio: A game theoretical approach," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 761–780, Feb. 2010.
- [33] J. Wang, G. Scutari, and D. P. Palomar, "Robust MIMO cognitive radio via game theory," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1183–1201, Mar. 2011.
- [34] D. Nguyen and T. Le-Ngoc, "Multiuser downlink beamforming in multicell wireless systems: A game theoretical approach," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3326–3338, July 2011.
- [35] D. Nguyen, H. Nguyen-Le, and T. Le-Ngoc, "Block-Diagonalization precoding in a multiuser multicell MIMO system: Competition and coordination," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 968–981, Feb. 2014.
- [36] T. Zhou, Y. Chen, and K. R. Liu, "Network formation games in cooperative MIMO interference systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 1140–1152, Feb. 2014.
- [37] W. Saad, Z. Han, M. Debbah, A. Hjørungnes, and T. Basar, "Coalitional game theory for communication networks," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 77–97, Sep. 2009.
- [38] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge university press, 2012.

- [39] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2740–2746, June 2007.
- [40] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Prentice-Hall, Inc., 1989.
- [41] S. Schaible and T. Ibaraki, "Fractional programming," *European Journal of Operational Research*, vol. 12, no. 4, pp. 325–338, 1983.
- [42] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.
- [43] S. Assumptions, "Parameters for FDD HeNB RF requirements," in *3GPP TSG-RAN WG4 R4-092042*, 2009.
- [44] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *IEEE Future Network and Mobile Summit*, 2010, pp. 1–8.
- [45] J. Huang, R. Berry, and M. Honig, "Distributed interference compensation for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 5, pp. 1074–1084, May 2006.
- [46] B. Chen and M. Gans, "MIMO communications in ad hoc networks," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2773–2783, Jul. 2006.
- [47] I. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Commun. Mag.*, vol. 43, no. 9, pp. S23–S30, Sep. 2005.
- [48] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.
- [49] M. J. Osborne, *A Course in Game Theory*. Cambridge, Mass.: MIT Press, 1994.
- [50] E. Wolfstetter, *Topics in Microeconomics: Industrial Organization, Auctions, and Incentives*. Cambridge University Press, 1999.
- [51] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 1995.
- [52] K. Jittorntrum, "An implicit function theorem," *Journal of Optimization Theory and Applications*, vol. 25, no. 4, pp. 575–577, 1978.
- [53] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*. Siam, 2000.
- [54] E. Wolfstetter, *Topics in microeconomics: Industrial organization, auctions, and incentives*. Cambridge University Press, 1999.
- [55] T. M. Apostol, *Mathematical analysis*. Addison Wesley Publishing Company, 1974.