Networked MIMO With Fractional Joint Transmission in Energy Harvesting Systems

Jie Gong, *Member, IEEE*, Sheng Zhou, *Member, IEEE*, and Zhenyu Zhou, *Member, IEEE*

***Abstract*— This paper considers two base stations (BSs) powered by renewable energy serving two users cooperatively. With different BS energy arrival rates, a fractional joint transmission (JT) strategy is proposed, which divides each transmission frame into two subframes. In the first subframe, one BS keeps silent to store energy, while the other transmits data, and then, they perform zero-forcing JT (ZF-JT) in the second subframe. We consider the average sum-rate maximization problem by optimizing the energy allocation and the time fraction of ZF-JT separately. First, the sum-rate maximization for given energy budgets in each frame is analyzed. We prove that the optimal transmit power can be derived in closed form, and the optimal time fraction can be found via bi-section search. Second, an approximate dynamic programming algorithm is introduced to determine the energy allocation among frames. We adopt a linear approximation with the features associated with system states and determine the weights of features by simulation. We also operate the approximation several times with random initial policy, named policy exploration, to broaden the policy search range. Numerical results show that the proposed fractional JT greatly improves the performance. In addition, appropriate policy exploration is shown to perform close to the optimal.**

***Index Terms*— Energy harvesting, cooperative communication, dynamic programming, power control.**

I. Introduction

WIRELESS communication with energy harvesting technology, which exploits renewable energy to power wireless devices, is expected as one of the promising trends to meet the target of green communications in the future. The advantages of energy harvesting include the sustainability with renewable energy source, the flexibility of network deployment without power line, and etc. Recently, wireless cellular networks with renewable energy are rapidly developing. For instance, China Mobile has built about 12,000 renewable energy powered base stations (BSs) by 2014 [1]. However, due to the randomness of the arrival process of the renewable energy and the limitation on the battery capacity, energy shortage or waste will occur when the energy arrival mismatches with the network traffic requirement. How to efficiently use the harvested energy is a big challenge.

In the literature, a lot of research work has focused on the energy harvesting based communications. For single-link case, the optimal power allocation structure, *directional water-filling*, is found in both single-antenna transceiver system [2], [3] and multiple-input multiple-output (MIMO) channel [4]. The research efforts have been further extended to the network case, and the power allocation policies are proposed for broadcast channel [5], multiple access channel [6], interference channel [7], as well as cooperative relay networks [8], [9]. Nevertheless, there lacks research effort on the effect of energy harvesting on the multi-node cooperation, i.e., network MIMO.

The network MIMO technology, which shares the user data and channel state information among multiple BSs, and coordinates the data transmission and reception by transforming the inter-cell interference into useful signals, has been extensively studied in the literature [10]–[12]. And it has been standardized in 3GPP as Coordinated Multi-Point (CoMP) [13]. By applying joint precoding schemes such as zero-forcing (ZF) [14], [15] among BSs for joint transmission (JT), the system sum-rate can be greatly increased. However, how the dynamic energy arrival influences the performance of network MIMO requires further study. Specifically, as the JT is constrained by the per-BS power budget, the performance of the network MIMO is limited if the power budgets are severely asymmetric among BSs. For example, if a solar-powered BS in a windless sunny day cooperates with a wind-powered BS, the latter will become the bottleneck of cooperation, while the harvested energy of the former is not efficiently utilized. To deal with this problem, people have introduced the concept of energy cooperation [16], [17], where BSs can exchange energy via either wired or wireless link with some loss of energy transfer. In this case, the JT problem with energy harvesting becomes a power allocation problem with weighted sum power constraint as shown in [18]. However, the cooperation in energy domain strongly depends on the existence and the efficiency of energy transfer link.

In this paper, we consider how to improve the utilization of harvested energy with cooperation between the wireless radio links. Intuitively, if the energy cannot be transferred between BSs, the BS with higher energy arrival rate should

Manuscript received February 2, 2016; revised June 1, 2016; accepted July 2, 2016. Date of publication July 9, 2016; date of current version August 12, 2016. This work is sponsored in part by the Fundamental Research Funds for the Central Universities, the National Basic Research Program of China (973 Program: No.2012CB316001), the Nature Science Foundation of China (61571265, 61321061, 61461136004), and Hitachi R&D Headquarter. This work was presented at the IEEE International Conference on Communication System [19]. The associate editor coordinating the review of this paper and approving it for publication was W. Zhang.

J. Gong is with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China (e-mail: gongj26@mail.sysu.edu.cn).

S. Zhou is with the Tsinghua National Laboratory for Information Science and Technology, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: sheng.zhou@tsinghua.edu.cn).

Z. Zhou is with the State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China (e-mail: zhenyu_zhou@ncepu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2016.2589267

use more energy in data transmission to avoid energy waste. While to use the energy more effectively, BS cooperation strategy should be carefully designed under the asymmetric energy constraints. Based on this, we propose a *fractional JT* strategy, where the network MIMO is only applied in a fraction of a transmission frame. Specifically, we consider two BSs cooperatively serving two users, and divide each transmission frame into two subframes. In the first subframe, one of the BSs serves one user while the other stores energy. In the second subframe, the two BSs perform JT to cooperatively serve the two users. With the stored energy, the power gap between two BSs in the second subframe is filled, and hence, JT can achieve higher sum-rate. Such a strategy avoids the potential energy waste in the BS with higher energy arrival rate, and hence can improve the energy utilization. The objective is to maximize the average sum-rate for given energy arrival rates, and the optimization parameters include the fraction of time for JT and the power allocation policy in each frame. Our preliminary work [19] has studied the greedy policy that tries to use all the available energy in each frame. In this paper, we further design the optimal policy as well as the low-complexity policy. The contributions of this paper are as follows.

- We propose the fractional JT strategy, and formulate the long-term average sum-rate maximization problem using Markov decision process (MDP) [20]. The problem is divided into two sub-problems, i.e., energy management among frames, and fractional JT optimization problem in each frame.
- We prove that to solve the average sum-rate maximization problem, in each frame, we only need to solve a power allocation problem with equality power constraints, which has closed-form expressions. Then the JT time fraction optimization problem is proved to be convex, and a bi-section search algorithm is proposed to find the optimal JT time fraction.
- We adopt the *approximate dynamic programming* (DP) [20] algorithm to reduce the computational complexity of determining the energy allocation among frames. The algorithm runs iteratively with two steps: *policy evaluation* and *policy improvement*. In the policy evaluation, the relative utility function in the Bellman's equation is approximated as a weighted summation of a set of features associated with system states. The weights are estimated by simulation. In the policy improvement, random initial policies are periodically selected to rerun the iteration to broaden the search range. Numerical simulations show the remarkable performance gain compared with the conventional network MIMO.

The rest of the paper is organized as follows. Section II describes the system model and Section III describes the MDP problem formulation. In Section IV, the per-frame optimization problem is analyzed. Then the approximate DP algorithm is proposed in Section V. Simulation study is presented in Section VI. Finally, Section VII concludes the paper.

Notations: Bold upper case and lower case letters denote matrices and vectors, respectively. $|\cdot|$ denotes the absolute value of a scalar, and $[x]^+ = \max\{x, 0\}$. $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and transpose conjugate of a matrix, respectively. $\mathcal{R}^+$ is the non-negative real number field. $\mathbb{E}$ represents the expectation operation.

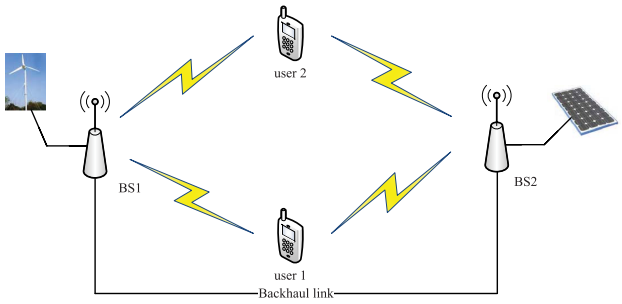

Fig. 1. System model for joint transmission with 2 BSs and 2 users.

II. System Model

We consider a wireless communication network consisting of two BSs powered by renewable energy (e.g., solar energy, wind energy, etc.) and two users as shown in Fig. 1. Assume the BSs are able to store the harvested energy in their battery for future usage. All the BSs and the users are equipped with a single antenna. The BSs are interconnected via an error-free backhaul link sharing all the data and the channel state information, so that they can perform JT to eliminate the interference. However, the energy cannot be transferred between the BSs as we consider the off-grid scenario. We consider the typical scenario for applying network MIMO, in which the two users are located at the cell boundary. In this case, the average channel gains are comparable, and hence cooperative transmission can achieve significant performance gain. The wireless channel is assumed block fading, i.e., the channel state is constant during each fading block, but changes from block to block. We define the transmission frame as a channel fading block with frame length T_f. The perfect channel state information is assumed known to the BSs at the beginning of each frame. If the backhaul capacity is limited, the two BSs can exchange quantized data and channel state information, and cooperate in the same way using the imperfect information.

In the t-th frame, if the JT technique is utilized, the received signals $\mathbf{y}_t = [y_{t,1}, y_{t,2}]^T$ at the users are

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{W}_t \mathbf{x}_t + \mathbf{n}_t, \tag{1}$$

where $\mathbf{H}_t$ is the channel matrix with components $H_{t,ik} = l_{ik}\tilde{H}_{t,ik}, 1 \le i, k \le 2$ indicating the channel coefficient from BS k to user i with large-scale fading factor l_{ik} and i.i.d. small-scale fading factor $\tilde{H}_{t,ik}$, $\mathbf{W}_t$ is the corresponding precoding matrix with components $w_{t,ki}$, $\mathbf{x}_t = [x_{i,1}, x_{t,2}]^T$ is the intended signals for the users with $\mathbb{E}(\mathbf{x}_t \mathbf{x}_t^H) = \text{diag}(p_{t,1}, p_{t,2})$, where $p_{t,i}, i = 1, 2$ is the power allocated to user i, and $\mathbf{n}_t$ is the additive white Gaussian noise with zero mean and variance $\mathbb{E}(\mathbf{n}_t \mathbf{n}_t^H) = \sigma_n^2 \mathbf{I}$, where $\mathbf{I}$ is a 2×2 unit matrix.

In this paper, the widely used ZF precoding scheme [14] is adopted to completely eliminate the interference by

channel inverse. The performance of ZF is sufficiently good, especially when the interference dominates the noise, and the decoding process at the users can be simplified. In addition, ZF precoding is a representative precoding scheme. The following analysis can be easily extended to other schemes. For ZF precoding scheme, we have

$$\mathbf{W}_t = \mathbf{H}_t^{-1}. \tag{2}$$

Hence, the data rate is

$$R_{t,i} = \log_2(1 + \frac{p_{t,i}}{\sigma_n^2}) \tag{3}$$

with per-BS power constraint

$$\sum_{i=1}^{2} |w_{t,ki}|^2 p_{t,i} \leq P_{t,k}, \quad k = 1, 2. \tag{4}$$

where $P_{t,k}$ is the maximum available transmit power of BS k in frame t. Notice that if the BSs and the users are equipped with multiple antennas, ZF precoding scheme should be replaced by the multi-cell block diagonalization (BD) [11] scheme which also nulls the inter-BS interference. As the multi-cell BD scheme is a generalization of ZF precoding scheme from single antenna case to multi-antenna case, it has similar mathematical properties with the latter. Hence, the following results can be extened to multi-antenna case.

As the BSs are powered by the renewable energy, $P_{t,k}$ is determined by the amount of harvested energy as well as the available energy in the battery. It is pointed out in [8] and [21] that in real systems, the energy harvesting rate changes in a much slower speed than the channel fading. Specifically, a fading block in current wireless communication systems is usually measured in the time scale of milliseconds, while the renewable energy such as solar power may keep constant for seconds or even minutes. Hence, the energy arrival rate (energy harvesting power) is assumed constant over a sufficient number of transmission frames, denoted by $E_k, k = 1, 2$. In this case, the key factor of the energy harvesting is the energy arrival causality constraint, i.e., the energy that has not arrived yet cannot be used in advance. In this paper, we mainly study the influence of the energy causality on the network MIMO.

Notice that in practice, the optimization over multiple energy coherence blocks is required as the energy arrival rate varies over time. If the future energy arrival information is unknown (i.e., purely random and unpredictable), we can monitor the energy harvesting rate and once it changes, we recalculate the optimal policy under the new energy constraint, and then apply the new policy. The policy optimization problem is considered in this paper. While if the energy arrival rate is predictable, the optimization should jointly consider multiple blocks in the prediction window, which is beyond the scope the this paper.

A. Fractional Joint Transmission Strategy

Notice that the energy arrival rates of different BSs may be different due to either utilizing various energy harvesting equipments (e.g., one with solar panel, the other with wind turbine) or encountering different environment conditions (e.g., partly cloudy). In this case, the conventional network MIMO may not be sum-rate optimal as the harvested energy is not efficiently utilized. Specifically, if the channel conditions of the two users are similar, applying network MIMO with on average the same energy usage can achieve the optimal cooperation performance. As a result, in the asymmetric energy arrival case, the energy of the BS with higher energy arrival rate may be not efficiently used. Hence, the performance of network MIMO may be greatly degraded. Notice that the above fact does not only hold for ZF precoding, but also holds for other precoding approaches (such as the approach based on dirty paper coding [22], [23]) as it is caused by the asymmetric per-BS power constraints, rather than the precoding scheme itself.

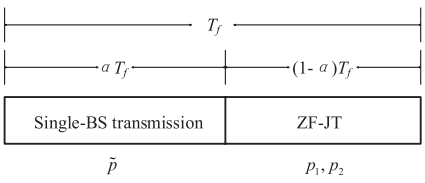

Fig. 2. Frame structure of fractional JT. The frame length is T_f.

To improve the utilization of the harvested energy, we propose a fractional JT strategy to adapt to the asymmetric energy arrival rates. Thanks to the energy storage ability, the BS can turn to sleep mode to store energy for a while, and then cooperatively transmits data with the other BS. In this way, it can provide higher transmit power when applying network MIMO. The strategy is detailed as follows. We divide the whole transmission frame into two subframes as shown in Fig. 2. In the first subframe, named as *single-BS transmission phase*, one of the BSs $k_t \in \{1, 2\}$ is selected to serve a user, while the other one, denoted by $\bar{k}_t \neq k_t$, turns to sleep mode to store energy. In the second subframe, named as *ZF-JT phase*, the two BSs jointly transmit to the two users with ZF precoding. Denote by $\alpha_t T_f$ the length of the single-BS transmission phase, where $0 \leq \alpha_t \leq 1$, and hence, the length of the ZF-JT phase is $(1 - \alpha_t)T_f$. To get the optimal fractional JT transmission strategy, we need to choose k_t and α_t carefully.

In the single-BS transmission phase, to be consistent with the objective of maximizing sum-rate, the active BS serves one of the users with higher instantaneous data rate. Specifically, the user $\tilde{i}$ is scheduled when satisfying $\tilde{i} = \arg\max_i \log_2(1 + \frac{\bar{P}|H_{t,ik_t}|^2}{\sigma_n^2})$, i.e., the user with the maximum expected data rate with transmit power $\bar{P} = E_{k_t}$. In practice, the proposed fractional JT transmission strategy can be supported by the CoMP [13], in which all the data is shared by the two BSs in both subframes. Notice that as only one BS is active in the first subframe, the data transferred to the inactive BS via the backhaul is useless, and such a backhaul data sharing protocol is inefficient.

However, when the backhaul capacity is limited, the proposed fractional JT strategy can make use of the backhaul capacity in the first subframe to enhance the performance. Since the shared data is required only in the second subframe,

the two BSs in the first subframe can proactively exchange the data to be jointly transmitted later. Thus, the quantization noise of the shared data can be reduced and the cooperation gain can be enhanced.

B. Sum-Rate Maximization Problem

Our objective is to optimize the sum-rate under the proposed fractional JT strategy. The power constraints in each frame are detailed as follows. The available energy in the battery of the active BS k_t at the beginning of each frame t is denoted by B_{t,k_t}. Then the power in the first subframe satisfies

$$\tilde{p}_t \le \frac{B_{t,k_t}}{\alpha_t T_f} + E_{k_t}. \tag{5}$$

At the beginning of the second subframe, the amounts of available battery energy in the two BSs become $B_{t,k_t} + \alpha_t T_f E_{k_t} - \alpha_t T_f \tilde{p}_t$ and $B_{i,\bar{k}_t} + \alpha_t T_f E_{\bar{k}_t}$, respectively. As a result, the power constraints (4) for ZF-JT become

$$\sum_{i=1}^{2} |w_{t,k_t i}|^2 p_{t,i} \le \frac{B_{t,k_t} + \alpha_t T_f (E_{k_t} - \tilde{p}_t)}{(1-\alpha_t) T_f} + E_{k_t}, \tag{6}$$

$$\sum_{i=1}^{2} |w_{t,\bar{k}_t i}|^2 p_{t,i} \le \frac{B_{t,\bar{k}_t} + \alpha_t T_f E_{\bar{k}_t}}{(1-\alpha_t) T_f} + E_{\bar{k}_t}. \tag{7}$$

The battery energy states are updated according to

$$B_{t+1,k_t} = B_{t,k_t} + T_f (E_{k_t} - \alpha_t \tilde{p}_t - (1-\alpha_t) \sum_{i=1}^{2} |w_{t,k_t i}|^2 p_{t,i}), \tag{8}$$

$$B_{t+1,\bar{k}_t} = B_{t,\bar{k}_t} + T_f (E_{\bar{k}_t} - (1-\alpha_t) \sum_{i=1}^{2} |w_{t,\bar{k}_t i}|^2 p_{t,i}), \tag{9}$$

with initial state $B_{1,1} = B_{1,2} = 0$. In (5), (6), and (7), we have $0 < \alpha_t < 1$ as the denominator cannot be zero. In fact, by multiplying α_t on both sides of (5) and $1 - \alpha_t$ on both sides of (6) and (7), the special case that $\alpha_t = 0$ or 1 can be included in a unified formulation. Denote by $\mathbf{k} = \{k_1, k_2, \cdots, k_N\}$, $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_N\}$, $\tilde{\mathbf{p}} = \{\tilde{p}_1, \tilde{p}_2, \cdots, \tilde{p}_N\}$, $\mathbf{p} = \{p_1, p_2, \cdots, p_N\}$, where $p_t = (p_{t,1}, p_{t,2})^T$, and N is the number of transmission frames. Our optimization problem can be formulated as

$$\max_{\mathbf{k},\alpha,\tilde{\mathbf{p}},\mathbf{p}} \lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N}\sum_{t=1}^{N}\left(\alpha_t \tilde{R}_{t,\tilde{i}} + (1-\alpha_t)\sum_{i=1}^{2} R_{t,i}\right)\right] \tag{10}$$

$$\text{s.t.} \quad \alpha_t \tilde{p}_t \le \frac{B_{t,k_t}}{T_f} + \alpha_t E_{k_t}, \tag{11}$$

$$(1-\alpha_t)\sum_{i=1}^{2} |w_{t,k_t i}|^2 p_{t,i} + \alpha_t \tilde{p}_t \le \frac{B_{t,k_t}}{T_f} + E_{k_t}, \tag{12}$$

$$(1-\alpha_t)\sum_{i=1}^{2} |w_{t,\bar{k}_t i}|^2 p_{t,i} \le \frac{B_{t,\bar{k}_t}}{T_f} + E_{\bar{k}_t}, \tag{13}$$

$$\tilde{p}_t, p_{t,1}, p_{t,2} \in \mathcal{R}^+, \qquad \forall t = 1, 2, \cdots, N. \tag{14}$$

$$0 \le \alpha_t \le 1, \tag{15}$$

where $\tilde{R}_{t,\tilde{i}} = \log_2(1 + \tilde{p}_t |H_{t,\tilde{i}k_t}|^2/\sigma_n^2)$, $R_{t,i}$ is expressed as (3), and the expectation is taken over all the possible channel matrix realizations. The optimization parameters include the transmit power $\tilde{p}_t$, $p_{t,k}$, $k = 1, 2$, the frame division parameter α_t, and the selection of BSs k_t for single-BS transmission phase. Notice that if $\alpha_t = 0$, the problem reduces to the conventional power allocation problem for network MIMO; if $\alpha_t = 1$, the problem becomes user selection and rate maximization problem for single-BS transmission. To find the optimal solution, we need to calculate the integration of the channel distribution over all the frames and exhaustively search all the possible power allocation and frame division policies, which is computationally overwhelming. In the work, we aim to design a low-complex algorithm to achieve close-to-optimal performance.

III. MDP Modeling and Optimization

In this section, we reformulate the stochastic optimization problem (10) based on the MDP framework [20]. Specifically, in each channel fading block, we need to decide which BS should turn to sleep to store energy in the first subframe, how long it should sleep, and how much power should be allocated. The decision in each frame will influence the decisions in the future, as it changes the remained energy in the battery. MDP is an effective mathematical framework to model such a time-correlated decision making problem. The formulation is detailed as follows.

A. MDP Problem Reformulation

A standard MDP problem contains the following elements: state, action, per-stage utility function and state transition. In our problem, the stage refers to the frame. In each stage, the system state includes the battery states of two BSs at the beginning of the frame and the channel states, i.e., $s_t = (B_{t,1}, B_{t,2}, \mathbf{H}_t)$. Denote the state space by $\mathcal{S}$. We model the action as the power budget of each frame, i.e., $a_t(s_t) = (A_{t,1}, A_{t,2})$ which satisfy $0 \le A_{t,1} \le \frac{B_{t,1}}{T_f} + E_1$ and $0 \le A_{t,2} \le \frac{B_{t,2}}{T_f} + E_2$. We denote the state-dependent action space by $\mathcal{A}(s_t) = \{(A_{t,1}, A_{t,2}) | 0 \le A_{t,1} \le \frac{B_{t,1}}{T_f} + E_1, 0 \le A_{t,2} \le \frac{B_{t,2}}{T_f} + E_2\}$. The per-stage sum-rate function can be expressed as

$$g(s_t, a_t) = \max_{k_t, \alpha_t, \tilde{p}_t, p_t} \alpha_t \log_2\left(1 + \frac{\tilde{p}_t |H_{t,\tilde{i}k_t}|^2}{\sigma_n^2}\right) + (1-\alpha_t)\sum_{i=1}^{2} \log_2\left(1 + \frac{p_{t,i}}{\sigma_n^2}\right), \tag{16}$$

where the maximization is taken under the constraints (11), (14), (15) and

$$(1-\alpha_t)\sum_{i=1}^{2} |w_{t,k_t i}|^2 p_{t,i} + \alpha_t \tilde{p}_t \le A_{t,k_t}, \tag{17}$$

$$(1-\alpha_t)\sum_{i=1}^{2} |w_{t,\bar{k}_t i}|^2 p_{t,i} \le A_{t,\bar{k}_t}, \tag{18}$$

The state transition of the battery energy is deterministic according to (8) and (9). The channel state of the next stage

is obtained according to the channel transition $\Pr(\mathbf{H}_{t+1}|\mathbf{H}_t)$, which is independent with the battery energy state.

Consequently, the original problem (10) can be reformulated as

$$\max_{\boldsymbol{a}} \ \lim_{N\to\infty} \mathbb{E}_{\mathbf{H}}\left[\frac{1}{N}\sum_{t=1}^{N} g(s_t, a_t(s_t))\right]. \tag{19}$$

The optimization is taken over all the possible policies $\boldsymbol{a} = \{a_1, a_2, \ldots\}$. It is obvious that for any two states, there is a stationary policy $\boldsymbol{a}$ so that one state can be accessed from the other with finite steps [20, Sec. 4.2]. Consequently, the optimal value is independent of the initial state and there exists an optimal stationary policy $\boldsymbol{a}^* = \{a^*(s)|s \in \mathcal{S}\}$.

B. Value Iteration Algorithm

According to [20, Proposition 4.2.1], there exists a scalar Λ^* together with some vector $\boldsymbol{h}^* = \{h^*(s)|s \in \mathcal{S}\}$ which satisfies Bellman's equation

$$\Lambda^* + h^*(s) = \max_{a\in\mathcal{A}(s)}\left[g(s,a) + \sum_{s'\in\mathcal{S}} p_{s\to s'|a} h^*(s')\right], \tag{20}$$

where Λ^* is the optimal average utility, and $h^*(s)$ is viewed as *relative or differential utility*.[1] It represents the maximum difference between the expected utility to reach a given state s_0 from state s for the first time and the utility that would be gained if the utility per stage was the average Λ^*. Furthermore, if $a^*(s)$ attains the maximum value of (20) for each s, the stationary policy $\boldsymbol{a}^*$ is optimal. Based on the Bellman's equation, instead of the long term average sum-rate maximization, we only need to deal with (20) which only relates with per-stage sum-rate $g(a, s)$ and state transition $p_{s\to s'|a}$. The value iteration algorithm [20, Sec. 4.4] can effectively solve the problem.

Specifically, we firstly initialize $h^{(0)}(s) = 0, \forall s \in \mathcal{S}$, and set a parameter $0 < \tau < 1$, which is used to guarantee the convergence of value iteration while obtaining the same optimal solution [20, Proposition 4.3.4]. Then we choose a state to calculate the relative utility. We choose a fixed state $s_0 = (0, 0, \mathbf{H}_0)$, and denote the output of the n-th iteration as $\boldsymbol{h}^{(n)} = \{h^{(n)}(s)|s \in \mathcal{S}\}$. For the $(n+1)$-th iteration, we first calculate

$$\Lambda^{(n+1)}(s_0) = \max_{a\in\mathcal{A}(s_0)}\left[g(s_0,a) + \tau\sum_{\mathbf{H}'}\Pr(\mathbf{H}'|\mathbf{H_0})h^{(n)}(s')\right], \tag{21}$$

where $s' = (B_1', B_2', \mathbf{H}')$, and B_1', B_2' are calculated according to (8) and (9), respectively. Then we calculate the relative utilities as

$$\begin{aligned} h^{(n+1)}(s) = {} & (1-\tau)h^{(n)}(s) \\ & + \max_{a\in\mathcal{A}(s)}\left[g(s,a) + \tau\sum_{\mathbf{H}'}\Pr(\mathbf{H}'|\mathbf{H})h^{(n)}(s')\right] \\ & - \Lambda^{(n+1)}(s_0). \end{aligned} \tag{22}$$

[1] In the textbook [20], $h^*(s)$ is defined as *relative cost* instead since the objective there is to minimize the average cost.

Recall that the parameter τ is used to guarantee the convergence of the relative value iteration. It can be viewed as replacing the relative utility $h(s)$ by $\tau h(s)$, which is proved not to change the optimal value. As the optimal average utility is irrelative with the initial state, $\Lambda^{(n+1)}(s_0)$ converges to Λ^*.

Notice that the states and the actions are all in the continuous space. By discretizing the state space and the action space, the MDP framework can be applied to solve the problem. However, to make the solution accurate, the granularity of the discretization needs to be sufficiently small, which results in a tremendous number of states, especially for the 2×2 MIMO channels (4 elements, each with two scalars: real part and imaginary part). As a consequence, we need to not only calculate the per-stage sum-rate function $g(s, a)$ that includes maximization operation for all states, but also iteratively update all the relative utilities $h(s)$. In this sense, solving the MDP problem encounters unaffordable high computational complexity, which is termed as the *curse of dimensionality* [20]. To reduce the computational complexity, on the one hand, the maximization problem in the per-stage sum-rate function should be solved efficiently. On the other hand, the complexity of the iteration algorithm should be reduced via some approximation. In the next two sections, we will discuss these two aspects in detail.

IV. Per-Frame Sum-Rate Maximization

In this section, we firstly consider the per-stage sum-rate function $g(s_t, a_t)$, i.e., the sum-rate maximization problem in each frame for the current state $s_t = (B_{t,1}, B_{t,2}, \mathbf{H}_t)$ and the given action $a_t = (A_{t,1}, A_{t,2})$. We ignore the time index t for simplicity. The per-frame optimization problem can be formulated as

$$\begin{aligned} \max_{k,\alpha,\tilde{p},p_1,p_2} \quad & \alpha\log_2\left(1+\frac{\tilde{p}|H_{\tilde{i}k}|^2}{\sigma_n^2}\right) \\ & + (1-\alpha)\sum_{i=1}^{2}\log_2\left(1+\frac{p_i}{\sigma_n^2}\right) \end{aligned} \tag{23}$$

$$\text{s.t.} \quad \alpha\tilde{p} \le \frac{B_k}{T_f} + \alpha E_k, \tag{24}$$

$$(1-\alpha)\sum_{i=1}^{2}|w_{ki}|^2 p_i + \alpha\tilde{p} \le A_k, \tag{25}$$

$$(1-\alpha)\sum_{i=1}^{2}|w_{\bar{k}i}|^2 p_i \le A_{\bar{k}}, \tag{26}$$

$$\tilde{p}, p_1, p_2 \in \mathcal{R}^+. \tag{27}$$

$$0 \le \alpha \le 1. \tag{28}$$

As $k \in \{1, 2\}$, the optimization over k can be done by solving the problem for all k, and selecting the one with larger sum-rate. Thus, we only need to consider the problem for a given k. Then the optimization problem can be rewritten as

$$\max_{\alpha,\tilde{p},p_1,p_2} \alpha\log_2\left(1+\frac{\tilde{p}|H_{\tilde{i}k}|^2}{\sigma_n^2}\right) + (1-\alpha)\sum_{i=1}^{2}\log_2\left(1+\frac{p_i}{\sigma_n^2}\right) \tag{29}$$

The problem (29) with constraints (24)-(28) is not convex in general. However, as shown later, given α, the power

allocation problem is a convex optimization, and the optimization over α given the optimal power allocation is also convex. According to these properties, we study the optimization of power allocation and subframe division separately.

A. Power Allocation Optimization

If we fix the variables k and α in (29), we obtain a power allocation optimization problem, which has the following property.

Theorem 1: For given k and α, the problem

$$\max_{\tilde{p},p_1,p_2} \ \alpha \log_2\left(1+\frac{\tilde{p}|H_{\bar{i}k}|^2}{\sigma_n^2}\right)+(1-\alpha)\sum_{i=1}^{2}\log_2\left(1+\frac{p_i}{\sigma_n^2}\right) \tag{30}$$

with constraints (24) - (27) is a convex optimization problem.

Proof: Once α is fixed, the objective function is the maximization of a summation of concave functions, and all the constraints are linear. As a result, the problem is convex. ∎

Theorem 1 tells us that for a given k and α, the optimal solution can be found by solving a convex optimization problem for power allocation. According to the convex optimization theory [24], we have the following observation.

Proposition 1: For a given k, when the optimal solution for the problem (29) with constraints (24)-(28) is achieved, either (25) or (26) is satisfied with equality.

Proof: See Appendix I. ∎

However, Proposition 1 cannot guarantee the equality holds in both (25) and (26). If both are satisfied with equality, the problem can be simplified and the solution can be given in closed-form. As a matter of fact, an equivalent problem can be formulated which only needs to solve the power allocation problem with equality held in (25) and (26). To get the result, we firstly provide a useful lemma as follows.

Lemma 1: The relative utility $h^*(s) = h^*(B_1, B_2, \mathbf{H})$ is nondecreasing w.r.t. B_1(or B_2) for given B_2(or B_1) and $\mathbf{H}$.

Proof: See Appendix II. ∎

Intuitively, more energy in the battery can support higher data rate. Hence, the utility increases with the increase of the battery energy. Based on Lemma 1, we have the following conclusion.

Theorem 2: Define $\bar{g}(s,a) = g(s,a)$ where the optimization is under the constraints (24), (27), (28) and the equality constraints

$$(1-\alpha)\sum_{i=1}^{2}|w_{ki}|^2 p_i + \alpha\tilde{p} = A_k, \tag{31}$$

$$(1-\alpha)\sum_{i=1}^{2}|w_{\bar{k}i}|^2 p_i = A_{\bar{k}}, \tag{32}$$

we have

$$\begin{aligned}\Lambda^* &= \max \lim_{N\to\infty} \mathbb{E}_{\mathbf{H}}\left[\frac{1}{N}\sum_{t=1}^{N} g(s_t, a_t(s_t))\right] \\ &= \max \lim_{N\to\infty} \mathbb{E}_{\mathbf{H}}\left[\frac{1}{N}\sum_{t=1}^{N} \bar{g}(s_t, a_t(s_t))\right]\end{aligned}$$

Proof: See Appendix III. ∎

Based on Theorem 2, we only need to solve the maximization problem under the equality constraints (31) and (32). The optimal power allocation solution as follows.

Theorem 3: For a given k and $0 < \alpha < 1$, we denote

$$\tilde{p}_{\min} = \max\left\{0, \frac{C_2}{\alpha|w_{\bar{k}1}|^2}\right\}, \tag{33}$$

$$\tilde{p}_{\max} = \min\left\{\frac{B_k}{\alpha T_f} + E_k, \frac{C_1}{\alpha|w_{\bar{k}2}|^2}\right\}, \tag{34}$$

define the set $\mathcal{P}_{k,\alpha} = \left\{\tilde{p}\middle|\tilde{p}_{\min} \le \tilde{p} \le \tilde{p}_{\max}\right\}$, and denote $\tilde{p}_0$ as the nonnegative root of

$$\begin{aligned}\frac{|H_{\bar{i}k}|^2}{\sigma_n^2+\tilde{p}|H_{\bar{i}k}|^2} &- \frac{(1-\alpha)|w_{\bar{k}2}|^2}{\sigma_n^2 C_0 + C_1 - \alpha|w_{\bar{k}2}|^2\tilde{p}} \\ &+ \frac{(1-\alpha)|w_{\bar{k}1}|^2}{\sigma_n^2 C_0 + C_2 + \alpha|w_{\bar{k}1}|^2\tilde{p}} = 0\end{aligned} \tag{35}$$

where $C_0 = (1-\alpha)(|w_{k1}|^2|w_{\bar{k}2}|^2 - |w_{k2}|^2|w_{\bar{k}1}|^2)$, $C_1 = A_k|w_{\bar{k}2}|^2 - A_{\bar{k}}|w_{k2}|^2$, $C_2 = A_k|w_{\bar{k}1}|^2 - A_{\bar{k}}|w_{k1}|^2$, and assume $C_0 > 0$. Then the solution for the problem (30) with constraints (24), (27), (31) and (32) is

- If $\mathcal{P}_{k,\alpha} = \emptyset$, the problem is infeasible.
- Otherwise, we have
  (1) if $\tilde{p}_0 \in \mathcal{P}_{k,\alpha}$, $\tilde{p}^* = \tilde{p}_0$ is the optimal power for the single-BS transmission phase;
  (2) if $\tilde{p}_0 > \tilde{p}_{\max}$, $\tilde{p}^* = \tilde{p}_{\max}$ is optimal;
  (3) if $\tilde{p}_0 < \tilde{p}_{\min}$, $\tilde{p}^* = \tilde{p}_{\min}$ is optimal;
  and the optimal $p_i^*, i = 1, 2$ can be obtained via

$$p_1^* = \frac{C_1 - \alpha|w_{\bar{k}2}|^2\tilde{p}^*}{C_0}, \tag{36}$$

$$p_2^* = \frac{\alpha|w_{\bar{k}1}|^2\tilde{p}^* - C_2}{C_0}. \tag{37}$$

Proof: See Appendix IV. ∎

Notice the solutions for $\alpha = 0$ and $\alpha = 1$ are not included in the proposition as they are trivial. For $\alpha = 0$, ZF-JT is applied in the whole frame. Then $\tilde{p} = 0$ and $p_i, i = 1, 2$ are obtained by solving (31) and (32). For $\alpha = 1$, the problem is feasible only when $A_{\bar{k}} = 0$, then $p_i = 0, i = 1, 2$ and $\tilde{p}$ can be obtained by solving (31). According to Theorem 3, for $0 < \alpha < 1$, the power allocation problem (30) for the fixed k and α with equality constraints (31) and (32) can be solved by calculating and comparing the values of $\tilde{p}_{\min}$, $\tilde{p}_{\max}$, and $\tilde{p}_0$. As they can be expressed in closed-form, the calculation is straightforward and simple. On the contrary, solving the original power allocation problem with inequality constraints (25) and (26) requires searching over the feasible set via iterations such as interior-point method [24, Ch. 11].

B. Optimization Over α

Besides the power allocation, we need to further determine optimal time ratio α. As a matter of fact, the following theorem tells us that the optimization over α is also convex.

Algorithm 1 Bi-section search algorithm to find the maximum $\bar{F}_k(\alpha)$

Initialize $\delta\alpha > 0$, $\underline{\alpha} = \alpha_{\min}$, $\bar{\alpha} = 1$, $I = 0$.
while $I = 0$ **do**
  Set $\hat{\alpha} = \frac{1}{2}(\underline{\alpha} + \bar{\alpha})$.
  if $\bar{F}_k(\hat{\alpha}) \geq \bar{F}_k(\hat{\alpha} - \delta\alpha)$ and $\bar{F}_k(\hat{\alpha}) \geq \bar{F}_k(\hat{\alpha} + \delta\alpha)$ **then**
    Set $I = 1$.
  else
    if $\bar{F}_k(\hat{\alpha} - \delta\alpha) \leq \bar{F}_k(\hat{\alpha}) \leq \bar{F}_k(\hat{\alpha} + \delta\alpha)$ **then**
      Set $\underline{\alpha} = \hat{\alpha}$.
    else
      Set $\bar{\alpha} = \hat{\alpha}$.
    end if
  end if
end while
The optimal solution is $\bar{F}_k(\hat{\alpha})$.

Theorem 4: For a given k, define a function

$$F_k(\alpha) = \max_{\tilde{p}, p_1, p_2} \alpha \log_2\left(1 + \frac{\tilde{p}|H_{\tilde{i}k}|^2}{\sigma_n^2}\right) + (1-\alpha)\sum_{i=1}^{2} \log_2\left(1 + \frac{p_i}{\sigma_n^2}\right), \quad (38)$$

where $0 \leq \alpha \leq 1$ and the maximization is constrained by (24)-(27). $F_k(\alpha)$ is a concave function.

Proof: See Appendix V. ∎

Corollary 1: The function $\bar{F}_k(\alpha) = F_k(\alpha)$, where the maximization is under constraints (24), (27), (31), and (32), is a concave function.

Proof: The proof simply follows the lines of Appendix V. ∎

Since $\bar{F}_k(\alpha)$ is a concave function, the optimal α either satisfies $\bar{F}_k'(\alpha) = 0$ or takes the boundary values $\alpha_{\min}$ or 1, where $\alpha_{\min} \leq 1$ is presented in (51) in Appendix IV. However, the closed-form solution for $\bar{F}_k'(\alpha) = 0$ is not easy to be obtained as the expression of $\bar{F}_k$ with respect to α is complex. Giving the condition that the value of $\bar{F}_k(\alpha)$ itself is easy to be computed, we can adopt the bi-section search algorithm and in each iteration check the monotonicity of $\bar{F}_k(\alpha)$ in a small neighborhood of α. The bi-section search algorithm is detailed in Algorithm 1.

In Algorithm 1, $\delta\alpha$ should be carefully selected to balance the accuracy of the optimal solution $\hat{\alpha}$ and the convergence speed of the iteration. Before running the bi-section algorithm, we need to firstly check if the optimal is obtained at the boundary points. Altogether, the algorithm for calculating $\bar{g}(s, a)$ is summarized in Algorithm 2.

Algorithm 2 Per-stage utility calculation algorithm

Initialize $\bar{g}(s, a) = 0$ and $\delta\alpha > 0$.
for all $k = 1$ to 2 **do**
  if $\mathcal{P}_{k,\alpha_{\min}} \neq \emptyset$, and $\bar{F}_k(\alpha_{\min}) > \bar{F}_k(\alpha_{\min} + \delta\alpha)$ **then**
    Update $\bar{g}(s, a) \leftarrow \max\{\bar{g}(s, a), \bar{F}_k(\alpha_{\min})\}$.
  else if $\mathcal{P}_{k,1} \neq \emptyset$, and $\bar{F}_k(1) > \bar{F}_k(1 - \delta\alpha)$ **then**
    Update $\bar{g}(s, a) \leftarrow \max\{\bar{g}(s, a), \bar{F}_k(1)\}$.
  else
    Run Algorithm 1, and then update $\bar{g}(s, a) \leftarrow \max\{\bar{g}(s, a), \bar{F}_k(\hat{\alpha})\}$.
  end if
end for

V. Approximate Dynamic Programming

In this section, we adopt the approximate DP [20, Ch. 6] to solve the policy optimization problem and deal with the complexity issue due to the large number of system states. The basic idea of the approximate DP is to estimate the relative utility $h(s)$ via a set of parameters $\boldsymbol{c} = (c_1, c_2, \cdots, c_M)^T$ rather than to calculate the exact value. In this way, we only need to train the parameter vector $\boldsymbol{c}$ based on a small set of simulation samples. Specifically, we apply *approximate policy iteration* algorithm as the convergence property can be guaranteed. Firstly, we briefly introduce the *policy iteration* algorithm and its approximation version. Then we will implement the algorithm to solve our problem.

A. Policy Iteration Algorithm

The policy iteration algorithm includes two steps in each iteration: *policy evaluation* and *policy improvement*. It starts with any feasible stationary policy, and improves the objective step by step. Suppose in the n-th iteration, we have a stationary policy denoted by $\boldsymbol{a}^{(n)} = \{a^{(n)}(s) | s \in \mathcal{S}\}$. Based on this policy, we perform policy evaluation step, i.e., we solve the following linear equations

$$\Lambda^{(n)} + h^{(n)}(s) = g(s, a^{(n)}(s)) + \sum_{s' \in \mathcal{S}} p_{s \to s' | a^{(n)}(s)} h^{(n)}(s') \quad (39)$$

for $\forall s \in \mathcal{S}$ to get the average cost $\Lambda^{(n)}$ and the relative utility vector $\boldsymbol{h}^{(n)}$. Notice that the number of unknown parameters $(\Lambda^{(n)}, \boldsymbol{h}^{(n)})$ is one more than the number of equations. Hence, more than one solutions exist, which are different with each other by a constant value for all $h^{(n)}(s)$. Without loss of generality, we can select a fixed state s_0 so that $h^{(n)}(s_0) = 0$, then the solution for (39) is unique.

The second step is to execute the policy improvement to find a stationary policy $\boldsymbol{a}^{(n+1)}$ which minimizes the right hand side of Bellman's equation

$$a^{(n+1)}(s) = \arg\max_{a \in \mathcal{A}(s)} \left[g(s, a) + \sum_{s' \in \mathcal{S}} p_{s \to s' | a} h^{(n)}(s') \right]. \quad (40)$$

If $\boldsymbol{a^{(n+1)} = a^{(n)}}$, the algorithm terminates, and the optimal policy is obtained $\boldsymbol{a^* = a^{(n)}}$. Otherwise, repeat the procedure by replacing $\boldsymbol{a^{(n)}}$ with $\boldsymbol{a^{(n+1)}}$. It is proved that the policy *does* improve the performance, i.e., $\Lambda^{(n)} \leq \Lambda^{(n+1)}$ [20, Proposition 4.4.2],[2] and the policy iteration algorithm terminates in finite number of iterations [20, Proposition 4.4.1].

[2]For the average cost minimization problem discussed in Bertsekas's book, the direction of the inequality reverses.

B. Approximate Policy Evaluation

For the policy evaluation step, the approximation DP tries to approximate the relative utility $h^{(n)}(s)$ by

$$\tilde{h}^{(n)}(s, \boldsymbol{c}^{(n)}) = \boldsymbol{\phi}(s)^T \boldsymbol{c}^{(n)}, \tag{41}$$

where $\boldsymbol{\phi}(s) = (\phi_1(s), \phi_2(s), \cdots, \phi_M(s))^T$ is an $M \times 1$ vector representing the features associated with state s, and $\boldsymbol{c}^{(n)} = (c_1^{(n)}, c_2^{(n)}, \cdots, c_M^{(n)})^T$ is an $M \times 1$ parameter vector. Instead of calculating all the relative utilities, we can train the parameter vector $\boldsymbol{c}^{(n)}$ using a relative small number of utility values and then estimate the others by (41). Based on the estimated relative utility, the approximation of parameter vector for the next iteration is obtained by minimizing the least square error based on a weighted Euclidean norm, i.e.,

$$\boldsymbol{c}^{(n+1)} = \arg \min_{\boldsymbol{c} \in \mathcal{R}^M} ||\hat{\boldsymbol{h}}^{(n+1)} - \Phi \boldsymbol{c}||_{\boldsymbol{\xi}}^2, \tag{42}$$

where $||\boldsymbol{J}||_{\boldsymbol{\xi}} = \sqrt{\sum_{s \in \mathcal{S}} \xi(s) J^2(s)}$ with a vector of positive weights $\xi(s), \forall s \in \mathcal{S}, \sum_s \xi(s) = 1$, $\mathcal{R}^M$ represents the M-dimensional real space, Φ is a matrix that has all the feature vectors $\phi(s)^T, \forall s \in \mathcal{S}$ as rows, and $\hat{\boldsymbol{h}}^{(n+1)} = F(\Phi \boldsymbol{c}^{(n)})$, where $F(\Phi \boldsymbol{c}^{(n)}) = (F(\boldsymbol{\phi}(s_1)^T \boldsymbol{c}^{(n)}), F(\boldsymbol{\phi}(s_2)^T \boldsymbol{c}^{(n)}), \cdots)^T$ and for each state s,

$$\begin{aligned} F(\boldsymbol{\phi}(s)^T \boldsymbol{c}^{(n)}) = {} & g(s, a^{(n)}(s)) - \Lambda^{(n)} \\ & + \sum_{s' \in \mathcal{S}} p_{s \to s' | a^{(n)}(s)} \phi(s')^T \boldsymbol{c}^{(n)}, \quad \forall s \in \mathcal{S}. \end{aligned} \tag{43}$$

For simplicity, the mapping F can be written in matrix form as in [20, Sec. 6.6], i.e., $F(\boldsymbol{h}) = \boldsymbol{g} - \Lambda \boldsymbol{e} + \boldsymbol{P}\boldsymbol{h}$, where Λ is the average utility, $\boldsymbol{P}$ is the transition probability matrix and $\boldsymbol{e}$ is the unit vector. Further more, the mapping F can be replaced by a parameterized mapping $F^{(\beta)} = (1 - \beta) \sum_{i=0}^{+\infty} \beta^i F^{i+1}$, where $\beta \in [0, 1)$, and $F^{i+1}(\boldsymbol{h}) = F^i(F(\boldsymbol{h}))$. The algorithm is called *least square policy evaluation with parameter β (LSPE(β))* [20, Ch. 6]. The benefit of introducing the parameter β is as follows. On the one hand, a higher convergence rate and smaller error bound can be obtained by setting larger β. On the other hand, when simulation is applied for approximation, larger β results in more pronounced simulation noise. Hence, tuning the parameter β helps to balance these factors. If $\beta = 0$, the mapping reduces to F.

Actually, we do not need to calculate samples of $\hat{h}(s)$ to estimate $\boldsymbol{c}$. Instead, the calculation can be done by simulation. Specifically, we generate a long simulated trajectory $s_0, s_1, \cdots$ based on the given action $\boldsymbol{a}^{(n)}$, and update $\boldsymbol{c}$ for each simulation realization according to the least square error metric. The advantage of simulation is that we only need a simulated trajectory rather than the state transition probability for a given policy. In reality, it means that we can use the simulated samples or the historical samples to directly calculate the estimated relative utility, instead of firstly estimate the transition probability and then estimate the utility. In the simulation-based LSPE(β) algorithm, $\boldsymbol{c}$ is updated iteratively according to each simulation sample. It can be expressed in matrix form [20, Sec. 6.6] as for the i-th sample,

$$\boldsymbol{c}_{i+1} = \boldsymbol{c}_i + \boldsymbol{B}_i^{-1}(\boldsymbol{A}_i \boldsymbol{c}_i + \boldsymbol{b}_i), \tag{44}$$

where

$$\begin{aligned} \boldsymbol{A}_i &= \frac{i}{i+1} \boldsymbol{A}_{i-1} + \frac{1}{i+1} z_i (\boldsymbol{\phi}(s_{i+1})^T - \boldsymbol{\phi}(s_i)^T), \\ \boldsymbol{B}_i &= \frac{i}{i+1} \boldsymbol{B}_{i-1} + \frac{1}{i+1} \boldsymbol{\phi}(s_i) \boldsymbol{\phi}(s_i)^T, \\ \boldsymbol{b}_i &= \frac{i}{i+1} \boldsymbol{b}_{i-1} + \frac{1}{i+1} z_i (g(s_i, a^{(n)}(s_i)) - \Lambda_i), \\ z_i &= \beta z_{i-1} + \boldsymbol{\phi}(s_i), \\ \Lambda_i &= \frac{1}{i+1} \sum_{j=0}^{i} g(s_j, a^{(n)}(s_j)), \end{aligned}$$

for all $i \geq 0$ and the boundary values $\boldsymbol{A}_{-1} = 0$, $\boldsymbol{B}_{-1} = 0, \boldsymbol{b}_{-1} = 0, z_{-1} = 0$. Note that there are two iterations in the approximate DP. The outer iteration runs policy evaluation and policy improvement to update the policy, the inner iteration runs the LSPE(β) algorithm to update the parameter vector $\boldsymbol{c}$. In the n-th policy evaluation, the policy $\boldsymbol{a}^{(n)}$ is viewed as an input to generate the simulation trajectory and calculate $\boldsymbol{c}_i$ according to (44) in the inner iteration. When the difference between $\boldsymbol{c}_{i+1}$ and $\boldsymbol{c}_i$ is small enough, the policy evaluation process terminates and we get $\boldsymbol{c}^{(n)} = \boldsymbol{c}_i$. Then the policy is updated using $\boldsymbol{c}^{(n)}$, i.e.,

$$a^{(n+1)}(s) = \arg \max_{a \in \mathcal{A}(s)} \left[g(s, a) + \sum_{s' \in \mathcal{S}} p_{s \to s' | a} \boldsymbol{\phi}(s')^T \boldsymbol{c}^{(n)} \right].$$

Generally, the length of the simulation trajectory is small than the number system state. Hence, the computational complexity of policy evaluation step can be reduced, especially when the number of states is large. Notice that the policy improvement step still needs to go through all the states due to the existence of the maximization operation.

C. Implementation Issues

To get an efficient approximate DP algorithm, the features of each state $\boldsymbol{\phi}(s)$ needs to be carefully selected. In our problem, we consider the following features.

- Energy-related features to indicate the influence of available energy on the utility. As the utility is represented in terms of data rate, the energy-related features are defined as $\log_2(1 + \frac{B_k / T_f + E_k}{\sigma_n^2}), k = 1, 2$.
- Channel-related features to indicate the influence of channel gain. Similarly, they are defined as $\log_2 (1 + |H_{ik}|^2), i = 1, 2, k = 1, 2$.
- Cooperation features to indicate the influence of JT. As a MIMO system, the eigenvalues are the key indicator of the MIMO link performance. Hence, we define this type of feature as $\log_2(1+\rho_i), i = 1, 2$, where $\rho_i, i = 1, 2$ are the eigenvalues of matrix $\mathbf{H}\mathbf{H}^H$.
- The 2nd-order features. As the actual data rate is calculated by the product of power and channel gain, we further consider the following features: $\log_2 (1 + \frac{(B_k / T_f + E_k)|H_{ik}|^2}{\sigma_n^2}), i = 1, 2, k = 1, 2$ and $\log_2 (1 + \frac{(B_k / T_f + E_k)\rho_k}{\sigma_n^2}), k = 1, 2$.

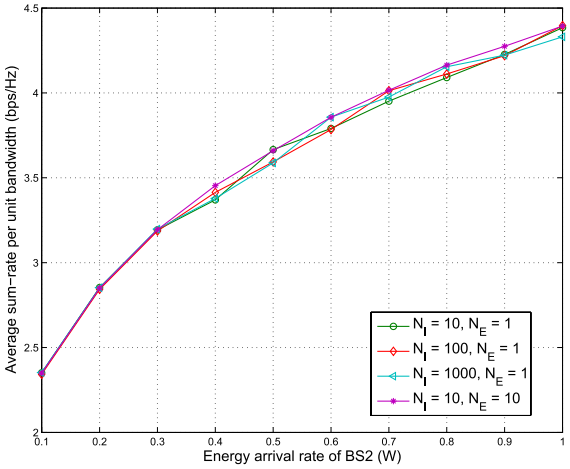

Fig. 3. The influence of number of iterations and number of policy explorations on the sum-rate performance of approximate DP. The energy arrival rate of BS1 is 0.1W.

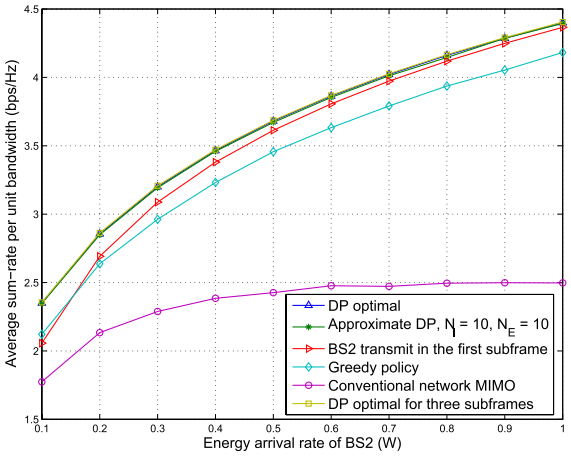

Fig. 4. Average sum-rate comparison of different algorithms. The energy arrival rate of BS1 is 0.1W.

The second issue concerning the approximate DP is that as the estimated relative utility is calculated based on the simulation samples generated for a given policy. Thus, some states that are unlikely to occur under this policy are under-represented. As a result, the relative utility estimation of these states may be highly inaccurate, causing potentially serious errors in the policy improvement process. This problem is known as *inadequate exploration* [20, Sec. 6.2] of the system dynamics. One possible way for guaranteeing adequate exploration of the state space is to frequently restart the simulation from a random state under a random policy. We call it as *policy exploration*. We will show later in the next section the influence of policy exploration on the performance.

VI. Simulation Study

We study the performance of the proposed algorithms by simulations. We adopt the outdoor pico-cell physical channel model from 3GPP standard [25]. The pathloss is $\text{PL} = 140.7 + 36.7 \log_{10} d$ (dB), where the distance d is measured in km. The distance between pico BSs is 100m. The shadowing fading follows log-normal distribution with variance 10dB. The small-scale fading follows Rayleigh distribution with zero mean and unit variance. The average SNR at the cell edge (50m to the pico BS) with transmit power 30dBm is set to 10dB. We set the two users are placed in the cell edge of the two pico BSs depicted in Fig. 1. Hence, they experience the same large-scale fading. The BSs are equipped with energy harvesting devices (e.g. solar panels). The transmit power of pico BSs is around hundreds of mW, and we set the energy harvesting rate accordingly.

Firstly, we evaluate the influence of number of iterations in the approximate DP on the performance. We fix the energy arrival rate of BS1 as 0.1W and change that of BS2. Denote the number of iterations for policy improvement by N_I, and the number of policy explorations which restarts the policy iteration by N_E. We set different values of N_I and N_E to run the approximate DP algorithm and compare the achievable sum-rate. The result is shown in Fig. 3. From this figure, we can see that if policy exploration is not considered, i.e., $N_E = 1$, the approximate DP reveals some random fluctuation. Solely increasing the number of policy iterations is not guaranteed to improve the performance. On the other hand, by increasing the number of policy explorations, the fluctuation can be efficiently reduced and the performance can be greatly improved, even with relatively small number of policy iterations. This validates the claim that the simulation-based policy iteration may be inaccurate, and it is quite important to adopt policy exploration in the approximate DP algorithm design.

Then we show the performance of approximate DP compared with the optimal policy obtained via DP optimal algorithm. And the following baselines are also considered for comparison. In the conventional network MIMO, the whole frame applies ZF-JT without sub-frame spitting. In the greedy policy, we do not optimize the energy allocation among frames, but greedily use all the available energy for sum-rate maximization in each frame. Mathematically, we solve the problem (23) under constraints (24)-(28) with $A_k = \frac{B_k}{T_f} + E_k$, $k = 1, 2$. Hence, instead of finding the policy for each state before the system runs, we can get the online solution based on current system state. According to Theorem 1 and Theorem 4, the problem can be solved by firstly applying bi-section search over α and then for each α calculating optimal power allocation via convex optimization. Besides, we consider always selecting the BS with higher energy arrival rate to transmit in the single-BS transmission subframe. Finally, we also consider a more general fractional JT scheme that divides each frame into three subframes: Each BS transmits individually in the first and second subframe, and then they jointly transmit in the third subframe. We also solve the sum-rate maximization problem via DP.

By fixing the energy arrival rate of BS1 as 0.1W and changing that of BS2, the results are shown in Fig. 4. It can be seen that the generalized fractional JT scheme with three subframes provides little performance gain compared with the scheme with two subframes, even with symmetric energy arrival rates. Intuitively, the fractional JT with three subframes may perform

better in symmetric case. However, the performance depends not only on the energy arrival rates of two BSs, but also on the channel states. When the energy arrival rates are asymmetric, dividing each frame into two subframes and letting the BS with higher energy arrival rate to transmit in the first subframe is sufficient. When the energy arrival rates are symmetric, the channel states become the key factor. In fact, the case with asymmetric channel gains is analogous to the case with asymmetric energy profiles. Hence, letting the BS with higher channel gain to transmit in the first subframe is sufficient. The scheme with three subframes may be better in symmetric case, which is however of low probability as it requires the energy arrival rates and the channel states are jointly symmetric. In addition, the optimization for three subframes is much more complex than that for two subframes. Therefore, the fractional JT with two subframes is preferred.

It can be also seen in Fig. 4 that the proposed approximate DP algorithm with $N_I = 10$, $N_E = 10$ performs very close to the optimal one. In addition, the greedy policy show a noticeable gap to the optimal policy, which illustrates the necessity of inter-frame energy allocation optimization. Always choosing BS2 to transmit in the first subframe degrades the performance compared with the proposed algorithm, while the gap diminishes as the energy asymmetry becomes stronger. This is also due to the dependence of performance on both the energy profiles and the channel states. When the channel state of the BS with more energy is much worse than the other, it would be preferred to sleep to wait for a better channel. Also, the proposed fractional JT algorithm dramatically outperforms the conventional network MIMO algorithm, especially when the asymmetry of energy arrival rate between two BSs becomes severe. Notice that the performance gain is remarkable even for the symmetric case (energy arrival rate of BS2 is also 0.1W). As mentioned before, the gain comes from the asymmetry of channel states, which is analogous to the asymmetry of energy arrival rates. With the increase of energy arrival rate in BS2, the sum-rate of conventional algorithm saturates to around 2.5bps/Hz. The reason is that according to the power constraint (4), the power constraint of BS2 associated with sufficiently large budget $P_{t,2}$ is usually satisfied with strict inequality. Then, increasing $P_{t,2}$ does not affect the optimization result. That is, the sum-rate does not increase as the higher energy arrival rate of BS2 does not contribute. On the other hand, the sum-rate of the fractional JT increases in the speed of log function. It also shows the importance of applying fractional JT in energy harvesting system.

We further simulate the case that the energy arrival rate is sufficient for transmission. We set the maximum transmit power per frame as 1.2W. The energy arrival rate of BS2 is equal to the maximum power per frame, and we vary the rate of BS1 to obtain the curves in Fig. 5. It can be seen that the performance gain of the proposed fractional JT strategy compared with the conventional network MIMO decreases as the energy arrival rate of BS1 becomes closer to that of BS2. And all the curves tend to be flat when the maximum transmit power can be satisfied by energy harvesting. Besides, always choosing BS2 to transmit in the first subframe approaches optimal then the energy asymmetry is strong. But it performs even worth than the greedy policy in symmetric case when the maximum transmit power is achieved in both BSs.

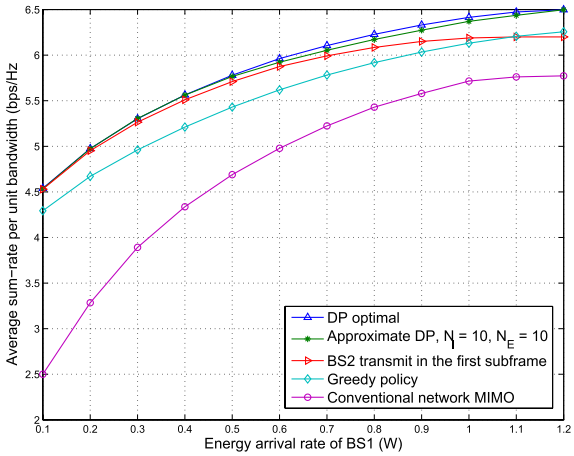

Fig. 5. Average sum-rate comparison of different algorithms. The energy arrival rate of BS2 is 1.2W.

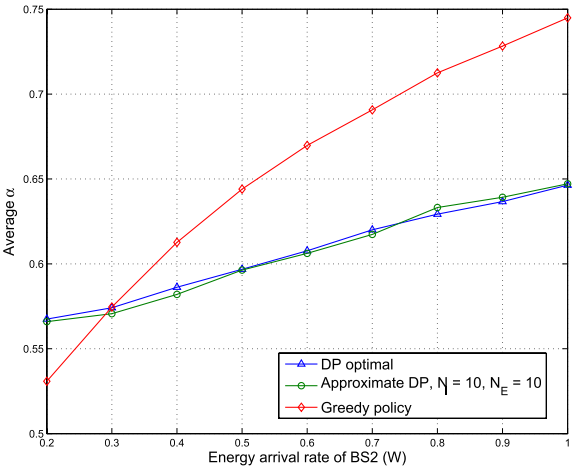

Fig. 6. Average time ratio α for single-transmission phase of different algorithms. The energy arrival rate of BS1 is 0.1W.

Fig. 6 shows the average time ratio α for single-transmission phase versus the energy arrival rate of BS2. It can be seen that average α increases as the asymmetry of energy arrival rates increases. Furthermore, the average α of DP optimal algorithm increases at the lowest speed, and the approximate DP algorithm performs very close to it. The greedy policy can only increase the time ratio for single-transmission to better utilize the higher energy arrival rate, and hence α increases at a higher speed w.r.t. the increase of energy arrival rate of BS2. On the contrary, by averaging the available energy over the transmission frames in the DP optimal and approximate DP algorithms, relatively more time ratio can be used to apply network MIMO to enhance the sum-rate.

Finally, the cumulative distribution function (CDF) of user data rate is depicted in Fig. 7 with energy arrival rates of the two BSs as 0.1W and 0.8W, respectively. It shows that the proposed fractional JT algorithm greatly enhances

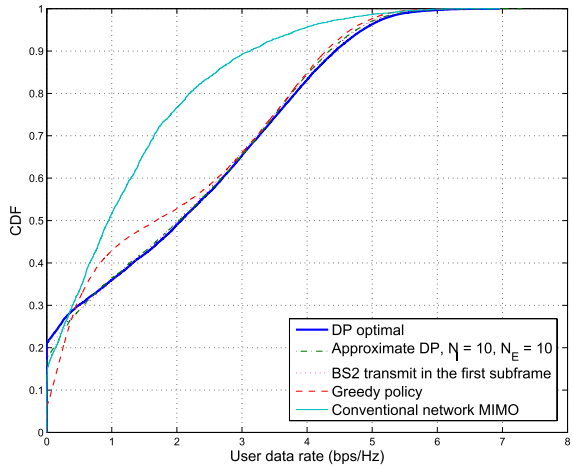

Fig. 7. Cumulative distribution function of user data rate with different algorithms. The energy arrival rate of BS1 is 0.1W, and that of BS2 is 0.8W.

the user data rate compared with the conventional network MIMO, and the proposed approximate DP algorithm achieves close-to-optimal performance. Since the energy arrival rate of BS2 is much larger than BS1, simply choosing BS2 to transmit in the first subframe also performs close to the optimal. Notice that the greedy policy reduces the percentage of zero data rate since it transmits with all the available energy in each frame, with the sacrifice of channel fading diversity for opportunistic inter-frame scheduling. As a result, the ratio of low data rate is much higher than the DP-based algorithms. For instance, about 43% of users' data rate is lower than 1bps/Hz. With DP-based algorithms, the ratio reduces by about 8%.

VII. Conclusion

In this paper, we have proposed a fractional JT scheme for BS cooperation that divides a transmission frame to firstly apply single-BS transmission and then adopt ZF-JT transmission to enhance the average sum-rate. The MDP-based problem is formulated and solved by firstly allocating energy among frames and then optimizing per-frame sum-rate. By analyzing the convexity of per-frame sum-rate optimization problem, and applying approximate DP algorithm, the computational complexity is greatly reduced. The proposed fractional JT scheme has been shown to achieve much higher sum-rate compared with the conventional ZF-JT only scheme. As the energy arrival asymmetry increases, the achievable rate of ZF-JT saturates (2.5bps/Hz in our settings), while the proposed scheme reveals a logarithmic increase. The proposed approximate DP algorithm can approach the DP optimal algorithm with sufficient number of policy explorations.

In this paper, fractional JT with two subframes is considered since we only consider the transmit power consumption. If the non-ideal circuit power is considered, more general frame structure is required to further save energy. Specifically, the BSs may turn to idle mode to reduce the circuit power consumption. This would be an interesting research direction for future work.

Appendix I
Proof of Proposition 1

For any given α, the power allocation solution satisfies the Karush-Kuhn-Tucker (KKT) conditions [24]. Define the Lagrangian function for any multipliers $\lambda \geq 0, \mu \geq 0$, $\eta \geq 0$ as

$$\begin{aligned}\mathcal{L} = &-\left(\alpha \log_2\left(1+\frac{\tilde{p}|H_{\tilde{i}k}|^2}{\sigma_n^2}\right)+(1-\alpha)\sum_{i=1}^{2}\log_2\left(1+\frac{p_i}{\sigma_n^2}\right)\right)\\ &+\lambda\left(\alpha\tilde{p}-\frac{B_k}{T_f}-\alpha E_k\right)\\ &+\mu\left((1-\alpha)\sum_{i=1}^{2}|w_{ki}|^2 p_i+\alpha\tilde{p}-A_k\right)\\ &+\eta\left((1-\alpha)\sum_{i=1}^{2}|w_{\bar{k}i}|^2 p_i-A_{\bar{k}}\right)\end{aligned} \tag{45}$$

with additional complementary slackness conditions

$$\lambda\left(\alpha\tilde{p}-\frac{B_k}{T_f}-\alpha E_k\right)=0,$$

$$\mu\left((1-\alpha)\sum_{i=1}^{2}|w_{ki}|^2 p_i+\alpha\tilde{p}-A_k\right)=0,$$

$$\eta\left((1-\alpha)\sum_{i=1}^{2}|w_{\bar{k}i}|^2 p_i-A_{\bar{k}}\right)=0.$$

Here, we ignore the non-negative power constraints in the above formulation to simplify the expression. It can be directly added to the result. We apply the KKT optimality conditions to the Lagrangian function (45). By setting $\partial\mathcal{L}/\partial\tilde{p}=\partial\mathcal{L}/\partial p_i=0$, we obtain

$$\tilde{p}^*=\left[\frac{1}{\lambda+\mu}-\frac{\sigma_n^2}{|H_{\tilde{i}k}|^2}\right]^+, \tag{46}$$

$$p_i^*=\left[\frac{1}{\mu|w_{ki}|^2+|w_{\bar{k}i}|^2\eta}-\sigma_n^2\right]^+, \quad i=1,2. \tag{47}$$

Notice that to guarantee the validity of (47), either μ or η should be non-zero, which means that at least one of (25) and (26) is satisfied with equality.

Appendix II
Proof of Lemma 1

Since $h^*(s)=\lim_{n\to+\infty}h^{(n)}(s)$, we prove the monotonicity property by induction. In addiction, we only need to prove the monotonicity for B_1. The proof for B_2 follows the same procedure.

Obviously, it is true for $n=0$ as $h^{(0)}(s)=0, \forall s\in\mathcal{S}$. Assume that $h^{(n)}(B_1,B_2,\mathbf{H})$ is nondecreasing w.r.t B_1, and the optimal action for state $s=(B_1,B_2,\mathbf{H})$ is $a^*=(A_1^*,A_2^*)$, i.e.,

$$\begin{aligned}&\max_{a\in\mathcal{A}(s)}\left[g(s,a)+\tau\sum_{\mathbf{H}'}\Pr(\mathbf{H}'|\mathbf{H})h^{(n)}(s')\right]\\ &\quad=g(s,A_1^*,A_2^*)+\tau\sum_{\mathbf{H}'}\Pr(\mathbf{H}'|\mathbf{H})h^{(n)}(B_1',B_2',\mathbf{H}').\end{aligned}$$

Then consider the state $s'' = (B_1+\delta B, B_2, \mathbf{H})$, where $\delta B > 0$. We have

$$\begin{aligned} h^{(n+1)}(s'') &= (1-\tau)h^{(n)}(s'') + \max_{a\in\mathcal{A}(s'')} \\ &\quad \times \left[g(s'', a) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) h^{(n)}(s') \right] - \Lambda^{(n+1)}(s_0) \\ &\overset{(a)}{\geq} (1-\tau)h^{(n)}(s'') + g(s'', A_1^*, A_2^*) \\ &\quad + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) h^{(n)}(B_1'+\delta B, B_2', \mathbf{H}') - \Lambda^{(n+1)}(s_0) \\ &\overset{(b)}{\geq} (1-\tau)h^{(n)}(s) + g(s, A_1^*, A_2^*) \\ &\quad + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) h^{(n)}(B_1', B_2', \mathbf{H}') - \Lambda^{(n+1)}(s_0) \\ &= h^{(n+1)}(s), \end{aligned}$$

where the inequality (a) holds as the action $(A_1^*, A_2^*) \in \mathcal{A}(s'')$, and (b) holds due to the following two reasons. Firstly, $g(s'', A_1^*, A_2^*) \geq g(s, A_1^*, A_2^*)$ as the constraint (24) for the latter is not looser than the former. Secondly, $h^{(n)}(B_1'+\delta B, B_2', \mathbf{H}') \geq h^{(n)}(B_1', B_2', \mathbf{H}')$ due to the monotonicity of $h^{(n)}(B_1, B_2, \mathbf{H})$ w.r.t. B_1. As a result, we prove that $h^{(n+1)}(B_1, B_2, \mathbf{H})$ is also nondecreasing w.r.t. B_1.

In summary, $h^{(n)}(B_1, B_2, \mathbf{H})$ is nondecreasing w.r.t. B_1 for all $n = 0, 1, 2, \cdots$. Hence, we also have that $h^*(B_1, B_2, \mathbf{H})$ is nondecreasing w.r.t. B_1. The same holds for B_2.

Appendix III
Proof of Theorem 2

Regarding the per-stage utility $\bar{g}$, the Bellman's equation also holds for a scalar $\bar{\Lambda}^*$ and some vector $\bar{\boldsymbol{h}}^* = \{\bar{h}^*(s) | s \in \mathcal{S}\}$, and the value iteration algorithm works in the same way. Hence, we only need to prove by induction that $\Lambda^{(n)}(s_0) = \bar{\Lambda}^{(n)}(s_0)$ and $h^{(n)}(s) = \bar{h}^{(n)}(s)$.

We initialize that $\Lambda^{(0)}(s_0) = \bar{\Lambda}^{(0)}(s_0) = 0$ and $h^{(0)}(s) = \bar{h}^{(0)}(s) = 0, \forall s \in \mathcal{S}$. Suppose that $\Lambda^{(n)}(s_0) = \bar{\Lambda}^{(n)}(s_0), h^{(n)}(s) = \bar{h}^{(n)}(s), \forall s \in \mathcal{S}$. For the $(n+1)$-th iteration and $\forall s = (B_1, B_2, \mathbf{H}), a = (A_1, A_2)$, we have

$$\begin{aligned} &\bar{g}(s,a) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) h^{(n)}(B_1', B_2', \mathbf{H}') \\ &\quad \overset{(c)}{\leq} g(s,a) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) h^{(n)}(B_1', B_2', \mathbf{H}') \\ &\quad \overset{(d)}{\leq} g(s,a) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) h^{(n)}(B_1'', B_2'', \mathbf{H}') \end{aligned}$$

where $B_k' = B_k + T_f E_k - A_k, \forall k = 1, 2$, while $B_k'', k = 1, 2$ are calculated via (8) and (9), respectively. Hence we have $B_k'' \geq B_k', \forall k = 1, 2$. Inequality (c) holds as the maximization of g has larger feasible region than that of $\bar{g}$, while (d) holds due to the monotonicity of the relative utility $h(s)$. As a result, we have

$$\begin{aligned} &\max_{a\in\mathcal{A}(s)} \left[\bar{g}(s,a) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) \bar{h}^{(n)}(s') \right] \\ &\quad \leq \max_{a\in\mathcal{A}(s)} \left[g(s,a) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) h^{(n)}(s') \right] \end{aligned} \tag{48}$$

On the other hand, there exists an action (A_1^*, A_2^*) such that

$$\begin{aligned} &\max_{a\in\mathcal{A}(s)} \left[g(s,a) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) h^{(n)}(s') \right] \\ &\quad = g(s, A_1^*, A_2^*) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) h^{(n)}(B_1^*, B_2^*, \mathbf{H}'), \\ &\quad \overset{(e)}{=} \bar{g}(s, A_1^*, A_2^*) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) \bar{h}^{(n)}(B_1^*, B_2^*, \mathbf{H}'), \\ &\quad \overset{(f)}{\leq} \max_{a\in\mathcal{A}(s)} \left[\bar{g}(s,a) + \tau \sum_{\mathbf{H}'} \Pr(\mathbf{H}'|\mathbf{H}) \bar{h}^{(n)}(s') \right], \end{aligned} \tag{49}$$

where $B_k^* = B_k + T_f E_k - A_k^*, \forall k = 1, 2$, and hence, equality (e) holds. Inequality (f) holds as $(A_1^*, A_2^*) \in \mathcal{A}(s)$. It can be seen by (48), (49) jointly with (21) and (22) that $\Lambda^{(n+1)}(s_0) = \bar{\Lambda}^{(n+1)}(s_0)$ and $h^{(n+1)}(s) = \bar{h}^{(n+1)}(s)$.

In summary, we have $\Lambda^{(n)}(s_0) = \bar{\Lambda}^{(n)}(s_0), h^{(n)}(s) = \bar{h}^{(n)}(s)$ for all $n = 0, 1, 2, \cdots$. Hence, we have $\Lambda^* = \max \lim_{N\to\infty} \mathbb{E}_{\mathbf{H}}\left[\frac{1}{N}\sum_{t=1}^{N} \bar{g}(s_t, a_t(s_t))\right] = \bar{\Lambda}^*$.

Appendix IV
Proof of Theorem 3

According to the equality constraints (31) and (32), p_i, $i = 1, 2$ can be represented as functions of $\tilde{p}$, i.e., $p_1 = \frac{C_1 - \alpha|w_{\bar{k}2}|^2\tilde{p}}{C_0}$, $p_2 = \frac{\alpha|w_{\bar{k}1}|^2\tilde{p} - C_2}{C_0}$, where C_0, C_1, C_2 are presented in the proposition. As the elements of $\mathbf{H}$ are i.i.d., we have $C_0 \neq 0$. Hence, the per-stage sum rate function can be written as a function of $\tilde{p}$:

$$\begin{aligned} f_{k,\alpha}(\tilde{p}) &= \alpha \log_2\left(1 + \frac{\tilde{p}|H_{\bar{i}k}|^2}{\sigma_n^2}\right) \\ &\quad + (1-\alpha)\left[\log_2\left(1 + \frac{C_1 - \alpha|w_{\bar{k}2}|^2\tilde{p}}{\sigma_n^2 C_0}\right) \right. \\ &\quad \left. + \log_2\left(1 + \frac{\alpha|w_{\bar{k}1}|^2\tilde{p} - C_2}{\sigma_n^2 C_0}\right) \right]. \end{aligned} \tag{50}$$

The constraints can be written as the feasible set of $\tilde{p}$. Without loss of generality, we assume $C_0 > 0$. The feasible set for $C_0 < 0$ can be derived in the similar way. With the non-negative constraints $p_i \geq 0, i = 1, 2$, we have $\frac{C_2}{\alpha|w_{\bar{k}1}|^2} \leq \tilde{p} \leq \frac{C_1}{\alpha|w_{\bar{k}2}|^2}$. Jointly with (24) and $\tilde{p} \geq 0$, the feasible set can be expressed as $\mathcal{P}_{k,\alpha} = \{\tilde{p} | \tilde{p}_{\min} \leq \tilde{p} \leq \tilde{p}_{\max}\}$, where $\tilde{p}_{\min}$ and $\tilde{p}_{\max}$ are expressed as (33) and (34), respectively. To guarantee that $\mathcal{P}_{k,\alpha} \neq \emptyset$, we have $\tilde{p}_{\min} \leq \tilde{p}_{\max}$, which results in $\alpha \geq \frac{1}{E_k}\left(\frac{C_2}{|w_{\bar{k}1}|^2} - \frac{B_k}{T_f}\right)$. We set

$$\alpha_{\min} = \max\left\{0, \frac{1}{E_k}\left(\frac{C_2}{|w_{\bar{k}1}|^2} - \frac{B_k}{T_f}\right)\right\}. \tag{51}$$

Hence, there are two cases so that $\mathcal{P}_{k,\alpha} = \emptyset$. The first is $\alpha_{\min} > 1$, and the second is that $0 < \alpha_{\min} \leq 1$ and $0 \leq \alpha < \alpha_{\min}$. Otherwise, the per-frame optimization problem can be reformulated as

$$\max_{\tilde{p} \in \mathcal{P}_{k,\alpha}} f_{k,\alpha}(\tilde{p}), \tag{52}$$

whose convexity still holds according to the following lemma.

Lemma 2: The problem (52) is a convex optimization problem.

Proof: As the log function is concave and the functions inside the log operation are linear function of $\tilde{p}$, the composition of a linear function with a concave function is still concave. Hence, $f_{k,\alpha}(\tilde{p})$ is a concave function. On the other hand, the feasible set $\mathcal{P}_{k,\alpha}$ is convex. Therefore, the considered problem is a convex optimization problem. ∎

Due to the concavity of the function $f_{k,\alpha}(\tilde{p})$, the optimal solution can be found by solving $f'_{k,\alpha}(\tilde{p}) = 0$, which is expressed as (35). It can be transformed into a quadratic equation, and hence, the nonnegative root can be easily solved. Denote the solution for $f'_{k,\alpha}(\tilde{p}) = 0$ by $\tilde{p}_0$. Then according to the concavity of the function $f_{k,\alpha}$, the optimal solution for the problem $\max_{\tilde{p} \in \mathcal{P}_{k,\alpha}} f_{k,\alpha}(\tilde{p})$ is either $\tilde{p}_0$ or the boundary points of the feasible set $\mathcal{P}_{k,\alpha}$ depending on whether $\tilde{p}_0 \in \mathcal{P}_{k,\alpha}$ or not.

APPENDIX V
PROOF OF THEOREM 4

For any $\alpha^{(1)}, \alpha^{(2)} \in [0, 1]$, we assume that

$$F_k(\alpha^{(j)}) = \alpha^{(j)} \log_2\left(1 + \frac{\tilde{p}^{(j)}|H_{\tilde{i}k}|^2}{\sigma_n^2}\right) + (1-\alpha^{(j)}) \sum_{i=1}^{2} \log_2\left(1 + \frac{p_i^{(j)}}{\sigma_n^2}\right),$$

for $j = 1, 2$, i.e., $\tilde{p}^{(j)}, p_i^{(j)}, i = 1, 2$ achieve the maximum sum-rate. For any $0 < \gamma < 1$, we have

$$\gamma\, F_k(\alpha^{(1)}) + (1-\gamma) F_k(\alpha^{(2)}) \leq \alpha' \log_2\left(1 + \frac{\tilde{p}'|H_{\tilde{i}k}|^2}{\sigma_n^2}\right) + (1-\alpha') \sum_{i=1}^{2} \log_2\left(1 + \frac{p'_i}{\sigma_n^2}\right) \tag{53}$$

where

$$\begin{aligned} \alpha' &= \gamma\, \alpha^{(1)} + (1-\gamma)\alpha^{(1)}, \\ \tilde{p}' &= \frac{\gamma\, \alpha^{(1)}}{\alpha'} \tilde{p}^{(1)} + \frac{(1-\gamma)\alpha^{(2)}}{\alpha'} \tilde{p}^{(2)}, \\ p'_i &= \frac{\gamma\,(1-\alpha^{(1)})}{1-\alpha'} p_i^{(1)} + \frac{(1-\gamma)(1-\alpha^{(2)})}{1-\alpha'} p_i^{(2)}, \quad i = 1, 2, \end{aligned} \tag{54}$$

and the inequality in (53) is due to the concavity of log function. In addition,

$$\begin{aligned} \alpha' \tilde{p}' &= \gamma\, \alpha^{(1)} \tilde{p}^{(1)} + (1-\gamma)\alpha^{(2)} \tilde{p}^{(2)} \\ &\leq \gamma \left(\frac{B_k}{T_f} + \alpha^{(1)} E_k\right) + (1-\gamma)\left(\frac{B_k}{T_f} + \alpha^{(2)} E_k\right) \\ &= \frac{B_k}{T_f} + \alpha' E_k, \end{aligned}$$

i.e., $\tilde{p}'$ satisfies the constraint (24). Similarly, $\tilde{p}'$ and p'_i, $i = 1, 2$ also satisfy the constraints (25) and (26). Hence, $\tilde{p}', p'_i, i = 1, 2$ is a feasible power allocation solution. As $F_k(\alpha)$ is maximal over all power allocation policies, we have

$$\alpha' \log_2\left(1 + \frac{\tilde{p}'|H_{\tilde{i}k}|^2}{\sigma_n^2}\right) + (1-\alpha') \sum_{i=1}^{2} \log_2\left(1 + \frac{p'_i}{\sigma_n^2}\right) \leq F_k(\alpha'). \tag{55}$$

Combining (53), (54) and (55), we have

$$\gamma\, F_k(\alpha^{(1)}) + (1-\gamma) F_k(\alpha^{(2)}) \leq F_k(\gamma\, \alpha^{(1)} + (1-\gamma)\alpha^{(2)}).$$

As a consequence, F_k is a concave function.

REFERENCES

[1] China Mobile, "Confronting climate change to have a win-win for green future," accessed on May 29, 2014. [Online]. Available: http://labs.chinamobile.com/news/105225

[2] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, Sep. 2011.

[3] J. Gong, S. Zhou, and Z. Niu, "Optimal power allocation for energy harvesting and power grid coexisting wireless communication systems," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 3040–3049, Jul. 2013.

[4] C. Hu, J. Gong, X. Wang, S. Zhou, and Z. Niu, "Optimal green energy utilization in MIMO systems with hybrid energy supplies," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3675–3688, Aug. 2015.

[5] J. Yang, O. Ozel, and S. Ulukus, "Broadcasting with an energy harvesting rechargeable transmitter," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 571–583, Feb. 2012.

[6] J. Yang and S. Ulukus, "Optimal packet scheduling in a multiple access channel with energy harvesting transmitters," *J. Commun. Netw.*, vol. 14, no. 2, pp. 140–150, Apr. 2012.

[7] K. Tutuncuoglu and A. Yener, "Sum-rate optimal power policies for energy harvesting transmitters in an interference channel," *J. Commun. Netw.*, vol. 14, no. 2, pp. 151–161, Apr. 2012.

[8] C. Huang, J. Zhang, P. Zhang, and S. Cui, "Threshold-based transmissions for large relay networks powered by renewable energy," in *Proc. IEEE Globel Commun. Conf. (Globecom)*, Dec. 2013, pp. 1921–1926.

[9] A. Minasian, S. ShahbazPanahi, and R. S. Adve, "Energy harvesting cooperative communication systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6118–6131, Nov. 2014.

[10] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[11] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, Jr., "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[12] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. J. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2983–2989, Jun. 2009.

[13] 3GPP, "Coordinated multi-point operation for LTE physical layer aspects (release 11)," 3GPP, Sohpia Antipolis, France, Tech. Rep. TR36.819, Mar. 2012.

[14] F. Boccardi and H. Huang, "Zero-forcing precoding for the MIMO broadcast channel under per-antenna power constraints," in *Proc. IEEE 7th Workshop Signal Process. Adv. Wireless Commun.*, Jul. 2006, pp. 1–5.

[15] S. Kaviani and W. A. Krzymień, "Optimal multiuser zero forcing with per-antenna power constraints for network MIMO coordination," *EURASIP J. Wireless Commun. Netw.*, vol. 2011, Jan. 2011, Art. no. 1.

[16] B. Gurakan, O. Ozel, J. Yang, and S. Ulukus, "Energy cooperation in energy harvesting communications," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4884–4898, Dec. 2013.

[17] Y.-K. Chia, S. Sun, and R. Zhang, "Energy cooperation in cellular networks with renewable powered base stations," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6996–7010, Dec. 2014.

[18] J. Xu and R. Zhang, "CoMP meets smart grid: A new communication and energy cooperation paradigm," *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2476–2488, Jun. 2015.
[19] J. Gong, S. Zho, Z. Zhou, and Z. Niu, "Downlink base station cooperation with energy harvesting," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, Nov. 2014, pp. 87–91.
[20] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol. 2, 3rd ed. Belmont, MA, USA: Athena Scientific, 2005.
[21] C. Huang, R. Zhang, and S. Cui, "Optimal power allocation for outage probability minimization in fading channels with energy harvesting constraints," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 1074–1087, Feb. 2014.
[22] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.
[23] M. K. Karakayali, "Network coordination for spectrally efficient communications in wireless networks," Ph.D. dissertation, Dept. Elect. Comput. Eng., Rutgers Univ., New Brunswick, NJ, USA, 2007.
[24] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[25] 3GPP, "Further advancements for E-UTRA physical layer aspects (release 9)," 3GPP, Sophia Antipolis, France, Tech. Rep. TR36.814, Mar. 2010.

Jie Gong (S'09–M'13) received the B.S. and Ph.D. degrees from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 2008 and 2013, respectively. From 2012 to 2013, he visited the Institute of Digital Communications, University of Edinburgh, Edinburgh, U.K. From 2013 to 2015, he was a Post-Doctoral Scholar with the Department of Electronic Engineering, Tsinghua University, Beijing. He is currently an Associate Research Fellow with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His research interests include cloud RAN, energy harvesting technology, and green wireless communications. He served as a Workshop Co-Chair of the IEEE ISADS 2015 and a TPC Member of the IEEE/CIC ICCC 2016. He was a co-recipient of the Best Paper Award from the IEEE Communications Society Asia-Pacific Board in 2013.

Sheng Zhou (S'06–M'12) received the B.E. and Ph.D. degrees in electronics engineering from Tsinghua University, Beijing, China, in 2005 and 2011, respectively. In 2010, he was a Visiting Student with the Department of Electrical Engineering, Wireless System Lab, Stanford University, Stanford, CA, USA. He is currently an Associate Professor with the Department of Electronic Engineering, Tsinghua University. His research interests include cross-layer design for multiple antenna systems, cooperative transmission in cellular systems, and green wireless communications.

Dr. Zhou co-received the Best Paper Award at the Asia-Pacific Conference on Communication in 2009 and 2013, respectively, the 23th IEEE International Conference on Communication Technology in 2011, and the 25th International Tele-traffic Congress in 2013.

Zhenyu Zhou (S'06–M'11) received the M.E. and Ph.D. degrees from Waseda University, Tokyo, Japan, in 2008 and 2011, respectively. From 2012 to 2013, he was the Chief Researcher with the Department of Technology, KDDI, Tokyo, Japan. Since 2013, he has been an Associate Professor with the School of Electrical and Electronic Engineering, North China Electric Power University, China. He has been a Visiting Scholar with Tsinghua-Hitachi Joint Lab, Environment-Harmonious ICT, University of Tsinghua, Beijing, since 2014. His research interests include green communications and smart grid. He served as the Workshop Co-Chair of the IEEE ISADS 2015, the Session Chair of the IEEE Globecom 2014, and the TPC member of the IEEE Globecom 2015, the ACM Mobimedia 2015, and the IEEE Africon 2015. He received the Young Researcher Encouragement Award from the IEEE Vehicular Technology Society in 2009. He is a member of IEICE and CSEE.