Interference Suppression and Energy Efficiency Improvement With Massive MIMO and Relay Selection in Cognitive Two-Way Relay Networks

Shashindra Silva, *Student Member, IEEE*, Masoud Ardakani, *Senior Member, IEEE*, and Chintha Tellambura, *Fellow, IEEE*

***Abstract*—We analyze a relay assisted wireless communication link between two underlay massive multiple-input multiple-output (MIMO) terminals. Specifically, an amplify and forwarding (AF) two-way relay is optimally selected to maximize the sum rate and to keep the interference on the primary user (PU) below an interference threshold. We first obtain asymptotic signal-to-interference-plus-noise ratio (SINR) values for two scenarios: (1) the relays and the two end nodes use transmit power scaling and (2) only the end nodes use transmit power scaling. For these two cases, we derive optimal power allocations subject to the PU interference constraints. With these optimal power allocations, we analyze how relay selection impacts the outage, the sum rate, and the energy efficiency of the network. For the first scenario, the outage can be reduced to zero with appropriate power allocation and the relay selection can be done offline. For the second scenario, outage will depend on the instantaneous channel state between the relays and the PU. Furthermore, for a realistic power consumption model, we analyze the energy efficiency and show that relay selection will increase it.**

***Index Terms*—Massive MIMO, energy efficiency, relay selection, cognitive radio.**

I. Introduction

A. Background and Motivation

GLOBAL mobile traffic data increased by 63% in 2016 and is expected to grow by seven fold during the next five years [2], contributing to high energy consumption and spectrum crunch. Hence, the fifth generation (5G) wireless design goals are to improve the energy efficiency (the number of bits transmitted per unit energy consumption [3], [4]) and the spectral efficiency (the number of bits transmitted per second over 1 Hz of bandwidth).

Three technologies have appeared to improve the spectral efficiency. First, in cognitive radios, primary user (PU) spectrum allocated can be temporarily released for use by an unlicensed secondary user subject to certain conditions [5]. In underlay cognitive radios, this means that the interference on PUs must be below a threshold [6]. If this is not possible, the underlay (aka secondary) network suffers an outage. Second technology is the use of two-way relay networks (TWRNs). In a TWRN, two end nodes communicate via a relay node, which then enables bi-directional data transfer between them [7]. These allow potential doubling of the spectral efficiency one-way relays, which penalize the spectral efficiency compared to that of direct links (see [8] for one-way relays in multiuser networks). TWRNs are included in the long-term evolution-advanced (LTE-A) standard [9].

Third, and most significant, technology is the use of multiple antennas at one or both ends of a link, also known as multiple-input multiple-output (MIMO). Massive MIMO (mMIMO) takes MIMO to the next level by using very large antenna arrays (100 to 500 antennas) at the base station [10]. A consequence of using such large arrays is the disappearance of the effect of small-scale fading (channel hardening) [11]. Due to this effect, the required signal processing simplifies; thus, simple linear beamformers and precoders such as zero forcing (ZF) and matched filtering (MF) become nearly optimal. Due to these advantages, mMIMO is a key enabler of the future 5G wireless [12] and provides very high spectral efficiency and energy efficiency [10], which may facilitate green communications [13].

Finally, the use of these three technologies will have major implications on energy efficiency. Moreover, when multiple relay nodes are present, relay selection extracts spatial diversity, which will enhance the reliability and the data rates [14]–[17]. Relay selection methods can be based on different objectives such as maximizing the sum rate or minimizing the bit error rate and can significantly increase the performance of TWRNs [18]. TWRNs can also be integrated with MIMO to further improve the achievable data rate and the reliability due to spatial multiplexing and diversity gains [19], [20]. So the overarching question is what benefits can be had by integrating all these together.

B. Previous Works on Cognitive TWRNs and mMIMO

In fact, cognitive relays have been analyzed in [21]–[24]. Specifically, [21] analyzes a system where the primary and the secondary systems mutually coordinate to enable relay assisted communications and [22] derives detection and outage probabilities in the overlay mode. Also, [23] analyzes an

Manuscript received November 19, 2018; revised April 9, 2019, June 30, 2019, August 10, 2019, and November 6, 2019; accepted December 31, 2019. Date of publication January 9, 2020; date of current version May 19, 2020. This work was presented in part at the IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, USA, 2017. The associate editor coordinating the review of this article and approving it for publication was R. Vaze. *(Corresponding author: Shashindra Silva.)*

The authors are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: jayamuni@ualberta.ca; ardakani@ualberta.ca; chintha@ualberta.ca).

Digital Object Identifier 10.1109/TGCN.2020.2965116

underlay CR system with filter-and-forward relays. Overlay systems with multiple relays have been analyzed in [25]. Furthermore, relay selection and optimal power allocation is analyzed for single antenna nodes [26]. In [27], cognitive mMIMO systems are analyzed for power allocation under pilot contamination. Furthermore, in [28], the performance of a MIMO secondary network under a MIMO primary network has been analyzed, with asymptotic results when the number of antennas at PU increases to infinity. However, this work does not analyze the effect of multiple relays in the secondary system. More recently, underlay mMIMO system have been analyzed in the presence of multiple primary MIMO nodes in [29]. This work analyses a secondary network with single-antenna users and a mMIMO base station. This work shows that a mMIMO PU can fully mitigate the co channel interference from the secondary network. However, none of these work focuses on relay selection for secondary mMIMO. Relay selection for cognitive full duplex one-way relay networks has been analyzed in [30] and relay selection for a full duplex energy harvesting system is analyzed in [31]. However, these result cannot be extended to two-way relay networks because of the interference constraints during the simultaneous data transmission of two nodes.

C. Problem Statement

Due to all these reasons, it makes sense to consider **an underlay TWRN** consisting of two mMIMO end nodes and a relay. However, several questions immediately arise. During the operation of the TWRN, interference on the PU is generated over two time slots. In the first one, both the end nodes generate it and in the second slot, the relay will generate it. How do we control this interference on the PU so that outage of the underlay network is avoided? What is the limiting behaviour of interference for a large number of antennas? Does relay selection help to reduce the interference? How do we scale down the transmit powers of the end nodes as a function of the number of antennas? What power allocation maximizes the sum rate while limiting the interference?

In order to answer some of these questions, we analyze the cognitive two-way relay selection problem of a particular network (Fig. 1). The source and destination nodes S_1 and S_2 (also called end nodes) are mMIMO enabled and the relay nodes (R_1 to R_k) are MIMO enabled. S_1 and S_2 perform zero forcing (ZF) based transmission and receiving while enabling multiple data sub-streams and the relay performs amplify-and-forwarding. Furthermore, S_1 and S_2 perform self-interference cancellation. S_1, S_2, and relay nodes all act as secondary users in the presence of a PU. Relay selection is performed to maximize the sum rate between S_1 and S_2 while limiting the interference on the PU. This system setup is ideal for Internet of Things (IoT) applications where networks are often deployed in an ad-hoc fashion. Such IoT networks operate in the underlay mode without affecting the licensed systems while improving energy efficiency [32].

In the considered network, the main interference issue arises during the first time slot when the two end nodes transmit simultaneously to the relay. This interference must be below the threshold at the PU. If not, the secondary system will go to outage. The key question is how to reduce or eliminate this outage. There are two immediate solutions. The first solution is to coordinate the transmissions of the two end nodes to reduce the joint interference. The problem is however that implementing coordination among two different end nodes is difficult. For instance, to do so, each end node will require the CSI between the other node and the PU. For example, S_1 needs to know the channel between S_2 and the PU node. This will require additional pilot overheads. The second solution is to use two separate time slots for S_1 and S_2 to transmit data to the relay. This solution avoids the cumulative interference, and it is proposed in [33], [34]. But the penalty is the 33% reduction of spectral efficiency as three time-slots are required for bidirectional data transfer.

Herein, we propose an alternative solution without constraining the end users to transmit in different time slots. It is based on the use of a large number of antennas, power scaling, and power allocation. Furthermore, relay selection is employed to obtain significant sum rate and energy efficiency gains. Specifically, our objective is to maximize the sum rate of an underlay TWRN through relay selection subject to the interference constraints on the PU.

D. Contributions:

- We obtain the signal-to-interference-plus-noise ratio (SINR) and the sum rate for a certain data sub-stream, when a certain relay is selected.
- We propose two possible power scaling scenarios to limit the interference at the PU. In power scaling one (PS-1), the transmit power at the end nodes and the relays are scaled down according to the number of antennas at the end nodes. In power scaling two (PS-2), only the power at the end nodes are scaled down according to the number of antennas at the end nodes.
- We derive asymptotic sum rates under the proposed power scaling scenarios and obtain the aggregated sum rate for the selected relay.
- We solve the optimum power allocation to maximize the sum rate and to satisfy the maximum interference constraint at PU for each relay under the two power scaling methods.
- We propose three relay selection criteria to maximize the sum rate. Specifically, we present one relay selection method for PS-1 and two relay selection methods for PS-2. We analyze these relay selection methods for their performance in terms of the achieved sum rate, outage of the secondary system, and the requirement of instantaneous channel information.
- We analyze the energy efficiency under PS-1 and PS-2 as the number of antennas at the end nodes are increased. For this analysis, we use a practical model which takes the power consumption in transceiver chains and the computational power consumption at the end nodes into account. This present work extends our related conference paper [1].

E. Organization and Notations

In Section II, we present the system, channel, and signal model and obtain the SINR and the sum rate for a given relay. In Section III, we present the required power scaling methods and obtain the asymptotic SINR and sum rates. By using these results, we allocate power to maximize the sum rate in Section IV. In Section V, we present relay selection strategies. In Section VI, we analyze the energy efficiency. Finally, Section VII compares the analytical and simulation results.

$\mathbf{Z}^{\mathrm{H}}$, $\mathbf{Z}^{\mathrm{T}}$, and $[\mathbf{Z}]_{k,k}$ denote the Hermitian-transpose, transpose, and the kth diagonal element of the matrix, $\mathbf{Z}$, respectively. A complex Gaussian random variable X with mean μ and standard deviation σ is denoted as $X \sim \mathcal{CN}(\mu, \sigma^2)$. $\mathbf{I}_M$ and $\mathbf{0}_{M\times N}$ are the $M \times M$ identity matrix and $M \times N$ matrix of all zeros, respectively. $\Gamma(z)$ is the Gamma function [35, Eqn. (8.310.1)], and $\Gamma(a, z)$ is the upper incomplete Gamma function [35, Eqn. (8.350.2)]. If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is a $p \times q$ matrix, then the Kronecker product is the $mp \times nq$ block matrix given as $\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$.

II. System, Channel, and Signal Model

A. System and Channel Model

Our system model consists of one PU and the secondary network (Fig. 1). Secondary TWRN consists of two user nodes (S_1 and S_2) and K relay nodes (R_k for $k \in \{1, \ldots, K\}$). PU is equipped with N antennas and user node S_i is equipped with N_i antennas for $i \in \{1, 2\}$, and the kth relay node has N_{R_k} antennas. All secondary nodes are assumed half-duplex terminals, and all channel amplitudes are assumed independently distributed, frequency flat-Rayleigh fading. The wireless channel from S_i to PU is defined as $\mathbf{F}_i = \hat{\mathbf{D}}_i^{1/2}\tilde{\mathbf{F}}_i$, where $\tilde{\mathbf{F}}_i \sim \mathcal{CN}_{N\times N_i}(\mathbf{0}_{N\times N_i}, \mathbf{I}_N \otimes \mathbf{I}_{N_i})$ captures the fast fading and $\hat{\mathbf{D}}_i = \hat{\eta}_i\mathbf{I}_N$ accounts for the pathloss. The channel between R_k and PU is defined as $\mathbf{G}_k = \mathbf{D}_k^{1/2}\tilde{\mathbf{G}}_k$, with $\tilde{\mathbf{G}}_k \sim \mathcal{CN}_{N\times N_{R_k}}(\mathbf{0}_{N\times N_{R_k}}, \mathbf{I}_N \otimes \mathbf{I}_{N_{R_k}})$ and $\mathbf{D}_k = \eta_k\mathbf{I}_N$. Similarly, the channel matrix from S_i to R_k is defined as $\mathbf{H}_{i,k} = \mathbf{D}_{i,k}^{1/2}\tilde{\mathbf{H}}_{i,k}$, with $\tilde{\mathbf{H}}_{i,k} \sim \mathcal{CN}_{N_{R_k}\times N_i}(\mathbf{0}_{N_{R_k}\times N_i}, \mathbf{I}_{N_{R_k}} \otimes \mathbf{I}_{N_i})$ and $\mathbf{D}_{i,k} = \eta_{i,k}\mathbf{I}_{N_{R_k}}$. The detailed system model is shown in Fig. 1. Here, the sets $\{\eta_k\}$, $\{\hat{\eta}_i\}$ and $\{\eta_{i,k}\}$ represent the pathloss components in the R_k-to-PU, S_i-to-PU and S_i-to-R_k channels.

The channel coefficients are assumed to be fixed during two consecutive time-slots (a time-slot is the time used for a single transmission between two wireless nodes), and hence, the reverse channels are assumed to be the transpose of forward channel by using the reciprocity property of wireless channels. The additive noise at all the receivers is modelled as complex zero mean additive white Gaussian (AWGN) noise. The direct channel between S_1 and S_2 is assumed to be unavailable due to large pathloss and heavy shadowing effects [7], [20].

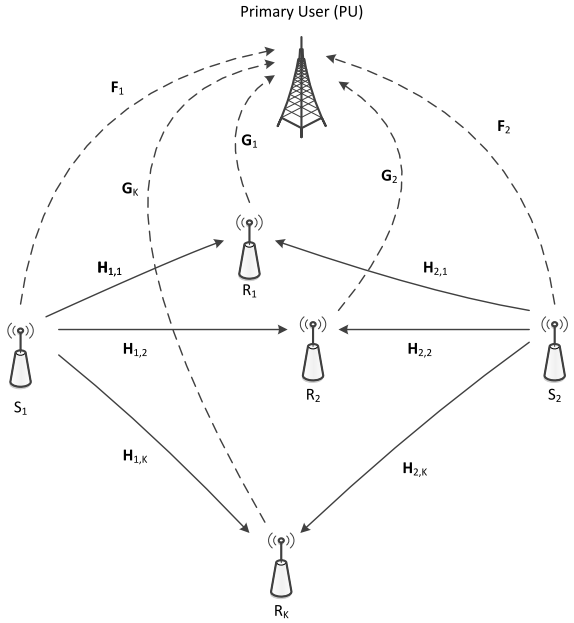

Fig. 1. The network with K relay nodes and two end nodes, S_1 and S_2.

B. Signal Model

S_1 and S_2 exchange their signal vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ by selecting one of the available relays using two time-slots. Here, we denote the selected relay as R_k for the analysis. First, S_1 and S_2 transmit $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, towards R_k by employing transmit-ZF precoding over the multiple access channel.[1] The received signal at R_k can then be written as

$$\mathbf{y}_{R_k} = m_{1,k}\mathbf{H}_{1,k}\mathbf{V}_{T_{1,k}}\mathbf{x}_1 + m_{2,k}\mathbf{H}_{2,k}\mathbf{V}_{T_{2,k}}\mathbf{x}_2 + \mathbf{n}_{R_k} + \mathbf{i}_{R_k}, \tag{1}$$

where the $N_{R_k} \times 1$ signal vector $\mathbf{x}_i$ satisfies $\mathcal{E}[\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}] = \mathbf{I}_{N_{R_k}}$ for $i \in \{1, 2\}$ and $k \in \{1, \ldots, K\}$. Thus, the $N_i \times 1$ precoded-transmit signal vector at S_i is given by $\mathbf{V}_{T_{i,k}}\mathbf{x}_i$. Here, $\mathbf{i}_{R_k}$ is the $N_{R_k} \times 1$ interference vector on the secondary relay by the PU, which is modelled as AWGN with average power $\sigma^2_{I_{R_k}} = P_U\eta_k$, where P_U is the transmit power of PU.[2] In (1), $m_{i,k}$ is the power normalizing factor at S_i and is designed to constrain its transmit power as follows [20]:

$$m_{i,k} = \sqrt{P_{i,k}/\mathrm{Tr}\Big(\mathbf{V}_{T_{i,k}}\mathbf{V}_{T_{i,k}}^{\mathrm{H}}\Big)}, \tag{2}$$

where $P_{i,k}$ is the transmit power at S_i to satisfy the interference constraints at PU. Further, in (1), $\mathbf{n}_{R_k}$ is the $N_{R_k} \times 1$ zero mean AWGN vector at R_k satisfying $\mathcal{E}(\mathbf{n}_{R_k}\mathbf{n}_{R_k}^{\mathrm{H}}) = \mathbf{I}_{N_{R_k}}\sigma^2_{N_{R_k}}$, and $\mathbf{V}_{T_{i,k}}$ is the transmit-ZF precoding matrix at S_i given by [36]

$$\mathbf{V}_{T_{i,k}} = \mathbf{H}_{i,k}^{\mathrm{H}}\Big(\mathbf{H}_{i,k}\mathbf{H}_{i,k}^{\mathrm{H}}\Big)^{-1}. \tag{3}$$

[1] In order to use transmit-ZF at S_1 and S_2, the constraint $\min(N_1, N_2) \geq \max_{k\in\{1,\ldots,K\}} N_{R_k}$ needs to be satisfied.

[2] Here, as we assume no coordination between the primary and secondary networks, we use the average interference power at the relay when the PU is transmitting. Similar assumptions are made in [26].

In the second time-slot, R_k amplifies $\mathbf{y}_{R_k}$ and broadcasts this amplified-signal towards both user nodes. Each node then performs ZF receiving and obtains the following signal vector:

$$\mathbf{y}_{S_{i,k}} = \mathbf{V}_{R_{i,k}}\big(M_k\mathbf{H}_{k,i}\mathbf{y}_{R_k} + \mathbf{n}_i + \mathbf{i}_i\big), \tag{4}$$

where M_k is the amplification factor at R_k and is defined as

$$M_k = \sqrt{P_{R_k}/\Big(m_{1,k}^2 + m_{2,k}^2 + \sigma_{N_{R_k}}^2 + \sigma_{I_{R_k}}^2\Big)}, \tag{5}$$

where $m_{i,k}$ is defined in (2) and P_{R_k} is the transmit power at R_k to satisfy the interference constraints at PU. Moreover, in (4), $\mathbf{H}_{k,i} = \mathbf{H}_{i,k}^T$, and $\mathbf{n}_i$ is the $N_i \times 1$ zero mean AWGN at S_i satisfying $\mathcal{E}(\mathbf{n}_i\mathbf{n}_i^{\mathrm{H}}) = \mathbf{I}_{N_i}\sigma_{N_i}^2$. Also, $\mathbf{i}_i$ is the interference on S_i by PU with average power $\sigma_{I_i}^2 = P_U\hat{\eta}_i$. Besides, $\mathbf{V}_{R_{i,k}}$ is the receive-ZF matrix at S_i and can be written as [36]

$$\mathbf{V}_{R_{i,k}} = \Big(\mathbf{H}_{k,i}^{\mathrm{H}}\mathbf{H}_{k,i}\Big)^{-1}\mathbf{H}_{k,i}^{\mathrm{H}}. \tag{6}$$

C. CSI Requirements

The CSI in the system can be categorized as (1) CSI within the secondary system and (2) CSI between the PU and the secondary system. Since we do not assume any coordination between the PU and the secondary system, acquiring type two CSI is harder than acquiring type one CSI (for related issues see [37]–[39] and references therein). The full CSI requirements can be listed as follows.

1) *CSI within the secondary system:* End node S_1 (S_2) require the CSI between it and the selected relay R_k, $\mathbf{H}_{1,k}$ ($\mathbf{H}_{2,k}$) to perform transmit beamforming and receive beamforming. The selected relay R_k only performs the AF operation. Thus, it will only require the information about the amplification coefficients at the end nodes $m_{1,k}$ and $m_{2,k}$. Although $m_{1,k}$ ($m_{2,k}$) depends on the CSI between the S_1 and R_k (S_2 and R_k), full CSI is not required at the R_k. For self interference cancellation, M_k should be known by both the end nodes. This knowledge can be acquired by the end nodes through a broadcast from the relay. Alternatively, all the nodes can decide to use the long term average values for $m_{i,k}$s and M_k. This will relax the CSI requirements on the system.
2) *CSI between the PU and secondary system:* Apart from the CSIs required for the data transmission between S_1 and S_2, the underlay operation requires all the nodes to limit the interference on the PUs. Thus when R_k transmits its signal, it should have the CSI between itself and the PU. Furthermore, when the end nodes transmit simultaneously in the first time slot, the end nodes must have the CSI between S_1 and PU and also between S_2 and PU (i.e., both $\mathbf{F}_1$ and $\mathbf{F}_2$ should be known by S_1 and S_2). This manuscript will show that this CSI is not required when massive MIMO and appropriate power scaling methods are used.

D. Effect on the Primary User

In underlay cognitive radio, the secondary users should transmit their data without exceeding the interference temperature at the PU. Thus, in this section we obtain the equations for the received interference at the PU during the two time-slots. During the first time-slot, the received interference signal at PU can then be written as

$$\mathbf{i}_{1,k} = m_{1,k}\mathbf{F}_1\mathbf{V}_{T_{1,k}}\mathbf{x}_1 + m_{2,k}\mathbf{F}_2\mathbf{V}_{T_{2,k}}\mathbf{x}_2. \tag{7}$$

Similarly, in the second time-slot, the received interference signal at PU can then be written as

$$\begin{aligned}\mathbf{i}_{2,k} &= M_k\mathbf{G}_k y_{R_k}\\ &= m_{1,k}M_k\mathbf{G}_k\mathbf{H}_{1,k}\mathbf{V}_{T_{1,k}}\mathbf{x}_1 + m_{2,k}M_k\mathbf{G}_k\mathbf{H}_{2,k}\mathbf{V}_{T_{2,k}}\mathbf{x}_2\\ &\quad + M_k\mathbf{G}_k\mathbf{n}_{R_k} + M_k\mathbf{G}_k\mathbf{i}_{R_k}.\end{aligned} \tag{8}$$

The interference constraints at the PU are given as

$$I_{1,k} = P_{1,k}\,\mathrm{Tr}\Big(\mathbf{F}_1^{\mathrm{H}}\mathbf{F}_1\Big) + P_{2,k}\,\mathrm{Tr}\Big(\mathbf{F}_2^{\mathrm{H}}\mathbf{F}_2\Big) \le I_t, \tag{9}$$

$$I_{2,k} = P_{R_k}\,\mathrm{Tr}\Big(\mathbf{G}_k^{\mathrm{H}}\mathbf{G}_k\Big) \le I_t, \tag{10}$$

where I_t is the interference threshold.

E. Outage of the Secondary System

The outage in the secondary TWRN with relay R_k will occur if the interference constraints (9) and (10) are not satisfied. Let the random variable $O_{S_k} = 1$ denote outage and $O_{S_k} = 0$ denote non outage of the secondary TWRN with relay R_k. The probability of outage can be defined as

$$P_{\mathrm{out,k}} = \Pr\big[O_{S_k} = 1\big] = \Pr\big[I_{1,k} > I_T \text{ or } I_{2,k} > I_T\big]. \tag{11}$$

Since $I_{1,k}$ and $I_{2,k}$ are independent of each other $P_{\mathrm{out,k}}$ can also be written as

$$P_{\mathrm{out,k}} = 1 - \Pr\big[I_{1,k} \le I_T\big]\Pr\big[I_{2,k} \le I_T\big]. \tag{12}$$

The outage of the whole system happens if all the relays are in a outage (i.e., $O_{S_k} = 1$ for all $k \in \{1,\ldots,K\}$). Let I_1 be the interference on the PU when the two end nodes transmit simultaneously in the first time slot. As I_1 is common for all the relays, this outage can be written as

$$P_{\mathrm{out}} = \Pr\big[I_1 > I_T \text{ or } \big(I_{2,1} > I_T \text{ and } \cdots I_{2,K} > I_T\big)\big]. \tag{13}$$

Since I_1, $I_{2,1},\ldots,I_{2,K}$ are independent of each other P_{out} can also be written as

$$\begin{aligned}P_{\mathrm{out}} &= 1 - \Pr[I_1 \le I_T]\\ &\quad\times\big(1 - \Pr\big[I_{2,1} \ge I_T\big]\cdots\Pr\big[I_{2,K} \ge I_T\big]\big).\end{aligned} \tag{14}$$

F. Exact Conditional End-to-End SINR

Theorem 1: The post processing end-to-end SINR of the *l*th data sub-stream at S_i under the transmit powers of $P_{1,k}, P_{2,k}$ and P_{R_k} when $O_{S_k} = 0$ and relay R_k is used, is given as

$$\gamma_{S_{i,k}^{(l)}|O_{S_k}=0} = \frac{M_k^2 m_{i',k}^2}{M_k^2\sigma_{R_k}^2 + \frac{\sigma_i^2}{\eta_{i,k}}\left[\Big(\tilde{\mathbf{H}}_{k,i}^{\mathrm{H}}\tilde{\mathbf{H}}_{k,i}\Big)^{-1}\right]_{l,l}}, \tag{15}$$

where $\{i,i'\} \in \{\{1,2\},\{2,1\}\}$, $l \in \{1,\ldots,N_{R_k}\}$, and $k \in \{1,\ldots,K\}$. Also we define the summed noise terms as $\sigma_{R_k}^2 = \sigma_{I_{R_k}}^2 + \sigma_{N_{R_k}}^2$ and $\sigma_i^2 = \sigma_{I_i}^2 + \sigma_{N_i}^2$.

Proof: By substituting (1), (3), and (6) into (4), the received signal vector at S_i can be written in an alternative form as follows:

$$\mathbf{y}_{S_{i,k}} = M_k\big(m_{i,k}\mathbf{x}_i + m_{i',k}\mathbf{x}_{i'} + \mathbf{n}_{R_k} + \mathbf{i}_{R_k}\big) + \tilde{\mathbf{n}}_i, \quad (16)$$

test where $\{i, i'\} \in \{\{1,2\},\{2,1\}\}$. Further, $\tilde{\mathbf{n}}_i$ is the filtered, colored noise and is given by $\tilde{\mathbf{n}}_i = \mathbf{V}_{R_{i,k}}(\mathbf{n}_i + \mathbf{i}_i)$. Next, by using self-interference cancellation to (16), the signal vector of $S_{i'}$ received at S_i can be extracted as follows:

$$\tilde{\mathbf{y}}_{S_{i,k}} = M_k\big(m_{i',k}\mathbf{x}_{i'} + \mathbf{n}_{R_k} + \mathbf{i}_{R_k}\big) + \tilde{\mathbf{n}}_i. \quad (17)$$

By using (17), the post-processing end-to-end SINR of the *l*th data sub-stream at S_i can be derived as (15). ■

Corollary 1: By substituting M_k (5), $m_{i',k}$ (2) and the result $\mathrm{Tr}(\mathbf{V}_{T_{i,k}}\mathbf{V}_{T_{i,k}}^{\mathrm{H}}) = \mathrm{Tr}([\mathbf{H}_{i,k}\mathbf{H}_{i,k}^{\mathrm{H}}]^{-1})$ into (15), the end-to-end conditional SINR in (15) can be written in a more insightful form as (18), shown at the bottom of this page.

G. Sum Rate Analysis When R_k is Selected

In this subsection, we obtain sum rate expressions when the relay R_k is selected assuming that the interference constraints at PU is satisfied. To correctly decode the data sub-streams by the receivers, each node needs to transmit the data with a common rate in MIMO TWRNs. Thus, the sum rate obtained by selecting the *k*th relay can be defined as follows:

$$\mathcal{R}_k = \begin{cases} 2\min\Big(\mathcal{R}_{S_{1,k}}, \mathcal{R}_{S_{2,k}}\Big), & \text{if } O_{S_k} = 0 \\ 0, & \text{if } O_{S_k} = 1 \end{cases} \quad (19)$$

where $\mathcal{R}_{S_i}$ is the sum of data sub-streams rates at $S_{k,i}$ for $i \in \{1,2\}$, and can be written as

$$\mathcal{R}_{S_{i,k}} = \begin{cases} \frac{1}{2}\sum_{l=1}^{N_{R_k}} \log\left(1 + \gamma_{S_{i,k}^{(l)}|O_{S_k}=0}\right), & \text{if } O_{S_k} = 0 \\ 0, & \text{if } O_{S_k} = 1 \end{cases} \quad (20)$$

The factor of two appears in (19) due to the presence of two user nodes in the TWRN of interest. Further, the pre-log factor of one-half in (20) is due to the use of two time-slots.

III. Asymptotic Analysis

We next provide the asymptotic analysis when the number of antennas at the end nodes goes to infinity (i.e., $N_1, N_2 \to \infty$). Specifically, while N_1 and N_2 grow unbounded, the number of antennas at relay R_k and PU remains constant. For simplicity, we assume that as $N_1, N_2 \to \infty$, the ratio between them remains constant:

$$\alpha = \frac{N_2}{N_1}. \quad (21)$$

We next obtain the asymptotic sum rate and the interference constraints for the two different power scaling scenarios.

A. Requirement of Power Scaling

Scaling the transmit power inversely with the number of antennas has been widely used in the literature to limit the total transmit power of a node. In [40] three power scaling scenarios has been analyzed for mMIMO two-way relay networks. These includes power scaling at the end nodes, power scaling at the relay, and power scaling at both the end and relay nodes. However, not all these power scaling methods can be used in an underlay setup. To clarify this situation, we rewrite the interference $I_{1,K}$ during the first time slot as

$$I_{1,k} = P_{1,k}\,\mathrm{Tr}\Big(\mathbf{F}_1^{\mathrm{H}}\mathbf{F}_1\Big) + P_{2,k}\,\mathrm{Tr}\Big(\mathbf{F}_2^{\mathrm{H}}\mathbf{F}_2\Big). \quad (22)$$

But if $P_{1,k}$ and $P_{2,k}$ are constants, as N_1 and N_2 become extremely large the value $\mathrm{Tr}(\mathbf{F}_1^{\mathrm{H}}\mathbf{F}_1)$ and $\mathrm{Tr}(\mathbf{F}_2^{\mathrm{H}}\mathbf{F}_2)$ reaches infinity (the asymptotic limit results in the Appendix). Thus without scaling down $P_{1,k}$ and $P_{2,k}$, the interference constraint at the PU can not be satisfied for a large number of transmit antennas. Thus, there are two ways to scale down $P_{1,k}$ and $P_{2,k}$, and they are analyzed in the next two subsections.

B. Power Scaling at the End Nodes and the Relay (PS-1)

Theorem 2: When, the transmit powers of user nodes S_1 and S_2 and relay nodes R_k for $k \in \{1,\ldots,K\}$ are scaled inversely with the number of antennas at the end nodes(i.e., PS-1) and condition (22) is satisfied, the asymptotic SINR is given as

$$\gamma_{S_{1,k}^{(l)}}^{\infty} = \frac{E_{R_k}E_{2,k}\eta_{1,k}\eta_{2,k}}{\sigma_1^2(E_{1,k}\eta_{1,k} + E_{2,k}\eta_{2,k}) + \sigma_{R_k}^2 N_{R_k}\big(E_{R_k}\eta_{1,k} + \sigma_1^2\big)} \quad (23)$$

and

$$\gamma_{S_{2,k}^{(l)}}^{\infty} = \frac{\alpha E_{R_k}E_{1,k}\eta_{1,k}\eta_{2,k}}{\sigma_2^2(E_{1,k}\eta_{1,k}+E_{2,k}\eta_{2,k})+\sigma_{R_k}^2 N_{R_k}\big(\alpha E_{R_k}\eta_{2,k}+\sigma_2^2\big)}. \quad (24)$$

Proof: The two end nodes and the relays can scale down their transmit power as follows:

$$P_{i,k} = \frac{E_{i,k}}{N_i} \text{ for } i \in \{1,2\} \text{ and } P_{R_k} = \frac{E_{R_k}}{N_1}, \quad (25)$$

where $E_{1,k}$, $E_{2,k}$ and E_{R_k} are fixed constants. To prove the asymptotic SINRs in (23) and (24), we first utilize the exact expression of SINR (18), which contains several matrix trace terms and the diagonal term of an inverse matrix. All these terms have asymptotic limits (see proof in the Appendix). ■

Interestingly, the asymptotic SINRs (23) and (24) are independent of the small-scale fading and only depend on the pathloss. Note that they are also independent of the data

$$\gamma_{S_{i,k}^{(l)}|O_{S_k}=0} = \frac{P_{R_k}P_{i',k}/\,\mathrm{Tr}\left(\left[\mathbf{H}_{i',k}\mathbf{H}_{i',k}^{\mathrm{H}}\right]^{-1}\right)}{P_{R_k}\sigma_{R_k}^2 + \left(\frac{P_{1,k}}{\mathrm{Tr}\left(\left[\mathbf{H}_{1,k}\mathbf{H}_{1,k}^{\mathrm{H}}\right]^{-1}\right)} + \frac{P_{2,k}}{\mathrm{Tr}\left(\left[\mathbf{H}_{2,k}\mathbf{H}_{2,k}^{\mathrm{H}}\right]^{-1}\right)} + \sigma_{R_k}^2\right)\frac{\sigma_i^2}{\eta_{i,k}}\left[\left(\tilde{\mathbf{H}}_{k,i}^{\mathrm{H}}\tilde{\mathbf{H}}_{k,i}\right)^{-1}\right]_{l,l}} \quad (18)$$

sub-stream index, l, and hence, we can denote $\gamma^{\infty}_{S^{(l)}_{i,k}} = \gamma^{\infty}_{S_{i,k}}$ for $l \in \{1, \ldots, N_{R_k}\}$.

We next show that the interference constraints are satisfied during the two slots given the power scaling scenario (25). First, let us consider the second time slot where only the selected relay is transmitting. By substituting $P_{i,k}$ and P_{R_k} we rewrite (10) as

$$I_{2,k} = E_{R_k} \frac{\mathrm{Tr}\big(\mathbf{G}_k^{\mathrm{H}} \mathbf{G}_k\big)}{N_1} \leq I_t, \tag{26}$$

and as $\frac{\mathrm{Tr}(\mathbf{G}_k^{\mathrm{H}} \mathbf{G}_k)}{N_1} \to 0$ as $N_1 \to \infty$, this condition is asymptotically satisfied for any value of E_{R_k}. Second, let us consider the first time slot where the two end nodes are transmitting. By using the limits on (22), we obtain the interference constraint as (see the Appendix)

$$\lim_{N_i \to \infty} I_{1,k} = E_{1,k} \hat{\eta}_1 N + E_{2,k} \hat{\eta}_2 N \leq I_t. \tag{27}$$

The constraint (27) does not contain any random quantities. Thus, it can be strictly satisfied by carefully selecting $E_{1,k}$ and $E_{2,k}$ at the end nodes of the secondary network. Therefore, the secondary system can operate without an outage (i.e., $P_{out} = 0$).

By using asymptotic SINR (23) and (24), we obtain the sum rate between U_i and R_k as

$$\mathcal{R}^{\infty}_{S_{i,k}} = \frac{1}{2} N_{R_k} \log\Big(1 + \gamma^{\infty}_{S_{i,k}}\Big). \tag{28}$$

Further, we obtain the total sum rate when R_k is selected as

$$\mathcal{R}^{\infty}_{k} = N_{R_k} \log\Big(1 + \min\Big(\gamma^{\infty}_{S_{1,k}}, \gamma^{\infty}_{S_{2,k}}\Big)\Big). \tag{29}$$

Remark 1: The results obtained in (23) and (24), only depend on the pathloss component and the average noise values. Thus, the total sum rate will be a fixed value for a given relay. Thus, relay selection can be done off-line and no instantaneous channel state is necessary for the relay selection process. Furthermore, outage $P_{out,k}$ can be set to zero by scaling the transmit powers of S_1 and S_2.

C. Power Scaling at the End Nodes - (PS-2)

Theorem 3: The asymptotic SINR when the transmit power at the user nodes S_1 and S_2 are scaled inversely proportional to the number of antennas at the user nodes (i.e., PS-2) is given as

$$\gamma^{\infty}_{S^{(l)}_{1,k}} = \frac{E_{2,k} \eta_{2,k}}{N_{R_k} \sigma^2_{R_k}} \tag{30}$$

and

$$\gamma^{\infty}_{S^{(l)}_{2,k}} = \frac{E_{1,k} \eta_{1,k}}{N_{R_k} \sigma^2_{R_k}}. \tag{31}$$

Proof: The transmit power scaling can be given as

$$P_{i,k} = \frac{E_{i,k}}{N_i} \text{ for } i \in \{1, 2\} \text{ and } \quad P_{R_k} = E_{R_k}, \tag{32}$$

where $E_{1,k}$, $E_{2,k}$ and E_{R_k} are fixed. The equations for SINRs are obtained by substituting $P_{i,k}$ and P_{R_k} values and by using the asymptotic limit results on (18). Proofs are omitted due to the similarity to the proof of SINRs in the Appendix. ■

Remark 2: Similar to (23) and (24), the SINR values in (30) and (31) only depend on the pathloss component and the average noise values. Furthermore, they are independent of the interference and noise levels at the end nodes (σ_i^2).

Proposition 1: The average sum rate of the secondary system under PS-2 is given as follows:

$$\begin{aligned} \bar{\mathcal{R}}^{\infty}_{k} &= (1 - P_{\mathrm{out,k}}) N_{R_k} \log\Big(1 + \min\Big(\gamma^{\infty}_{S_{1,k}}, \gamma^{\infty}_{S_{2,k}}\Big)\Big) \\ &= \left(1 - \frac{\Gamma\Big(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}}\Big)}{\Gamma\big(NN_{R_k}\big)}\right) N_{R_k} \\ &\quad \times \log\left(1 + \min\left(\frac{E_{2,k}\eta_{2,k}}{N_{R_k}\sigma^2_{R_k}}, \frac{E_{1,k}\eta_{1,k}}{N_{R_k}\sigma^2_{R_k}}\right)\right). \end{aligned} \tag{33}$$

Proof: As mentioned before, (27) provides the interference constraint during the first time-slot. This constraint can be satisfied by selecting $E_{1,k}$ and $E_{2,k}$ appropriately. However, the PU interference from the selected relay during the second time slot must also satisfy the interference constraint:

$$I_{2,k} = E_{R_k} \mathrm{Tr}\Big(\mathbf{G}_k^{\mathrm{H}} \mathbf{G}_k\Big) = \eta_k E_{R_k} \mathrm{Tr}\Big(\tilde{\mathbf{G}}_k^{\mathrm{H}} \tilde{\mathbf{G}}_k\Big) \leq I_t. \tag{34}$$

Unlike the PS-1 case, this is a random event, whose likelihood depends on channel quality between the selected relay and the PU. Thus the outage of the secondary system (i.e., $P_{out,k}$) is unpredictable. The outage probability may be expressed as

$$P_{\mathrm{out,k}} = \Pr\big[I_{2,k} \geq I_t\big] = \Pr\left[\mathrm{Tr}\Big(\tilde{\mathbf{G}}_k^{\mathrm{H}} \tilde{\mathbf{G}}_k\Big) \geq \frac{I_t}{\eta_k E_{R_k}}\right]. \tag{35}$$

It is well known that $\mathrm{Tr}(\tilde{\mathbf{G}}_k^{\mathrm{H}} \tilde{\mathbf{G}}_k)$ is Gamma distributed with shape parameter of $N \times N_{R_k}$ [41]. Thus $P_{out,k}$ can be expressed as

$$P_{\mathrm{out,k}} = \frac{1}{\Gamma\big(NN_{R_k}\big)} \Gamma\Big(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}}\Big). \tag{36}$$

By using the obtained SINR values, the sum rate can be written as

$$\mathcal{R}^{\infty}_{k} = \begin{cases} N_{R_k} \log\Big(1 + \min\Big(\gamma^{\infty}_{S_{1,k}}, \gamma^{\infty}_{S_{2,k}}\Big)\Big), & \text{if } O_{S_k} = 0 \\ 0, & \text{if } O_{S_k} = 1, \end{cases} \tag{37}$$

and (33) can be obtained. ■

IV. Optimal Power Allocation

We develop optimal power allocation to maximize the sum rates given in (29) and (37) under the interference constraint in (27) for PS-1 and PS-2. We represent the optimal power allocations at S_1 and S_2 as $E^*_{1,k}$ and $E^*_{2,k}$. Furthermore, we assume that the transmit power at the selected relay E_{R_k} is a fixed value for the following two cases.

A. Power Scaling at the End Nodes and the Relay (PS-1)

Maximizing R^{∞}_k in (29) corresponds to maximizing the minimum of $\gamma^{\infty}_{S_{1,k}}$ and $\gamma^{\infty}_{S_{2,k}}$ by selecting $E_{1,k}$ and $E_{2,k}$ which satisfies (27). It can be seen that both $\gamma^{\infty}_{S_{1,k}}$ and $\gamma^{\infty}_{S_{2,k}}$ are

increasing with respect to $E_{1,k}$ and $E_{2,k}$. Thus, the maximum of R_k^{∞} occurs when (27) becomes an equality. Thus

$$E_{1,k}^{*} = \frac{1}{\hat{\eta}_1}\left(\frac{1}{(1+\delta)N}I_t - E_{2,k}^{*}\hat{\eta}_2\right), \tag{38}$$

where $0 \leq \delta \ll 1$ is used to make sure that the interference on PU is significantly less than the threshold interference level. Furthermore, it can be seen that the maximum value for the minimum between two asymptotic SINR values occur when $\gamma_{S_{1,k}}^{\infty} = \gamma_{S_{2,k}}^{\infty}$. Thus, we obtain the following equation:

$$\begin{aligned}\sigma_2^2\eta_{2,k}E_{2,k}^{*2} - \alpha\sigma_1^2\eta_{1,k}E_{1,k}^{*,2} + \left(\sigma_2^2\eta_{1,k} - \alpha\sigma_1^2\eta_{2,k}\right)E_{1,k}^{*}E_{2,k}^{*} \\ + \left(\alpha E_{R_k}\eta_{2,k} + \sigma_2^2\right)N_{R_k}\sigma_{R_k}^2E_{2,k}^{*} \\ - \alpha\left(E_{R_k}\eta_{1,k} + \sigma_1^2\right)N_{R_k}\sigma_{R_k}^2E_{1,k}^{*} = 0.\end{aligned} \tag{39}$$

Then solving (38) and (39) we can obtain the optimal values $E_{1,k}^{*}$ and $E_{2,k}^{*}$. By using these values we can obtain $\gamma_{S_{1,k}}^{\infty,*}$, $\gamma_{S_{1,k}}^{\infty,*}$, and $\mathcal{R}_k^{\infty,*}$ which are the asymptotic optimal values of SINRs and sum rate, respectively.

As an special case, if we assume that $\alpha = 1$ and $\eta_{1,k} = \eta_{2,k}$, then we can conclude that $E_{1,k}^{*} = E_{2,k}^{*}$. And the optimum power allocation values are given by using (38), as

$$E_{1,k}^{*} = E_{2,k}^{*} = \frac{1}{(1+\delta)(\hat{\eta}_1+\hat{\eta}_2)N}I_t. \tag{40}$$

B. Power Scaling at the End Nodes (PS-2)

In this section we provide the optimal power allocation scenario for the power scaling at the end nodes case. Unlike in the previous scenario, the sum rate also depends on the outage probability $P_{out,k}$ which depends on the transmit power at the relay node P_{R_k}. However, the asymptotic SINR values in (30) and (31) are independent of E_{R_k}. Thus, similar to the previous case this corresponds to the case where the equality condition in (27) is satisfied. Thus,

$$\begin{aligned}\gamma_{S_{1,k}^{(l)}}^{\infty} &= \gamma_{S_{2,k}^{(l)}}^{\infty} \\ \frac{E_{2,k}\eta_{2,k}}{N_{R_k}\sigma_{R_k}^2} &= \frac{E_{1,k}\eta_{1,k}}{N_{R_k}\sigma_{R_k}^2}.\end{aligned} \tag{41}$$

By using value for $E_{1,k}^{*}$ from (38) in (41), we obtain the following:

$$E_{1,k}^{*} = \frac{\eta_{2,k}}{(1+\delta)\left(\hat{\eta}_1\eta_{2,k} + \hat{\eta}_2\eta_{1,k}\right)N}I_t, \tag{42}$$

$$E_{2,k}^{*} = \frac{\eta_{1,k}}{(1+\delta)\left(\hat{\eta}_1\eta_{2,k} + \hat{\eta}_2\eta_{1,k}\right)N}I_t. \tag{43}$$

Thus, the optimal average sum rate is written as

$$\begin{aligned}\bar{\mathcal{R}}_k^{\infty,*} = \left(1 - \frac{\Gamma\left(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}}\right)}{\Gamma\left(NN_{R_k}\right)}\right)N_{R_k} \\ \times \log\left(1 + \frac{\eta_{1,k}\eta_{2,k}}{(1+\delta)(\hat{\eta}_1\eta_{2,k} + \hat{\eta}_2\eta_{1,k})N_{R_k}N\sigma_{R_k}^2}I_t\right).\end{aligned} \tag{44}$$

V. Relay Selection With Optimum Values

We now present the relay selection criteria for the two power scaling strategies PS-1 and PS-2. For both these cases, the objective of relay selection is to maximize the achievable asymptotic sum rate of the secondary system. The relay selection criterion may be stated as:

$$K^{*} = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}}\left[\mathcal{R}_k^{\infty,*}\right], \tag{45}$$

where K^{*} is the index of the selected relay. In the next two subsections, we analyze the relay selection criteria under PS-1 and PS-2.

A. Power Scaling at the End Nodes and the Relay (PS-1)

The relay selection can further be simplified to the following:

$$K^{*} = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}}\left[\mathcal{R}_k^{\infty,*}\right] = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}}\left[N_{R_k}\log\left(1+\gamma_{S_{1,k}}^{\infty}\right)\right], \tag{46}$$

Thus, relay selection is deterministic as all the $R_k^{\infty,*}$ are fixed values. Further, (46) shows that the choice of optimal relay highly depends on the maximum value of N_{R_k}. Since the log of $(1+\gamma_{S_{1,k}}^{\infty})$ is taken, the impact of N_{R_k} is higher than that of the logarithm. Thus, for most cases the relay selection simply boils down to the selection of the relay with the maximum number of antennas.

B. Power Scaling at the End Nodes (PS-2)

Unlike in the previous case, here the sum rate of R_k^{∞} and outage are inter-related. Since outage is random, the relay selection will no longer be a deterministic process. Yet, motivated by the simplicity of relay selection for deterministic sum rates (46), we propose two relay selection methods where, one of them does not require the instantaneous channel knowledge. The proposed relay selection methods are to

1) Maximize the instantaneous asymptotic sum rate (SS-1),
2) Maximize the average sum rate (SS-2).

We next discuss the advantages and disadvantages of these two relay selection schemes, and their performances are compared in Section VII.

1) Maximize the Instantaneous Sum Rate: The relay selection criterion may be stated as

$$K^{*} = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}}\left[\mathcal{R}_k^{\infty,*}\right]. \tag{47}$$

However, when some of the relays are in outage, the optimal selected relay will be among the relays that are not in outage. Thus, in this case relay selection clearly provides higher sum rates and lower outage probabilities. However, the knowledge of the channel states is necessary for the relay selection process, whose complexity will thus increase.

As we can ensure $\Pr[I_1 \leq I_T] = 1$ with power allocation, by using the outages related to $I_{2,k}$s from (36), we can rewrite (14) as

$$P_{\text{out}} = \prod_{k=1}^{K}\frac{\Gamma\left(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}}\right)}{\Gamma\left(NN_{R_k}\right)}. \tag{48}$$

As a special case, if all the relay nodes are identical (i.e., $N_{R_k} = N_R$ and $E_{R_k} = E_R$ for all $k \in (1, \ldots, K)$), the average sum rate with relay selection is given as[3]

$$\bar{\mathcal{R}}^{\infty,*} = \left(1 - \frac{\prod_{k=1}^{K} \Gamma\left(NN_R, \frac{I_t}{\eta_k E_R}\right)}{\Gamma(NN_R)^K}\right) N_R \times \log\left(1 + \frac{\eta_{1,k}\eta_{2,k}}{(1+\delta)\left(\hat{\eta}_1\eta_{2,k} + \hat{\eta}_2\eta_{1,k}\right) N_R N \sigma_R^2} I_t\right). \tag{49}$$

Eq. (48) shows that relay selection significantly decreases the outage probability and hence increases the average sum rate compared to a single relay system. However, instantaneous channel knowledge for the relay selection will cost bandwidth, energy and complexity. The reason is that channel estimation requires the transmission of pilot sequences and channel state feed back to the transmit node and to the relay selector.

2) Maximize the Average Sum Rate: For this case, the relay selection equation can be given as

$$K^* = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}} \left[\bar{\mathcal{R}}_k^{\infty,*}\right] = \underset{k\in\{1,\ldots,K\}}{\operatorname{argmax}} \left[\left(1 - \frac{\Gamma\left(NN_{R_k}, \frac{I_t}{\eta_k E_{R_k}}\right)}{\Gamma\left(NN_{R_k}\right)}\right) N_{R_k} \times \log\left(1 + \frac{\eta_{1,k}\eta_{2,k}}{(1+\delta)\left(\hat{\eta}_1\eta_{2,k} + \hat{\eta}_2\eta_{1,k}\right) N_{R_k} N \sigma_{R_k}^2} I_t\right)\right]. \tag{50}$$

Since we use the average sum rate which is a deterministic value, the relay selection becomes deterministic. Thus the relay selection complexity for this scenario is significantly less than the previous case as the instantaneous channel conditions of all the relays are not required for the relay selection. However, the channel conditions of the required relay is still required for the beamforming and interference control. Similar to the power scaling at the end nodes and the relay case, the number of antennas at the relay N_{R_k} has the highest impact on the relay selection.

As a special case, if we consider the case with identical relays, the average sum rate is given by

$$\bar{\mathcal{R}}_k^{\infty,*} = \left(1 - \frac{\Gamma\left(NN_R, \frac{I_t}{\eta_k E_R}\right)}{\Gamma(NN_R)}\right) N_R \times \log\left(1 + \frac{\eta_{1,k}\eta_{2,k}}{(1+\delta)\left(\hat{\eta}_1\eta_{2,k} + \hat{\eta}_2\eta_{1,K}\right) N_R N \sigma_R^2} I_t\right). \tag{51}$$

And it is evident from (51) and (49) that the increased sum rate in the previous case is due to the increased relay selection complexity.

Remark 3: As evident from the above two cases, the number of antennas at a given relay N_{R_k} has the highest impact on the achievable sum rate of a system. Thus, if all other parameters are approximately equal, then selecting the relay with the highest number of antennas will provide the highest sum rate for the secondary system.

Remark 4: It can be seen that the sum rate $R_k^{\infty,*}$ is a deterministic variable. Thus, the relay selection also becomes a deterministic task. Thus, relay selection can be done during the system design stage. That is, the end and relay nodes need not implement real-time relay selections.

VI. Energy Efficiency Analysis

In Section IV, we analyzed the transmit power allocations at S_1, S_2 and the relay to help control the PU interference. Here, we analyze the energy efficiency under these conditions.

Some analyses are based on the assumption that the power consumption at the nodes is only due to the wireless data transfers between nodes. In this case, energy efficiency of PS-1 approaches infinity and the energy efficiency of PS-2 approaches an asymptotic limit. However, the processing done at the nodes consumes more than 50% of the total in mMIMO [3]. Thus, both power consumption at circuit components and that for ZF computations done at the end nodes must be considered. Therefore, a more accurate mMIMO energy efficiency analysis is proposed in [42]. According to this, the power consumption during time-slot 1, $P_{\mathrm{tot,ts1}}$ can be given as

$$P_{\mathrm{tot,ts1}} = P_{1,k} + P_{2,k} + P_{\mathrm{ts1,TC}} + P_{\mathrm{ts1,C}} + P_{1,k,\mathrm{LP}} + P_{2,k,\mathrm{LP}}, \tag{52}$$

where $P_{i,k}$ is the transmit power at the *i*th end node, $P_{\mathrm{ts1,TC}}$ is the power consumed in transceiver chains in time-slot 1, $P_{\mathrm{ts1,C}}$ is the power consumed for coding in time-slot 1, and $P_{i,k,\mathrm{LP}}$ is the power consumed in node *i* for linear processing (in this case ZF). The values for these power consumption components are given as follows:

$$P_{\mathrm{ts1,TC}} = N_1 P_{\mathrm{EN}} + N_2 P_{\mathrm{EN}} + K P_{\mathrm{Rel}} + P_{\mathrm{SYN}}, \tag{53}$$

$$P_{\mathrm{ts1,C}} = 2\mathcal{R}_k P_{\mathrm{COD}}, \tag{54}$$

$$P_{i,k,\mathrm{LP}} = \frac{2BN_i N_{\mathrm{R_k}}}{L_{\mathrm{BS}}} + \frac{B}{U}\left(\frac{N_{R_k}^3}{3L_{\mathrm{BS}}} + \frac{3N_i N_{R_k}^2 + N_i N_{R_k}}{L_{\mathrm{BS}}}\right), \tag{55}$$

where P_{EN} and P_{Rel} are the powers required to run the circuit components at the end nodes and relays, P_{SYN} is the power of the local oscillator, P_{COD} is the coding power consumption at the end nodes, *B* is the bandwidth, L_{BS} is the computational efficiency (given in flops/W) of the end nodes, and *U* is the coherence block. Similarly, the power consumption during time-slot 2 is given as

$$P_{\mathrm{tot,ts2}} = P_{R_k} + P_{\mathrm{ts2,TC}} + P_{\mathrm{ts2,D}} + P_{k,1,\mathrm{LP}} + P_{k,2,\mathrm{LP}}, \tag{56}$$

where $P_{\mathrm{ts2,TC}} = P_{\mathrm{ts1,TC}}$, $P_{k,i,\mathrm{LP}} = P_{i,k,\mathrm{LP}}$, and

$$P_{\mathrm{ts2,D}} = 2\mathcal{R}_k P_{\mathrm{DEC}}, \tag{57}$$

where P_{DEC} is the decoding power consumption at the end nodes. Based on theses values the average energy efficiency

[3]Even in this case, the parameter values $\eta_{1,k}$ and $\eta_{2,k}$ will be different. However, for simplification we have assumed $\eta_{1,k}$ and $\eta_{2,k}$ are approximately equal for all *k*.

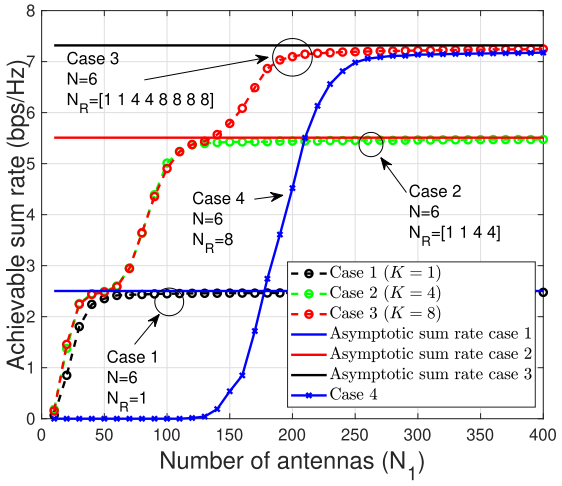

Fig. 2. Sum rate as a function of the number of S_1 antennas.

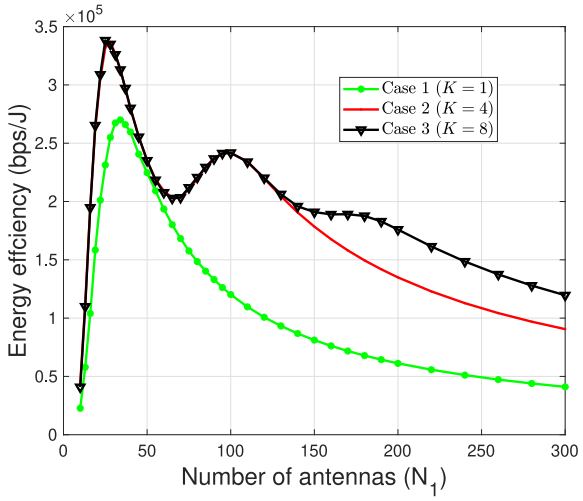

Fig. 3. Energy efficiency comparisons as function of the number of S_1 antennas for PS-1.

of the secondary system is defined as

$$\rho_k = \frac{\mathcal{R}_k}{P_{\text{tot,ts1}} + P_{\text{tot,ts2}}}. \tag{58}$$

Based on (58), the energy efficiency under PS-1 can be written as (59), shown at the bottom of this page. Here in (59), $P_{1,k,\text{LP}}$ and $P_{2,k,\text{LP}}$ is given by (55). Note that the consumed power does not go to zero as the number of antennas are increased. Instead, the consumed power increase with N_1 and N_2 due to the second term in the above equation. Similarly, the energy efficiency under PS-2 is written as (60), shown at the bottom of this page.

VII. Numerical Results

Herein, we present numerical and simulation results of the proposed selection strategies. Specifically, the average and asymptotic sum rates are computed and simulated. The interference threshold at PU is $I_T = 10$ dB, the pathloss components between PU and S_1 and S_2 as $\hat{\eta}_1 = \hat{\eta}_2 = 1/16$, and the noise powers $\sigma^2_{N_i} = \sigma^2_{R_k} = 1$. We also set $\alpha = 1$ and transmit power at the relay R_k as $E_k = 25$ dBm. The primary user has 6 antennas ($N = 6$). Figs. 2 to 7 represent the results for PS-1 while the Figs. 8 to 12 show the simulation results for PS-2.

In Fig. 2, we analyze four cases with different number of relays for end-node and relay power scaling. Case∼1 consists of a single relay node ($K = 1$) with a single antenna. In Case∼2, $K = 4$ and the numbers of relay antennas (N_{R_k}) are 1, 1, 4, 4, respectively. In Case∼3, $K = 8$ and the numbers of relay antennas are 1, 1, 4, 4, 8, 8, 8, 8, respectively. In Case∼4, we plot the sum rate, when the relay with the highest number of antennas is selected (relay with 8 antennas). We use the optimal power allocation scheme with $\delta = 0.1$. We see that multiple relays bring about significant sum rate gains. For example a single relay with a single antenna can only achieve a sum rate of 2.5bps/Hz while the four-relay system, which has 1, 1, 4, and 4 antennas each can obtain a sum rate of 5.5bps/Hz. Also, with eight relays, the achievable sum rate is 7.2bps/Hz, which is a significant increase compared to previous two cases. Our asymptotic analysis perfectly matches the simulated sum rates when the number of antennas are increased. Selecting the relay with the highest number of antennas (Case∼4) provides the same asymptotic sum rate as Case∼3, which uses relay selection. However, without relay selection, more antennas are required to achieve asymptotic performance.

In Fig. 3, we plot the energy efficiency given in (59) for system configuration cases that were analyzed in the previous figure. The parameter values for P_{EN}, P_{Rel}, P_{SYN}, P_{COD}, P_{DEC}, B, U, and L_{BS} are adopted from [42]. For Case∼1, the energy efficiency increases as the number of antennas are increased up-to around 40. However, after that, energy efficiency reduces as the number of antennas are increased and asymptotically reach zero. For Case∼2 and Case∼3, there are multiple local maxima for the energy efficiency. The reason behind this is that different relays have different number of antennas (i.e., for Case∼3, two relays have single antennas and two relays have 4 antennas and the other relays have 8 antennas). Regardless of the number of antennas available at the relays, more relays always improve energy efficiency.

To investigate the effect of identical antenna configurations, we plot the energy efficiency when all the relays have 8 antennas for PS-1 in Fig. 4. It clearly shows that more relays

$$\rho_{k,\text{PS1}} = \frac{\mathcal{R}_k}{\frac{E_{1,k}}{N_1} + \frac{E_{2,k}}{N_2} + \frac{E_{R_k}}{N_1} + 2((N_1+N_2)P_{\text{EN}} + KP_{\text{Rel}} + P_{\text{SYN}}) + 2\mathcal{R}_k(P_{\text{COD}} + P_{\text{DEC}}) + 2P_{1,k,\text{LP}} + 2P_{2,k,\text{LP}}} \tag{59}$$

$$\rho_{k,\text{PS2}} = \frac{\mathcal{R}_k}{\frac{E_{1,k}}{N_1} + \frac{E_{2,k}}{N_2} + E_{R_k} + 2((N_1+N_2)P_{\text{EN}} + KP_{\text{Rel}} + P_{\text{SYN}}) + 2\mathcal{R}_k(P_{\text{COD}} + P_{\text{DEC}}) + 2P_{1,k,\text{LP}} + 2P_{2,k,\text{LP}}} \tag{60}$$

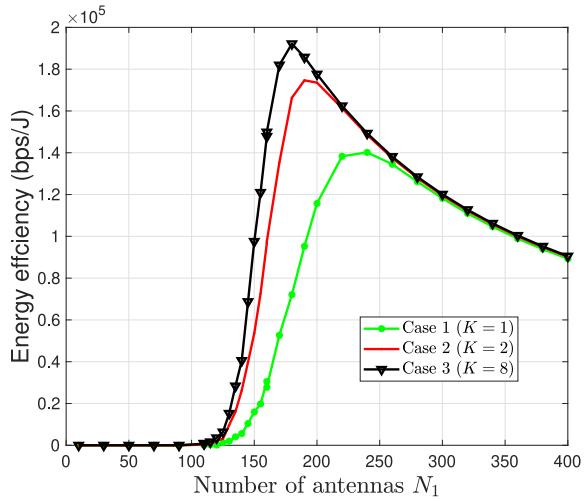

Fig. 4. Energy efficiency as a function of the number of S_1 antennas. All the relays are equipped with 8 antennas each.

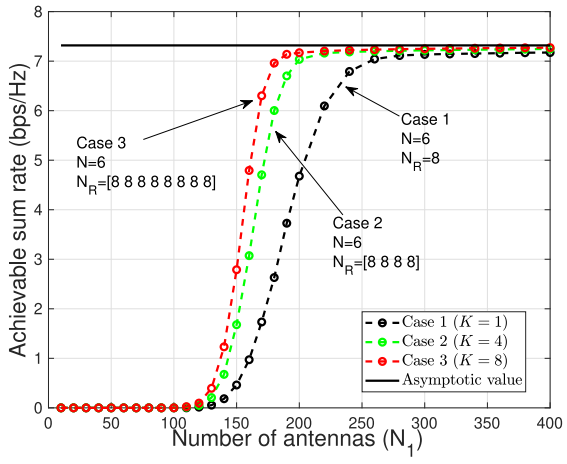

Fig. 5. Sum rate comparisons as a function of the number of S_1 antennas. All relays are equipped with 8 antennas.

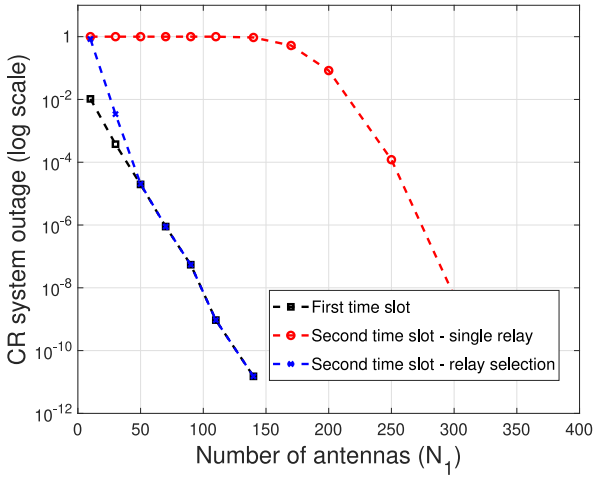

Fig. 6. CR TWRN outage during different time slots with and without relay selection.

improve energy efficiency at any number of antennas at S_1 and also, the peak energy efficiency value can be obtained for fewer antennas at S_1. For instance, the peak energy efficiency of 1.9×10^5bps/J can be obtained with around 170 antennas at the end nodes for Case∼3 while the peak energy efficiency of 1.7×10^5bps/J can be obtained with around 190 antennas for Case∼2. The energy efficiency increase with the number of relays (especially in the low antenna regime) is due to the fact that more relays increase the sum rate of the system.

In Fig. 5, we analyze the sum rate with different number of relays, but when those relays have the same number of antennas ($N_{R_k} = 8$ for all the relays). Cases one, two, and three corresponds to 1, 4, and 8 relays respectively. Optimal power allocation with $\delta = 0.1$ is used. Unlike in Fig. 2, we can see that the achievable asymptotic sum rates for different cases are the same. The reason is that the achievable sum rate depends on the relay with the highest number of antennas. However, with multiple relays, the asymptotic performance can be obtained by using a smaller number of antennas at the end nodes. Also it can be seen that until the number of antennas at the end nodes surpasses a threshold (in this case around 120), the achievable sum rate is zero. This is due to the interference constraint at the PU. Although (22) is satisfied asymptotically for any E_{R_k}, when N_1 is low this constraint may not be satisfied and the secondary system will remain in outage state. Thus, until the number of antennas at the end nodes increases to a certain limit, the secondary network cannot start the transmission.

In Fig. 6, we have plotted the outage when the PU interference exceeds the threshold. We have plotted (1) interference during the first time-slot (when both the end nodes are transmitting) causes the outage, (2) interference during the second time-slot causes the outage when having 8 antennas, and (3) interference during the second time-slot causes the outage when the relay with the highest number of antennas is selected. As seen from Fig. 6, the first-time slot outage rapidly reduces with the number of antennas. For instance, with 200 antennas the outage is less than 10^{-12}. This shows that, with a large antenna array at the end nodes and with power scaling, the interference on the PU approaches zero and thus, the end nodes do not need to know each others channel matrices to decide whether to transmit or not. Also, the interference during the second time-slot approaches zero but with a lower convergence rate. As an example with 250 antennas, relay selection provides less than 10^{-12} outage while the outage with a single relay is around 10^{-5}.

In Fig. 7, we plot the achievable sum rates for three cases for different power allocations at the end nodes. In Case∼1, we use the derived optimal power allocation scheme in Section IV-A. In Case∼2, we use $\frac{E_1}{E_2} = 3/2$ and in Case∼3, we use $\frac{E_1}{E_2} = 7/3$. It can be seen from the plot that the sum rate approaches the asymptotic sum rates we obtained through equation (29) when the number of antennas at the end nodes increases. Furthermore, our optimal power allocation scheme in Section IV-A obtains the highest achievable sum rate while the sum rates of other power allocations are significantly less than that for the optimal power allocation.

In Figs. 8, 9, and 10 we plot the outage, sum rate, and the energy efficiency for end-node power scaling (PS-2)

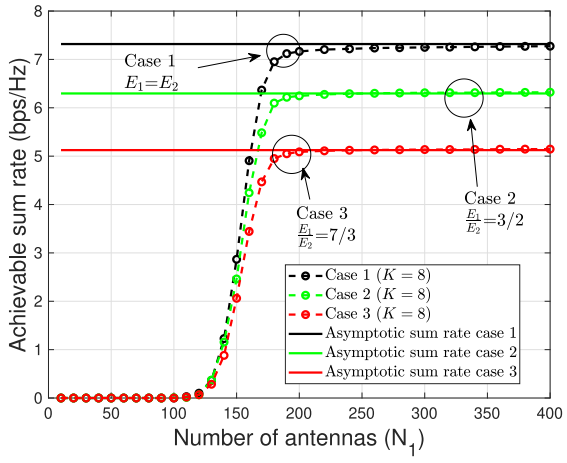

Fig. 7. The achievable sum rate against the number of antennas at the end nodes for different power allocation cases.

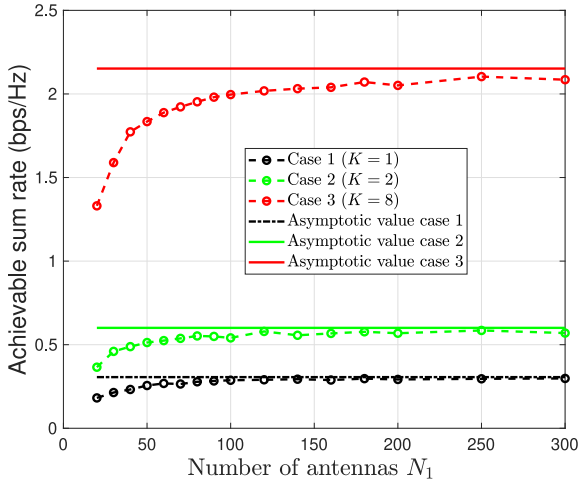

Fig. 9. Achievable sum rate vs the number of antennas at S_1 for end-node power scaling.

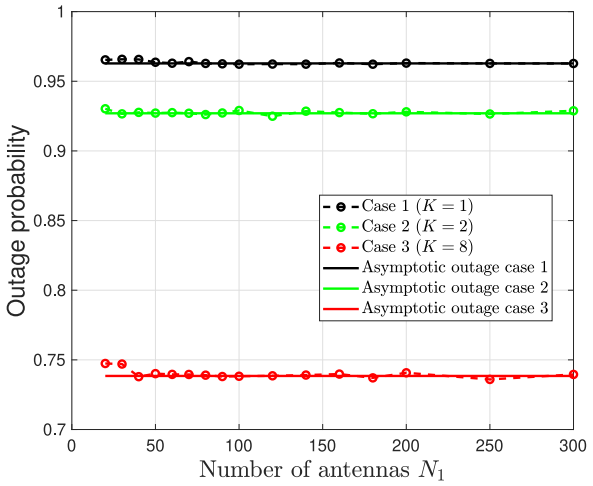

Fig. 8. Outage as a function the number of antennas at S_1 for end-node power scaling.

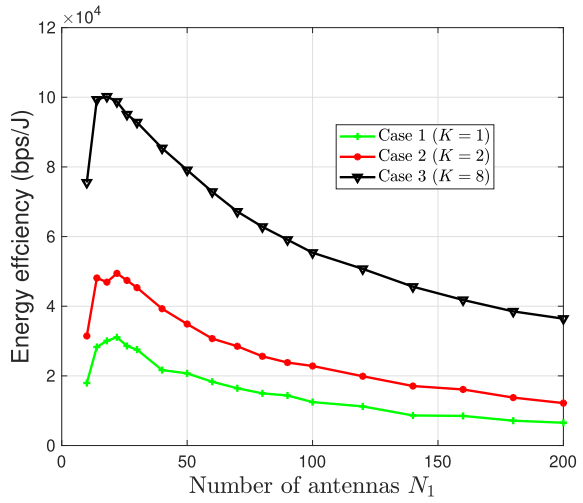

Fig. 10. Energy efficiency vs the number of antennas at S_1 for end-node power scaling.

where different number of relays are available for each case. Specifically, Case∼1 has a single relay ($L = 1$), whereas Case∼2 and Case∼3 involve 2 and 8 relays, respectively. In all the three cases, each relay has 8 antennas. We have used the relay selection to maximize the instantaneous sum rate, and the simulation results are compared with the analytical asymptotic results obtained in (48) and (49). These two figures show that the analytical results match with the simulation results. Also, more relays will reduce the outage and increase the sum rate. For instance, just one relay ($L = 1$) will experience an outage of 0.96 while the outage of a system with 8 relays is 0.74. With further calculations, we can show that the outage will drop to 0.47, if 20 relays are deployed. Furthermore, system with 8 relays can provide sum rate of 2.1bps/Hz while a system with a single relay can only provide a sum rate of 0.3 bps/Hz. This is significantly different than that for the previous case with all-node power scaling. In that case, the achievable asymptotic sum rate is not increased when the number of relays are increased while all of them had the same number of antennas. However, when power scaling is only used at the end nodes, more relays will significantly improve the sum rate.

The energy efficiency (60) is plotted in Fig. 10. Similar to the case in PS-1, the energy efficiency increases while the number of antennas are increased and then reduces after a certain threshold. For all the cases, this threshold is around 25 antennas. This is in contrast to the energy efficiency based only on the transmit power where the energy efficiency approaches infinity for extremely large number of antennas. Furthermore, more relays increases the energy efficiency and thus, relay selection helps to make green communications a reality.

In Fig. 11, we plot the sum rate against the number of antennas at the end nodes S_1 when power scaling is only used at the end nodes. In Case∼1, we use a relay with 8 antennas and in Case∼2 we use two relays with 8 antennas each. However in Case∼3, we use 8 relays where 2 relays contain 8 antennas and 6 relays with 4 antennas each. It can be seen that multiple relays increase the sum rate. When the number of antennas at different relays are not equal, no closed-form solution exists for the achievable asymptotic sum rate.

In Fig. 12, we plot the outage against the number of antennas at the end nodes S_1 when power scaling is only used at S_1 and S_2. We also plot the outage when the relay is selected

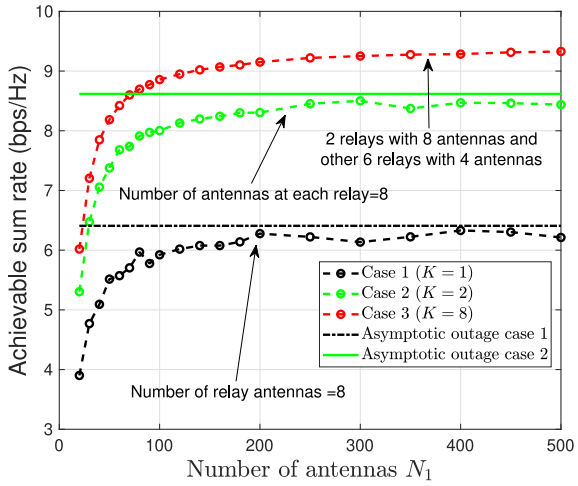

Fig. 11. Achievable sum rate vs the number of S_1 antennas for end-node power scaling.

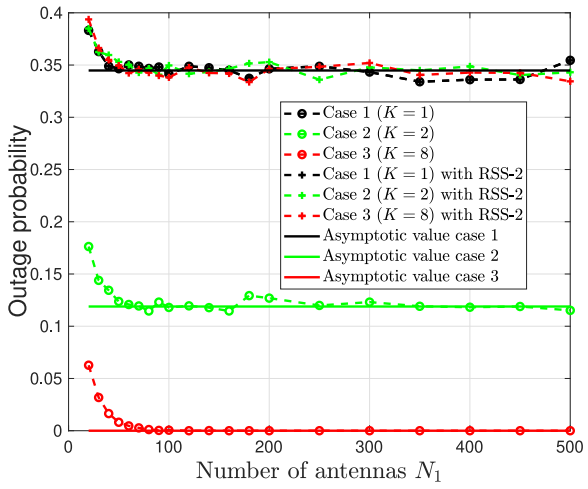

Fig. 12. Outage vs the number of S_1 antennnas for end-node power scaling.

based on the average sum rate, labeled as relay selection scheme 2 (RSS-2). Although the outage is significantly reduces when the number of relays are increased for relay selection to maximize the instantaneous sum rate scenario, RSS-2 can only obtain the performance equal to the case with a single relay antenna. As an example, with 8 relay nodes, the relay selection scheme can reduce the outage asymptotically to zero while the outage is 0.35 with just a single relay. Furthermore, two relays will reduce the outage to 0.12. If RSS-2 is used, the outage value does not change even if the number of relays is increased to 8. However, the overhead used for the relay selection for RSS-2 is significantly low as the same relay is selected every time regardless of the instantaneous channel conditions. The trade-off for this simplicity will be the increased outage and eventually the reduced achievable sum rate. Moreover, when all the relays have the same number of antennas, the advantage of having more relays will be absent if the relay selection is based on the average sum rate.

VIII. Conclusion

In this paper, we analyzed a cognitive mMIMO TWRN with relay selection to maximize the sum rate. We identified two power scaling scenarios: (1) (PS-1) the transmit powers of S_1, S_2, and the relay are scaled and (2) (PS-2) only those of S_1 and S_2 are scaled. Furthermore, optimal power allocation was derived for S_1 and S_2 subject to interference constraints at PU. Next, the power consumption was analyzed realistically for the proposed system. Our study reveals the following.

1) It is well known that mMIMO flattens the effect of small-scale fading in mMIMO, and that holds for cognitive mMIMO as well. In particular, the asymptotic SINR and sum rate do not depend on small-scale fading for PS-1. Thus, the PS-1 relay selection can be done offline. Further, the interference on the PU can easily be reduced or eliminated when the number of antennas at the end nodes becomes extremely large. Thus, the secondary system can function with zero outage.
2) In PS-2, the relay will have a finite number of antennas. Then, the outage of the TWRN will depend on the channel conditions between the relays and the PU. Thus, although the end-to-end SINR is independent of instantaneous channel state, the sum rate will still depend on the instantaneous values.
3) Relay selection improves the sum rate of the secondary system and reduces the interference on the PU. Our first relay selection method is based on the instantaneous sum rate maximization and thus, will require instantaneous CSI between relays and PU. Our second relay selection aims for the highest average sum rate and does not require instantaneous CSI. Thus, for this case, relay selection can be performed offline.
4) Also, as the number of available relays are increased, the energy efficiency is improved significantly for both PS-1 and PS-2. However, as the number of antennas at the end nodes are increased, energy efficiency peaks and then declines. The reduced transmit power due to power scaling improves energy efficiency while the increased power consumption in the transceiver chains reduces energy efficiency as the number of antennas at the end nodes are increased.

A potential future topic is to consider multiple PUs and establish the resulting outage conditions.

Appendix
Asymptotic SINR With Power Scaling at All Nodes

In this Appendix, we provide the sketch of the proof for SINR values (23) and (24). We use the following limit results for this proof [43]. For two independent complex Gaussian vectors $\mathbf{p} \sim \mathcal{CN}_{N\times 1}(0, \sigma_p^2)$ and $\mathbf{q} \sim \mathcal{CN}_{N\times 1}(0, \sigma_q^2)$, due to the law of large numbers, we readily find that

$$\mathbf{p}^{\mathrm{H}}\mathbf{p}/N \xrightarrow[N\to\infty]{a.s.} \sigma_P^2 \quad \text{and} \quad \mathbf{p}^{\mathrm{H}}\mathbf{q}/N \xrightarrow[N\to\infty]{a.s.} 0, \tag{61}$$

$$\mathbf{p}^{\mathrm{H}}\mathbf{q}/\sqrt{N} \xrightarrow[N\to\infty]{d} \mathcal{CN}\left(0, \sigma_p^2\sigma_q^2\right), \tag{62}$$

where subscripts *a.s.* and *d* stands for almost sure convergence and the convergence of distributions, respectively. Based

$$\gamma_{S_{1,k}^{(l)}|O_{S_k}=0} = \frac{\dfrac{E_{R_k}E_{2,k}}{N_2\,\mathrm{Tr}\left(\left[\mathbf{H}_{2,k}\mathbf{H}_{2,k}^{\mathrm{H}}\right]^{-1}\right)}}{E_{R_k}\sigma_{R_k}^2 + \left(\dfrac{E_{1,k}}{N_1\,\mathrm{Tr}\left(\left[\mathbf{H}_{1,k}\mathbf{H}_{1,k}^{\mathrm{H}}\right]^{-1}\right)} + \dfrac{E_{2,k}}{N_2\,\mathrm{Tr}\left(\left[\mathbf{H}_{2,k}\mathbf{H}_{2,k}^{\mathrm{H}}\right]^{-1}\right)} + \sigma_{R_k}^2\right)\dfrac{\sigma_1^2}{\eta_{1,k}}\left[N_1\left(\tilde{\mathbf{H}}_{k,1}^{\mathrm{H}}\tilde{\mathbf{H}}_{k,1}\right)^{-1}\right]_{l,l}} \tag{68}$$

on (61) and (62), we find the following limit results:

$$\frac{\mathbf{H}_{i,k}\mathbf{H}_{i,k}^{\mathrm{H}}}{N_i} = \mathbf{D}_{i,k}^{\frac{1}{2}}\left(\frac{\tilde{\mathbf{H}}_{i,k}\tilde{\mathbf{H}}_{i,k}^{\mathrm{H}}}{N_i}\right)\mathbf{D}_{i,k}^{\frac{1}{2}} \xrightarrow[N_i\to\infty]{a.s.} \mathbf{D}_{i,k}, \tag{63}$$

$$\frac{\mathbf{G}_k\mathbf{G}_k^{\mathrm{H}}}{N_i} = \mathbf{D}_k^{\frac{1}{2}}\left(\frac{\tilde{\mathbf{G}}_k\tilde{\mathbf{G}}_k^{\mathrm{H}}}{N_i}\right)\mathbf{D}_k^{\frac{1}{2}} \xrightarrow[N_i\to\infty]{a.s.} \mathbf{0}_{N\times N}, \tag{64}$$

$$\frac{\mathbf{F}_i^{\mathrm{H}}\mathbf{F}_i}{N_i} = \hat{\mathbf{D}}_i^{\frac{1}{2}}\left(\frac{\tilde{\mathbf{F}}_i^{\mathrm{H}}\tilde{\mathbf{F}}_i}{N_i}\right)\hat{\mathbf{D}}_i^{\frac{1}{2}} \xrightarrow[N_i\to\infty]{a.s.} \hat{\mathbf{D}}_i. \tag{65}$$

The asymptotic limit of (63) shows that the left hand side converges to the diagonal matrix. Since the trace of the inverse of a diagonal matrix can be readily determined, we obtain the following limit:

$$\lim_{N_i\to\infty} N_i\,\mathrm{Tr}\left(\left[\mathbf{H}_{i,k}\mathbf{H}_{i,k}^{\mathrm{H}}\right]^{-1}\right) = N_{R_k}\eta_{i,k}^{-1}. \tag{66}$$

Furthermore, (64) and (65) immediately lead to the following limits:

$$\lim_{N_i\to\infty}\frac{\mathrm{Tr}\left(\mathbf{G}_k^{\mathrm{H}}\mathbf{G}_k\right)}{N_i} = 0 \quad \text{and} \quad \lim_{N_i\to\infty}\frac{\mathrm{Tr}\left(\mathbf{F}_i^{\mathrm{H}}\mathbf{F}_i\right)}{N_i} = \hat{\eta}_i N, \tag{67}$$

Now the proofs of (23) and (24) can be outlined as follows. First, scaled transmit powers (25) are substituted on the SINR expression (18) to yield (68), shown at the top of this page. Here in (68), the limit of the matrix $(\tilde{\mathbf{H}}_{k,1}^{\mathrm{H}}\tilde{\mathbf{H}}_{k,1})/N_1$ as N_1 tends to infinity is the identity matrix of size $N_{R_k}\times N_{R_k}$, which follows from (61) and (62). By using this result and by replacing the trace terms in (68) by (66) we obtain (23). With similar steps, we obtain (24). Also, by using the second limit of (67) on (22), we obtain the interference constraint (27).

REFERENCES

[1] S. Silva, M. Ardakani, and C. Tellambura, "Relay selection for cognitive massive MIMO two-way relay networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[2] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021*. Cisco Syst., San Jose, CA, USA, 2017. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11–520862.pdf

[3] X. Ge, J. Yang, H. Gharavi, and Y. Sun, "Energy efficiency challenges of 5G small cell networks," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 184–191, May 2017.

[4] K. M. S. Huq, S. Mumtaz, J. Bachmatiuk, J. Rodriguez, X. Wang, and R. L. Aguiar, "Green HetNet CoMP: Energy efficiency analysis and optimization," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4670–4683, Oct. 2015.

[5] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[6] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[7] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.

[8] X. Gong, S. A. Vorobyov, and C. Tellambura, "Joint bandwidth and power allocation with admission control in wireless multi-user networks with and without relaying," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1801–1813, Apr. 2011.

[9] K. Loa *et al.*, "IMT-advanced relay standards," *IEEE Commun. Mag.*, vol. 48, no. 8, pp. 40–48, Aug. 2010.

[10] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[11] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[12] J. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[13] J. Chen, X. Chen, T. Liu, and L. Lei, "Toward green and secure communications over massive MIMO relay networks: Joint source and relay power allocation," *IEEE Access*, vol. 5, pp. 869–880, 2017.

[14] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 659–672, Mar. 2006.

[15] S. Atapattu, Y. Jing, H. Jiang, and C. Tellambura, "Relay selection schemes and performance analysis approximations for two-way networks," *IEEE Trans. Commun.*, vol. 61, no. 3, pp. 987–998, Mar. 2013.

[16] S. Atapattu, Y. Jing, H. Jiang, and C. Tellambura, "Relay selection and performance analysis in multiple-user networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 8, pp. 1517–1529, Aug. 2013.

[17] G. Amarasuriya, M. Ardakani, and C. Tellambura, "Output-threshold multiple-relay-selection scheme for cooperative wireless networks," *IEEE Trans. Veh. Commun.*, vol. 59, no. 6, pp. 3091–3097, Jul. 2010.

[18] S. Silva, G. Amarasuriya, C. Tellambura, and M. Ardakani, "Relay selection strategies for MIMO two-way relay networks with spatial multiplexing," *IEEE Trans. Commun.*, vol. 63, no. 12, pp. 4694–4710, Dec. 2015.

[19] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[20] G. Amarasuriya, C. Tellambura, and M. Ardakani, "Performance analysis of zero-forcing for two-way MIMO AF relay networks," *IEEE Wireless Commun. Lett.*, vol. 1, no. 2, pp. 53–56, Apr. 2012.

[21] S. Masrour, A. H. Bastami, and P. Halimi, "Spectrum sharing in cognitive radio networks using beamforming and two-path successive relaying," in *Proc. Iran. Conf. Elect. Eng. (ICEE)*, May 2017, pp. 1810–1814.

[22] G. Wang, Y. Zou, J. Lu, and C. Tellambura, "Cognitive transmission and performance analysis for amplify-and-forward two-way relay networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 5843–5848.

[23] M. Shamailzadeh and A. H. Bastami, "Cooperative beamforming in cognitive radio network with network-coded bidirectional filter-and-forward relaying," in *Proc. 24th Iran. Conf. Elect. Eng. (ICEE)*, May 2016, pp. 1454–1459.

[24] A. Advaita, M. M. Gali, T. M. C. Chu, and H. J. Zepernick, "Outage probability of MIMO cognitive cooperative radio networks with multiple AF relays using orthogonal space-time block codes," in *Proc. IEEE 13th Int. Conf. Wireless Mobile Comput. Netw. Commun. (WiMob)*, Oct. 2017, pp. 84–89.

[25] Y. Zou, Y.-D. Yao, and B. Zheng, "Cognitive transmissions with multiple relays in cognitive radio networks," *IEEE Trans. Commun.*, vol. 10, no. 2, pp. 648–659, Feb. 2011.

[26] P. Ubaidulla and S. Aissa, "Optimal relay selection and power allocation for cognitive two-way relaying networks," *IEEE Wireless Commun. Lett.*, vol. 1, no. 3, pp. 225–228, Jun. 2012.
[27] S. Kusaladharma and C. Tellambura, "Secondary user interference characterization for spatially random underlay networks with massive MIMO and power control," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 7897–7912, Sep. 2017.
[28] N. Wang, E. Hossain, and V. K. Bhargava, "Joint downlink cell association and bandwidth allocation for wireless backhauling in two-tier HetNets with large-scale antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3251–3268, May 2016.
[29] H. Al-Hraishawi and G. A. A. Baduge, "Sum rate analysis of cognitive massive MIMO systems with underlay spectrum sharing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–7.
[30] C. Yin, T. X. Doan, N. P. Nguyen, T. Mai, and L. D. Nguyen, "Outage probability of full-duplex cognitive relay networks with partial relay selection," in *Proc. Int. Conf. Recent Adv. Signal Process. Telecommun. Comput. (SigTelCom)*, Jan. 2017, pp. 115–118.
[31] D. Wang, R. Zhang, X. Cheng, L. Yang, and C. Chen, "Relay selection in full-duplex energy-harvesting two-way relay networks," *IEEE Trans. Green Commun. Netw.*, vol. 1, no. 2, pp. 182–191, Jun. 2017.
[32] A. A. Khan, M. H. Rehmani, and A. Rachedi, "Cognitive-radio-based Internet of Things: Applications, architectures, spectrum related functionalities, and future research directions," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 17–25, Jun. 2017.
[33] T. T. Duy and H. Y. Kong, "Exact outage probability of cognitive two-way relaying scheme with opportunistic relay selection under interference constraint," *IET Commun.*, vol. 6, no. 16, pp. 2750–2759, Nov. 2012.
[34] K. B. Fredj and S. Aissa, "Performance of spectrum-sharing constrained two-way relaying," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2014, pp. 845–850.
[35] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. Waltham, MA, USA: Academic, 2007.
[36] R. W. Heath, S. Sandhu, and A. Paulraj, "Antenna selection for spatial multiplexing systems with linear receivers," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 142–144, Apr. 2001.
[37] G. Wang, Q. Liu, R. He, F. Gao, and C. Tellambura, "Acquisition of channel state information in heterogeneous cloud radio access networks: Challenges and research directions," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 100–107, Jun. 2015.
[38] G. Wang, F. Gao, Y.-C. Wu, and C. Tellambura, "Joint CFO and channel estimation for OFDM-based two-way relay networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 456–465, Feb. 2010.
[39] G. Wang, F. Gao, W. Chen, and C. Tellambura, "Channel estimation and training design for two-way relay networks in time-selective fading environments," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2681–2691, Aug. 2011.
[40] S. Silva, G. A. A. Baduge, M. Ardakani, and C. Tellambura, "Performance analysis of massive MIMO two-way relay networks with pilot contamination, imperfect CSI, and antenna correlation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4831–4842, Jun. 2018.
[41] T. Pham-Gia, D. N. Thanh, and D. T. Phong, "Trace of the Wishart matrix and applications," *Open J. Stat.*, vol. 5, no. 3, p. 173, 2015.
[42] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, Jun. 2015.
[43] H. Cramér, *Random Variables and Probability Distributions*. Cambridge, U.K.: Cambridge Univ. Press, 1970.

Shashindra Silva (Student Member, IEEE) received the B.Sc. degree (First-Class Hons.) in engineering from the Department of Electronic and Telecommunication Engineering, University of Moratuwa, Moratuwa, Sri Lanka, in 2013, and the M.Sc. degree in electrical engineering from the Department of Electrical and Computer Engineering, University of Alberta, AB, Canada, in 2015, where he is currently pursuing the Ph.D. degree with the Electrical and Computer Engineering Department. His current research interests include massive MIMO, machine learning for wireless systems, and cooperative MIMO relay networks.

Masoud Ardakani (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the Isfahan University of Technology in 1994, the M.Sc. degree in electrical engineering from Tehran University in 1997, and the Ph.D. degree in electrical engineering from the University of Toronto, Canada, in 2004, where he was a Postdoctoral Fellow with the University of Toronto from 2004 to 2005. He is currently a Professor of electrical and computer engineering with the University of Alberta, Canada. His research interests are in the general area of information theory. He serves as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE WIRELESS COMMUNICATIONS and as a Senior Editor for the IEEE COMMUNICATION LETTERS.

Chintha Tellambura (Fellow, IEEE) received the B.Sc. degree (First-Class Hons.) from the University of Moratuwa, Sri Lanka, the M.Sc. degree in electronics from the King's College, University of London, U.K., and the Ph.D. degree in electrical engineering from the University of Victoria, Canada. He was with Monash University, Australia, from 1997 to 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta. He has authored or coauthored over 500 journal and conference papers with an H-index of 74 (Google Scholar). His current research interests include the design, modeling and analysis of cognitive radio, heterogeneous cellular networks, and 5G wireless networks. He served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS from 1999 to 2011 and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2001 to 2007 and for the latter he was the Area Editor for *Wireless Communications Systems and Theory* from 2007 to 2012. He has received Best Paper Awards in the Communication Theory Symposium in 2012 IEEE International Conference on Communications (ICC) in Canada and 2017 ICC in France. He is the winner of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. In 2011, he was elected as an IEEE Fellow for his contributions to physical layer wireless communication theory. In 2017, he was elected as a Fellow of the Canadian Academy of Engineering.