Optimal Resource Allocation and EE-SE Trade-Off in Hybrid Cognitive Gaussian Relay Channels

Xuemin Hong, *Member, IEEE*, Chao Zheng, Jing Wang, Jianghong Shi, and Cheng-Xiang Wang, *Senior Member, IEEE*

***Abstract*—Recent literature has suggested the benefits of integrating licensed radio and cognitive radio into a hybrid cooperative communication system. The fundamental properties of such hybrid systems, however, have not been thoroughly investigated. This paper studies the hybrid cognitive Gaussian relay channel (HCGRC), which uses licensed radio resource (RR) and cognitive/unlicensed RR for forward and relay transmissions, respectively. HCGRC fundamentally differs from conventional relay channels in that the licensed and cognitive RRs are not subject to a total resource constraint and that the cognitive RR is opportunistic in nature. With respect to both the upper and lower bounds, we derive the optimal power-bandwidth allocation strategies for the cognitive relay to maximize the capacity, spectrum efficiency (SE), and energy efficiency (EE). The Pareto-optimal EE-SE tradeoff curve is also derived analytically. Our study leads to two key observations. First, the multi-objective power-bandwidth allocation problem is characterized by five regions, each representing a unique performance tradeoff. Second, the reliability of cognitive RR has no impact on the EE-SE tradeoff given unlimited bandwidth and power.**

***Index Terms*—Cognitive radio, energy efficiency, spectral efficiency, trade-off.**

I. Introduction

DUE to the proliferation of wireless communication systems, available wireless spectrum resource is becoming increasingly scarce. Meanwhile, a large portion of the allocated spectrum is found to be highly underutilized [1]. The contradiction of "spectrum shortage" and "spectrum underutilization" motivated the concept of cognitive radio (CR) [2], [3], which enables opportunistic access of the underutilized spectrum to improve overall spectrum utilization. Incumbents and CR users using the same frequency band are differentiated by their priorities to access the spectrum, hence they are also called primary users (PUs) and secondary users (SUs), respectively.

The radio resource (RR) available to incumbents and CR systems are called licensed RR and cognitive (or, secondary) RR, respectively. The licensed RR is typically featured with a relatively small bandwidth, high transmit power, and high reliability. On the contrary, the cognitive RR is characterized by its potentially broad bandwidth, low transmit power, and low reliability. It is obvious that these two types of RRs are complementary in nature and demand different approaches for system design and optimization.

Most literature on CR networks assumed that only the cognitive RR is utilized, resulting in "pure" CR networks [4]–[18]. The performance of a pure CR network, however, is fundamentally unreliable due to the opportunistic nature of cognitive RRs. To overcome this drawback, a few articles have studied hybrid CR networks that jointly utilize both the licensed and cognitive RRs [19]–[25]. For example, a cellular operator can use its own cellular spectrum and leased CR spectrum from other bands to form a hybrid network. It was found that hybrid CR networks have the potential to outperform incumbent networks and pure CR networks by exploiting the complementary natures of licensed and cognitive RRs [19], [20].

Two basic architectures can be defined for hybrid CR networks: non-cooperative and cooperative [19], [20]. The former creates two separated radio interfaces operating at the licensed and cognitive RRs, respectively. In other words, licensed and cognitive RRs are used to build two networks that overlap in coverage, separated in physical layers, but integrated in upper layers to perform joint scheduling. In contrast, the cooperative architecture utilizes both licensed and cognitive RRs to design a single integrated physical layer using principles of cooperative communications, resulting in a hybrid cooperative CR network. It was found that the cooperative architecture has advantages over the non-cooperative one [19], [20]. The underlying rationale is that significant performance gains can be achieved when the heterogeneous RRs are carefully assigned to match the heterogeneous channels created by cooperative communication schemes. For example, the licensed RR is better used for long-range communications, while the cognitive RR is better used for short-range communications to facilitate local cooperation.

Manuscript received August 30, 2014; revised January 11, 2015; accepted March 11, 2015. Date of publication March 30, 2015; date of current version August 10, 2015. The authors gratefully acknowledge the support of this work from the National Natural Science Foundation of China (Grant No. 61201195 and 61201196), the Fundamental Research Funds for the Central Universities, Ministry of Science and Technology of China through the 863 Project in 5G Wireless Networking (Grant No. 2014AA012101), EU FP7 QUICK project (Grant No. PIRSES-GA-2013-612652), and the Opening Project of the Key Laboratory of Cognitive Radio and Information Processing (Guilin University of Electronic Technology), Ministry of Education (Grant No. 2013KF01). The associate editor coordinating the review of this paper and approving it for publication was M. Bennis.

X. Hong, C. Zheng, J. Wang and J. Shi are with the Department of Communications Engineering, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China (e-mail: xuemin.hong@xmu.edu.cn; c_zheng_xmu@163.com; j_wang_xmu@163.com; shijh@xmu.edu.cn).

C.-X. Wang is with the Institute of Sensors, Signals and Systems, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, U.K., and also with the School of Information Science and Engineering, Shandong University, Jinan 250100, China (e-mail: cheng-xiang.wang@hw.ac.uk).

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/TWC.2015.2417550

TABLE I
SUMMARY OF MAJOR THEORETICAL RESULTS

		Capacity	SE	EE	EE-SE trade-off
Lower bound	Performance Bound	(6)	(16)	(25)	(34)
	Optimal resource allocation	(12)	(22)	(31)	(12)
Upper bound	Performance Bound	(7)	(17)	(26)	(35)
	Optimal resource allocation	(13)	(24)	(33)	(13)

Our study in this paper focuses on hybrid cooperative CR systems. Previous works on hybrid cooperative CR networks have focused on system-level studies [19]–[24]. Such system-level studies are network-engineering oriented and focus on the performance of large scale networks where multiple users share the given resources. On the contrary, the link-level study of hybrid cooperative CR networks has so far received less attention. The link-level study is typically information-theoretic and concerns the fundamental performance bounds of a single transmit-receive pair. For "pure" CR networks, there exists a wealth of link-level studies. The capacity of Gaussian CR channels were studied in [26]–[30]. Extensions of the above results to fading channels and multiple-input multiple-output (MIMO) channels were presented in [31]–[36], respectively. In contrast, for hybrid cooperative CR networks, the link-level study is still by large an open field.

To our best knowledge, our recent conference article [25] made the first attempt to study the link-level properties of hybrid cooperative CR networks. As link-level study typically starts with the Gaussian channel, we formulated a new type of relay channel called hybrid cognitive Gaussian relay channel (HCGRC). HCGRC shares similar structures with the conventional Gaussian relay channel [37] and the orthogonal Gaussian relay channel [38], but also differs from them in two critical aspects: First, the source and relay are subject to separate resource constraints rather than a total resource constraint. Second, the cognitive RR is opportunistic in nature, hence it should be characterized by not only power and bandwidth, but also availability/reliability.

In [25], we made the first step to study performance lower bounds and single-objective resource allocation problems in HCGRC. This paper makes systematic extensions beyond [25] in three aspects. First, the performance upper bounds and corresponding resource allocation schemes are presented. Second, the fundamental energy efficiency (EE)-spectral efficiency (SE) tradeoff and the corresponding multi-objective resource allocation problem are studied for the first time. Third, some limiting values regarding performance metrics and resource allocation are derived for the first time. The main theoretical results are summarized in Table I to give readers an overview about the systematic structure of this paper. Such a systematic extension enables us to make critical comparisons and reveal much deeper insights into the basic properties of HCGRC.

The remainder of the paper is organized as follows. Section II motivates our study and describes the system model. Section III solves the single-objective resource allocation problems with respect to capacity, SE, and EE. Section IV further considers multi-objective resource allocation and discusses the EE-SE tradeoff. Numerical results and discussions are given in Section V, followed by conclusions in Section VI.

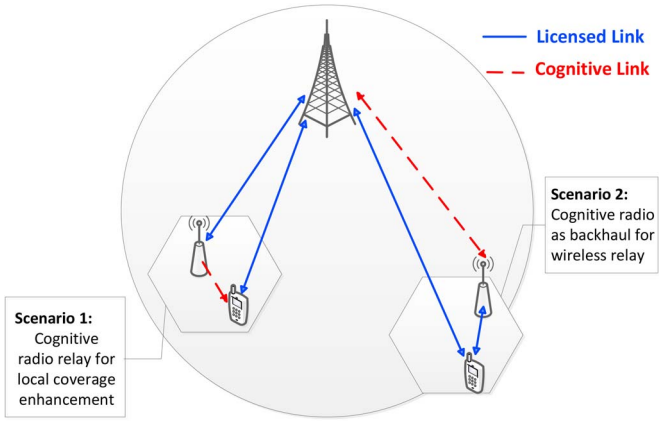

Fig. 1. Hybrid cooperative CR network in cellular systems.

II. MOTIVATION AND SYSTEM MODEL

The motivating application scenario of our study is illustrated in Fig. 1, where we envision a hybrid cooperative CR network for cellular communication systems. In Scenario 1, a cognitive relay is deployed for coverage extension or capacity enhancement. The cognitive relay communicates with the BS using the licensed RR (e.g., licensed cellular spectrum) and provides a local coverage using the cognitive RR (e.g., borrowed spectrum from TV white space). Such a cognitive relay is able to work in a duplex fashion to outperform conventional relays. In Scenario 2, a cognitive relay uses the cognitive RR for backhaul and the licensed RR for local coverage. A unique merit of this configuration is that no modification is needed for user devices, i.e., the network-side upgrade of deploying CR is transparent to users.

The performance of a wireless communication system can be evaluated by three metrics: capacity, spectral efficiency (SE), and energy-efficiency (EE). In recent years, the EE-SE tradeoff analysis has gained popularity [39]–[43]. The EE measures how efficiently the energy is consumed, while the SE indicates how efficiently the bandwidth is utilized. It is well-known that maximizing EE and SE are conflicting objectives and there exists a fundamental tradeoff between them [39]–[43]. The EE-SE tradeoff, therefore, reveals the theoretical performance boundary of a communication system. All these three metrics have unique practical values for CR-enabled cellular systems. Capacity is the most important metric because the main purpose of using CR is for capacity enhancement. On the other hand, the EE-SE tradeoff study also has a unique practical value. It is envisioned

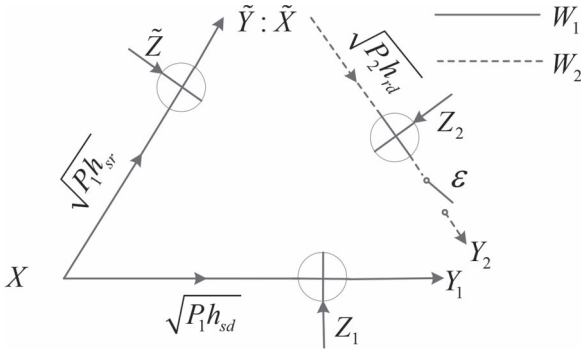

Fig. 2. Heterogeneous cognitive Gaussian relay channel (HCGRC) model.

that spectrum leasing bills (related to bandwidth and SE) and electricity bills (related to EE) will be two significant parts in the operational expenditure (OPEX) of a real hybrid cooperative CR system. Therefore, understanding the EE-SE tradeoff can provide direct guidelines to the OPEX management of commercial hybrid CR networks. Our paper will study all these three metrics and investigate the pair-wise tradeoffs among them.

As a general abstraction of the above scenarios, we consider a simple three-node HCGRC with a source, a destination, and a relay. The source broadcasts to the relay and destination using the licensed RR, while the relay forwards information to the destination using the cognitive RR. This abstraction characterizes the downlink of Scenario 1 and the uplink of Scenario 2 in Fig. 1. Although our paper does not specifically address the uplink of Scenario 1 and the downlink of Scenario 2, extensions can be easily made.

Because the licensed and cognitive RRs use different frequency bands, we assume that the relay node can work in a full-duplex fashion, i.e., receive and transmit at the same time. It should be noted that the characteristics of cognitive RR depend on the primary system/incumbent as well as the coexisting mechanism, which includes transmitter-oriented (sensing and database) approaches and receiver-oriented (interference temperature) approaches [3]. There exists a wealth of literature that investigates the available cognitive RR with respect to various primary systems under different coexisting mechanisms. Without going into such specifics, we propose a higher level abstraction of the cognitive RR that characterizes it with power, bandwidth, and reliability. This abstraction is useful in decoupling the resource-acquisition and resource-allocation problems in CR systems and allows us to focus on the latter problem.

The model of HCGRC is illustrated in Fig. 2. The solid and dashed lines indicate transmissions in the licensed and cognitive bands, respectively. We are interested in the capacity, SE, and EE of the HCGRC. Denote the transmit power in the licensed and cognitive RRs as P_1 and P_2, respectively. Similarly, denote the bandwidth of the licensed and cognitive RRs as W_1 and W_2, respectively. To clearly indicate the relationship between licensed and cognitive RRs, we define bandwidth ratio θ and power ratio φ as

$$\theta = W_2/W_1, \quad \varphi = P_2/P_1 \tag{1}$$

respectively. The HCGRC signaling is characterized by the following equations

$$\begin{aligned} Y_1 &= \sqrt{P_1 h_{sd}} X + Z_1 \\ Y_2 &= \varepsilon \left(\sqrt{P_2 h_{rd}} \tilde{X} + Z_2 \right) \\ \tilde{Y} &= \sqrt{P_1 h_{sr}} X + \tilde{Z} \end{aligned} \tag{2}$$

where X and $\widetilde{X}$ are inputs of the licensed and cognitive channels, respectively, h_{sr}, h_{rd} and h_{sd} are channel gains from source-to-relay, relay-to-destination and source-to-destination, respectively. Moreover, Z_1, Z_2, and $\widetilde{Z}$ are zero-mean independent white Gaussian noises, whose variances are given by $W_1 N_0$, $W_2 N_0$, and $W_1 \tilde{N}_0$, respectively. Here, N_0 and $\tilde{N}_0$ are the noise power spectrum densities at the destination and relay, respectively. They are treated as different to reflect the potential difference in receiver noise figures at the destination and relay.

The transmit signal-to-noise ratios (SNRs) of the source-to-destination link and source-to-relay link can be written as

$$\begin{aligned} \rho_1 &= \frac{P_1}{N_0 W_1} \\ \rho_3 &= \frac{P_1}{\tilde{N}_0 W_1} \end{aligned} \tag{3}$$

respectively, and the transmit SNR of the relay-to-destination link can be expressed as $\rho_2 = \rho_1 \varphi / \theta$. In (2), ε is a binary random variable defined on probability space [0,1] to represent the opportunistic nature of the cognitive channel. When the cognitive band is unavailable, we have $\varepsilon = 0$. In this case both the CR transmitter and receiver in the cognitive band stop working and do not consume any extra power. When the cognitive band is available, we have $\varepsilon = 1$ to give a conventional Gaussian channel. The mean of ε is $\bar{\varepsilon}$, which can be interpreted as the reliability measure of the cognitive channel or the fraction of time that the cognitive channel is available. It is assumed that ε varies at a much slower rate than the transmit symbols. Furthermore, the channel capacity is evaluated over a long time period so that ε becomes statistically relevant.

The signaling procedure of the HCGRC takes four steps. 1) When the source initiates a connection, bandwidth W_1 and power P_1 are allocated to the source from the licensed band. This allocation is done in the legacy cellular network and independent from the CR relay; 2) As the source communicates to the destination in the licensed band, a CR relay also receives the user's transmitted signal and stores the information; 3) Meanwhile, the CR relay senses the cognitive band for secondary access. When this band is available (i.e., $\varepsilon = 1$), the CR relay decides a bandwidth W_2 and power P_2 to relay information to the destination. Otherwise the CR relay transmits nothing. 4) The destination receives both the continuous signal from the licensed band and the intermittent signal from the cognitive band to perform joint decoding over a sufficiently long codebook. The codebook is long-enough to permit a statistical characterization of ε using its mean value $\bar{\varepsilon}$. Based on the above procedure, this paper aims to address a unique research problem: given W_1, P_1, and $\bar{\varepsilon}$, how should the cognitive relay adjust W_2 and P_2 to optimize the overall performance? This problem differs from the classic problems of CR resource

allocation because we consider the overall performance over both licensed and cognitive RRs, resulting in a fundamentally different objective function for optimization.

III. Performance Analysis and Resource Allocation

This section focuses on the performance analysis of HCGRC under three different metrics: capacity, SE, and EE. For each metric, we will first obtain its theoretical upper and lower bounds, followed by the derivation of the optimal bandwidth and power allocation in the cognitive band.

A. Capacity

The HCGRC model is similar to the Gaussian orthogonal relay model studied in [38]. Both types of channels have the same capacity bounds given by [38]

$$C_{lower} = \sup_{p(x)p(\tilde{x})} \min\left\{I(X;Y_1)+I(\tilde{X};Y_2), I(X;\tilde{Y})\right\} \quad (4)$$

$$C_{upper} = \sup_{p(x)p(\tilde{x})} \min\left\{I(X;Y_1)+I(\tilde{X};Y_2), I(X;\tilde{Y},Y_1)\right\} \quad (5)$$

where C_{lower} and C_{upper} are the lower and upper bounds of the capacity, respectively, $p(\cdot)$ indicates the probability distribution function, and $I(\cdot)$ returns the mutual information. However, unlike [38], the source node and relay node in HCGRC are not subject to a total bandwidth constraint. This leads to different capacity formulas as follows (Appendix A).

Proposition 1: The lower and upper capacity bounds of HCGRC are given by

$$C_{lower}(\theta,\varphi) = \min\left\{C_{1,low}(\theta,\varphi), C_{2,low}\right\} \quad (6)$$

$$C_{upper}(\theta,\varphi) = \min\left\{C_{1,up}(\theta,\varphi), C_{2,up}\right\} \quad (7)$$

respectively, where

$$C_{1,low} = W_1\log(1+\rho_1 h_{sd}) + W_1\theta\bar{\varepsilon}\log\left(1+\frac{\rho_1\varphi h_{rd}}{\theta}\right) \quad (8)$$

$$C_{2,low} = W_1\log(1+\rho_3 h_{sr}) \quad (9)$$

$$C_{1,up} = C_{1,low} \quad (10)$$

$$C_{2,up} = W_1\log(1+\rho_3 h_{sr}+\rho_1 h_{sd}). \quad (11)$$

For convenience to our subsequent discussion, the capacity bounds are expressed as functions of bandwidth ratio θ and power ratio φ in (6) and (7). As shown in (6) and (7), when θ and φ increase, the capacity bounds will eventually saturate to $C_{2,low}$ and $C_{2,up}$. Readers can also refer to Fig. 3(a) for an intuitive understanding. In practice, we are interested in the optimal allocation of θ and φ given other parameters. Let (θ_c^*,φ_c^*) and (θ_c^0,φ_c^0) denote the optimal solutions that maximize the lower bound and upper bound of the capacity, respectively. A joint power and bandwidth allocation (θ,φ) is said to be optimal when the capacity bound is maximized while the least bandwidth (or power) is used given the power (or bandwidth). Mathematically, the optimal solution

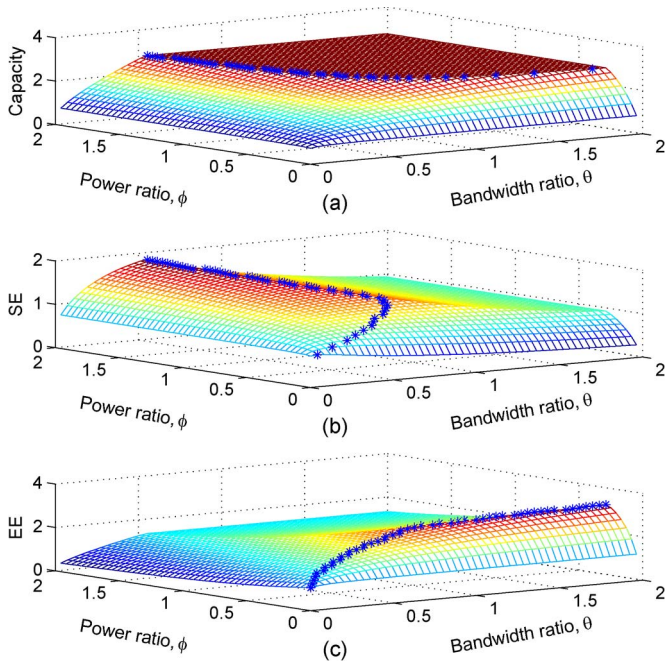

Fig. 3. Capacity, SE and EE lower bound as functions of θ and φ ($\bar{\varepsilon}=1$, $r_{sr}=1/2$).

with respect to the lower bound is defined as the (θ_c^*,φ_c^*) pair that satisfies $\theta_c^* = \min\arg\sup_{\theta}(C_{lower}(\theta,\varphi_c^*))$ and $\varphi_c^* = \min\arg\sup_{\varphi}(C_{lower}(\theta_c^*,\varphi))$. Similar definition applies to the upper bound. The following proposition can be obtained.

Proposition 2: The optimal resource allocation solution (θ_c^*,φ_c^*) that achieves the capacity lower bound is

$$\varphi_c^* = \Psi(\theta_c^*) = \frac{\theta_c^*}{\rho_1 h_{rd}}\left[\left(\frac{1+\rho_3 h_{sr}}{1+\rho_1 h_{sd}}\right)^{\frac{1}{\theta_c^*\bar{\varepsilon}}} - 1\right] \quad (12)$$

and the optimal resource allocation solution (θ_c^0,φ_c^0) that achieves the capacity upper bound is

$$\varphi_c^0 = \Phi(\theta_c^0) = \frac{\theta_c^0}{\rho_1 h_{rd}}\left[\left(\frac{1+\rho_3 h_{sr}+\rho_1 h_{sd}}{1+\rho_1 h_{sd}}\right)^{\frac{1}{\theta_c^0\bar{\varepsilon}}} - 1\right]. \quad (13)$$

A sketch of the proof of Proposition 2 is given in Appendix B. Proposition 2 shows that to obtain the maximum capacity, power and bandwidth are exchangeable to certain extent. When power ratio φ tends to infinity, the required bandwidth ratio θ approaches zero. It should be noted that the reverse, however, does not hold. It can be easily shown that when θ tends to infinity, a minimum value of φ should be satisfied to achieve the maximum capacity. The minimum values are given by

$$\varphi_{low}^{\min} = \ln\left(\frac{1+\rho_3 h_{sr}}{1+\rho_1 h_{sd}}\right)\Big/\bar{\varepsilon}\rho_1 h_{rd} \quad (14)$$

$$\varphi_{up}^{\min} = \ln\left(\frac{1+\rho_1 h_{sd}+\rho_3 h_{sr}}{1+\rho_1 h_{sd}}\right)\Big/\bar{\varepsilon}\rho_1 h_{rd} \quad (15)$$

for the lower and upper bounds, respectively.

B. Spectral Efficiency

In this subsection, we address another performance metric of the HCGRC: SE, which measures the average number of bits per Hertz. The classic definition of SE is the ratio of capacity and bandwidth. This definition should be modified in the context of HCGRC to accommodate two new features: First, the exact capacity cannot be obtained. Therefore, capacity will be evaluated by its upper and lower bounds. Second, the cognitive bandwidth W_2 is only available for a fraction of time $\bar{\varepsilon}$. Intuitively, when the cognitive bandwidth is not available, it should not be taken into account as part of the utilized system resources. To this end, we define the "effective bandwidth" of the total system as $W_1 + \bar{\varepsilon}W_2$, which is the average utilized bandwidth per unit time over a long period. It follows that we can define the lower and upper bounds of the SE as $S_{lower} = C_{lower}/(W_1 + \bar{\varepsilon}W_2)$ and $S_{upper} = C_{upper}/(W_1 + \bar{\varepsilon}W_2)$, respectively. Such a modified definition of SE has a straightforward physical meaning: consider a time interval T, the SE measures the maximum number of bits transmitted in T against the total spectral-temporal degrees of freedom, which is the product of bandwidth and the actual duration of spectrum utilization. From (6)–(11) we get

$$S_{lower}(\theta,\varphi) = \min\{S_{1,low}(\theta,\varphi), S_{2,low}(\theta)\} \tag{16}$$
$$S_{upper}(\theta,\varphi) = \min\{S_{1,up}(\theta,\varphi), S_{2,up}(\theta)\} \tag{17}$$

respectively, where

$$S_{1,low} = \frac{\log(1+\rho_1 h_{sd}) + \theta\bar{\varepsilon}\log(1+\varphi\rho_1 h_{rd}/\theta)}{(\theta\bar{\varepsilon}+1)} \tag{18}$$
$$S_{2,low} = (\log(1+\rho_3 h_{sr}))/(\theta\bar{\varepsilon}+1) \tag{19}$$
$$S_{1,up} = S_{1,low} \tag{20}$$
$$S_{2,up} = (\log(1+\rho_3 h_{sr}+\rho_1 h_{sd}))/(\theta\bar{\varepsilon}+1). \tag{21}$$

It is easy to see that given θ, SE is a non-decreasing function on φ. On the other hand, given φ, there exists an optimal θ that maximizes SE. Let (θ_s^*, φ) and (θ_s^0, φ) denote the optimal solutions with respect to the lower bound and upper bound of the SE, respectively. A joint power and bandwidth allocation (θ, φ) is said to be optimal when the maximum SE is achieved for the given φ. Mathematically, the optimal solution satisfies $\theta_s^*(\varphi) = \arg\max_{\theta}(S_{lower}(\theta,\varphi))$ for the lower bound. Similar definition applies to the upper bound.

Proposition 3: The SE lower bound is a monotonically increasing function of φ. The optimal resource allocation pair (θ_s^*, φ) with respect to the SE lower bound is a piecewise function given by

$$\theta_s^*(\varphi) = \begin{cases} \theta_{s(1)}^*(\varphi) = \frac{\varphi\rho_1 h_{rd} W_0(k_1)}{\varphi\rho_1 h_{rd}\bar{\varepsilon} - 1 - W_0(k_1)} & \varphi < \varphi_{th}^* \\ \Psi^{-1}(\varphi) & \varphi > \varphi_{th}^* \end{cases} \tag{22}$$

where

$$k_1 = \frac{(\varphi\rho_1 h_{rd}\bar{\varepsilon} - 1)}{e(1+\rho_1 h_{sd})} \tag{23}$$

and $W_0(x)$ is the Lambert W function [46] which satisfies $W(x)e^{W(x)} = x$ for any complex number x, e is the base of the natural logarithm, and $\Psi^{-1}(\cdot)$ is the inverse function of $\Psi(\cdot)$ defined in (12). The threshold φ_{th}^* is the unique solution for $\theta_{s(1)}^*(\varphi) = \Psi^{-1}(\varphi)$. Although φ_{th}^* cannot be solved analytically, it can be easily calculated via numerical methods.

Similarly, for the SE upper bound, the optimal resource allocation (θ_s^0, φ) is given by

$$\theta_s^0(\varphi) = \begin{cases} \theta_{s(1)}^*(\varphi) & \varphi < \varphi_{th}^0 \\ \Phi^{-1}(\varphi) & \varphi > \varphi_{th}^0 \end{cases} \tag{24}$$

where $\theta_{s(1)}^*(\varphi)$ is defined in (22), $\Phi(\cdot)$ is defined in (13), and $\Phi^{-1}(\cdot)$ denotes its inverse function. The threshold φ_{th}^0 is the unique solution for $\theta_{s(1)}^*(\varphi) = \Phi^{-1}(\varphi)$. The proof of Proposition 3 is given in Appendix C.

C. Energy Efficiency

The EE metric evaluates the average number of bits per Joule spent. In this paper, we consider the total energy consumption of the source and relay. When the cognitive spectrum is unavailable (i.e., $\varepsilon = 0$), the relay consumes no power. It follows that we can define the lower and upper bounds of the EE as $E_{lower} = C_{lower}/(P_1 + \bar{\varepsilon}P_2)$ and $E_{upper} = C_{upper}/(P_1 + \bar{\varepsilon}P_2)$, respectively. According to (6)–(11), the corresponding lower and upper bounds of EE are given by

$$E_{lower}(\theta,\varphi) = \min\{E_{1,low}(\theta,\varphi), E_{2,low}(\varphi)\} \tag{25}$$
$$E_{upper}(\theta,\varphi) = \min\{E_{1,up}(\theta,\varphi), E_{2,up}(\varphi)\} \tag{26}$$

respectively, where

$$E_{1,low} = \frac{W_1(\log(1+\rho_1 h_{sd}) + \theta\bar{\varepsilon}\log(1+\varphi\rho_1 h_{rd}/\theta))}{(P_1 + P_1\varphi\bar{\varepsilon})} \tag{27}$$
$$E_{2,low} = (W_1\log(1+\rho_3 h_{sr}))/(P_1 + P_1\varphi\bar{\varepsilon}) \tag{28}$$
$$E_{1,up} = E_{1,low} \tag{29}$$
$$E_{2,up} = (W_1\log(1+\rho_3 h_{sr}+\rho_1 h_{sd}))/(P_1 + P_1\varphi\bar{\varepsilon}). \tag{30}$$

It is easy to see that EE is a non-decreasing function on θ. Therefore, a joint power and bandwidth allocation (θ, φ) is said to be EE-optimal when the maximum EE is achieved for the given θ. In other words, the optimal solution satisfies $\varphi_e^*(\theta) = \arg\max_{\varphi}(E_{lower}(\theta,\varphi))$ for the lower bound. Similar definition applies to the upper bound.

Proposition 4: The SE lower bound is a monotonically increasing function of θ. The optimal power allocation pair (θ, φ_e^*) that achieves the EE lower bound is a piecewise function given by

$$\varphi_e^*(\theta) = \begin{cases} \varphi_{e(1)}^*(\theta) = \frac{\rho_1 h_{rd} - \theta_{\bar{\varepsilon}} - \theta\bar{\varepsilon}W_0(k_2)}{\rho_1 h_{rd}\bar{\varepsilon}W_0(k_2)} & \theta < \theta_{th}^* \\ \Psi(\theta) & \theta > \theta_{th}^* \end{cases} \tag{31}$$

where

$$k_2 = \frac{(\rho_1 h_{rd} - \theta\bar{\varepsilon})\exp\left(\frac{-\theta\bar{\varepsilon}+\ln(1+\rho_1 h_{sd})}{\theta\bar{\varepsilon}}\right)}{\theta\bar{\varepsilon}} \tag{32}$$

and $\Psi(\cdot)$ is defined in (12). The threshold θ_{th}^{*} is the unique solution for $\varphi_{e(1)}^{*}(\theta)=\Psi(\theta)$. Similarly, with regard to the EE upper bound, the optimal allocation is given by

$$\varphi_e^0(\theta)=\begin{cases}\varphi_{e(1)}^{*}(\theta) & \theta<\theta_{th}^0\\ \Phi(\theta) & \theta>\theta_{th}^0\end{cases} \tag{33}$$

where $\varphi_{e(1)}^{*}(\theta)$ is defined in (31) and $\Phi(\cdot)$ is defined in (13). The threshold θ_{th}^0 is the unique solution for $\varphi_{e(1)}^{*}(\theta)=\Phi(\theta)$. The proof of Proposition 4 is given in Appendix D.

IV. EE-SE Trade-Off and Resource Allocation

In the previous section, we have focused on bandwidth and power allocation with respect to a single performance metric (capacity, SE, and EE). Ideally, it is desirable to maximize all the three metrics simultaneously. However, optimizing EE and SE are conflicting objectives in most cases and there is a fundamental tradeoff between them [39]–[43]. The EE-SE tradeoff can be studied as a multi-objective optimization problem $\max_{\theta,\phi}(SE,EE)$, which aims to maximize SE and EE across all possible values of θ and φ. We are interested in the Pareto-optimal solution of this problem. In multi-objective optimization, a solution is called Pareto optimal if none of the objective functions can be improved without degrading other objective values. In the case of bi-objective problems, the Pareto front is also called tradeoff curve. The EE-SE tradeoff curve can be expressed by writing EE as a function of SE. It is important to find the Pareto optimal EE-SE tradeoff curve because it reveals the theoretical performance boundary of the HCGRC and serves as a benchmark for practical system designs. Without additional subjective preference for EE or SE, all Pareto optimal solutions can be considered equally good. In practice, once certain preference is decided (e.g., in terms of weights), the optimal EE-SE operating point can be easily obtained from the tradeoff curve.

Proposition 5: The Pareto-optimal EE-SE tradeoff curves with respect to capacity lower and upper bounds are given by

$$E_{low}(S)=\frac{K_1}{K_2-\frac{K_3}{S}+K_3\left(\frac{1}{S}-\frac{1}{K_5}\right)K_4^{\frac{S}{K_5-S}}} \tag{34}$$

$$E_{up}(S)=\frac{T_1}{T_2-\frac{T_3}{S}+T_3\left(\frac{1}{S}-\frac{1}{T_5}\right)T_4^{\frac{S}{T_5-S}}} \tag{35}$$

respectively, where $K_1=W_1\log(1+\rho_3 h_{sr})$, $K_2=P_1\left(1+\frac{1}{\rho_1 h_{rd}}\right)$, $K_3=\frac{P_1\log(1+\rho_3 h_{sr})}{\rho_1 h_{rd}}$, $K_4=\frac{1+\rho_3 h_{sr}}{1+\rho_1 h_{sd}}$, $K_5=\log(1+\rho_3 h_{sr})$, $T_1=W_1\log(1+\rho_3 h_{sr}+\rho_1 h_{sd})$, $T_2=P_1\left(1+\frac{1}{\rho_1 h_{rd}}\right)$, $T_3=\frac{P_1\log(1+\rho_3 h_{sr}+\rho_1 h_{sd})}{\rho_1 h_{rd}}$, $T_4=\frac{1+\rho_3 h_{sr}+\rho_1 h_{sd}}{1+\rho_1 h_{sd}}$, and $T_5=\log(1+\rho_3 h_{sr}+\rho_1 h_{sd})$. The proof of the above proposition is given in Appendix E.

The above proposition gives the maximum EE value for any feasible value of SE. Such a tradeoff is a fundamental one in the sense that it is fully determined by fixed parameters that characterize the system scenario (in contrast to θ and φ that can be adjusted dynamically). From a theoretical perspective, we are also interested in the global maximum values of SE and EE. Let $(S_{low}^{\max},E_{low}^{\max})$ and $(S_{up}^{\max},E_{up}^{\max})$ denote the global maximum values of SE and EE with respect to capacity lower and upper bounds, respectively. The following corollary can be obtained.

Corollary 1: The global maximum values of SE and EE with respect to the capacity lower bound are

$$\begin{cases}S_{low}^{\max}=\log(1+\rho_3 h_{sr})\\ E_{low}^{\max}=\frac{W_1\log(1+\rho_3 h_{sr})}{P_1\left[1+\frac{1}{\rho_1 h_{rd}}\ln\left(\frac{1+\rho_3 h_{sr}}{1+\rho_1 h_{sd}}\right)\right]}\end{cases} \tag{36}$$

respectively. Similarly, the global maximum values of SE and EE with respect to the capacity upper bound are

$$\begin{cases}S_{up}^{\max}=\log(1+\rho_3 h_{sr}+\rho_1 h_{sd})\\ E_{up}^{\max}=\frac{W_1\log(1+\rho_1 h_{sd}+\rho_3 h_{sr})}{P_1\left[1+\frac{1}{\rho_1 h_{rd}}\ln\left(\frac{1+\rho_1 h_{sd}+\rho_3 h_{sr}}{1+\rho_1 h_{sd}}\right)\right]}\end{cases} \tag{37}$$

respectively. The proof of the above corollary is given in Appendix F.

Finally, we are interested in the bandwidth and power allocation corresponding to the Pareto front, i.e., the optimal resource allocation resulting in the Pareto optimal EE-SE tradeoff curve. This leads to the following corollary.

Corollary 2: The optimal resource allocations leading to the Pareto-optimal EE-SE solutions are given in (12) and (13) for upper and lower bounds, respectively.

Proof: As shown in Appendix E, the Pareto optimality is achieved when $S_1=S_2$ and $E_1=E_2$ are satisfied. This can be easily shown to be equivalent to $C_1=C_2$. In other words, the Pareto optimal power-bandwidth allocation is the same as the allocation in Proposition 2, which maximizes the capacity with minimum bandwidth and power.

V. Numerical Results and Discussions

This section presents numerical results based on our previous analysis. For purpose of illustration, we assume that $h_{sr}=r_{sr}^{\alpha}$, $h_{sd}=r_{sd}^{\alpha}$, $h_{rd}=r_{rd}^{\alpha}$, where α is the path-loss exponent, r_{sr}, r_{sd}, and r_{rd} are distances between the source-to-relay, source-to-destination, and relay-to-destination, respectively. Without loss of generality, we set $W_1=1$, $P_1=1$, $r_{sd}=1$, and $\alpha=4$. We also assume that the relay lies on the line between the source and destination for simplicity.

First of all, to give readers an intuitive understanding on the nature of our problem, Fig. 3 shows the capacity, SE, and EE as functions of bandwidth ratio θ and power ratio φ based on (6), (16), and (25), respectively. The lower bounds are used and we take $\bar{\varepsilon}=1$ for example. Fig. 3(a) shows that the capacity increases monotonically with θ and φ until it reaches a maximum value. Fig. 3(b) shows that SE increases monotonically with φ, while Fig. 3(c) shows that EE increases monotonically with θ. The corresponding optimal (θ-φ) curves are obtained by numerical methods and highlighted in Fig. 3. These curves are what Propositions 2 to 4 aim to derive.

Based on Propositions 2 to 4, Fig. 4 shows the optimal power-bandwidth allocation curves with respect to the upper and lower bounds of capacity, SE, and EE, respectively. Based on (14)

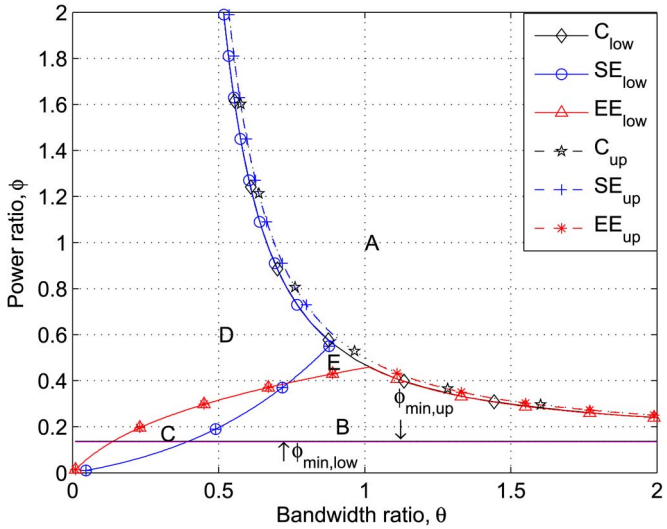

Fig. 4. Upper and lower bounds of optimal bandwidth-power allocation curves ($\bar{\varepsilon} = 1, r_{sr} = 1/2$).

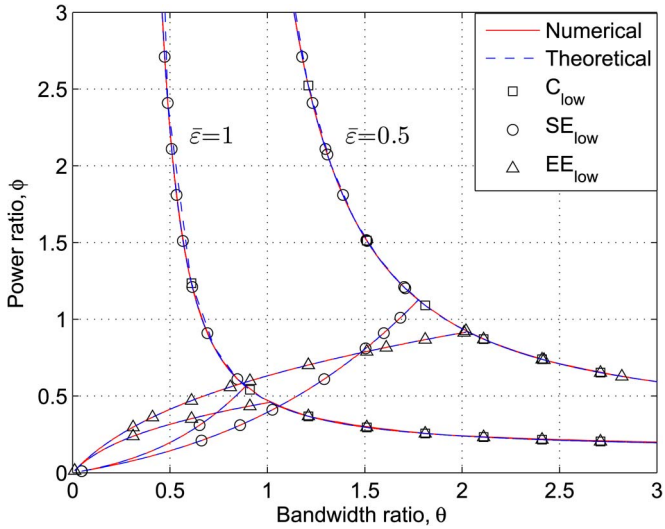

Fig. 5. Lower bound of optimal bandwidth-power allocation curves with varying $\bar{\varepsilon}$ ($r_{sr} = 1/2$).

and (15), the minimum values of $\varphi_{low}^{\min} = 0.1930$ and $\varphi_{up}^{\min} = 0.1981$ required to achieve maximum capacity bounds are also shown. Two key observations are made. First, the upper and lower bounds do not lead to significant differences in optimal resource allocation curves. In fact, a close look reveals that such differences will remain small as long as the SNR of the source-to-relay channel dominants the SNR of the source-to-destination channel. Second, the curves for SE and EE partly overlap with the curve for capacity. These curves divide the power-bandwidth plane into five areas (A to E). Each area has a unique implication as follows:

A) Resource excess area: Power and bandwidth are over-provisioned and have caused negative impacts on SE and EE.
B) Power hungry area: increasing power will improve all three metrics, while increasing bandwidth will improve capacity and EE but degrade SE.
C) Power and bandwidth hungry area: increasing either power or bandwidth will improve all three metrics.
D) Bandwidth hungry area: increasing bandwidth will improve all three metrics, while increasing power will improve capacity and SE but degrade EE.
E) Trade-off area: increasing power will improve capacity and SE but degrade EE, while increasing bandwidth will improve capacity and EE but degrade SE.

These five regions shown in Fig. 4 can provide useful guidelines for the cognitive relay to adjust its bandwidth and power according to performance requirements.

Similar to Fig. 4, Fig. 5 shows the optimal power-bandwidth allocation curves with varying reliability parameter $\bar{\varepsilon}$. Numerical results obtained by grid searching methods (see Fig. 3) are shown to collaborate the theoretical results calculated based on Propositions 2 to 4. As cognitive RR is featured by its unreliability, we are particularly interested in the impact of $\bar{\varepsilon}$ on power-bandwidth allocation. Taking the intersections of curves as key reference points, we observe that the optimal power-bandwidth allocation scales roughly linearly with $1/\varepsilon$. In other words, for a k-percent reduction of reliability, both the power and bandwidth should increase by $1/(1-k)$ to compensate for the reliability loss.

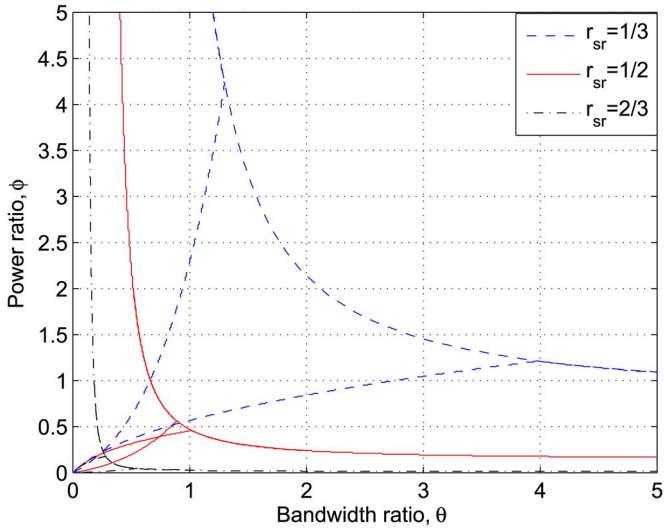

Fig. 6. Lower bound of optimal bandwidth-power allocation curves with varying relay location ($\bar{\varepsilon} = 1$).

Fig. 6 illustrates the impact of relay location on optimal power-bandwidth allocation. The normalized source-relay distances are set to vary from 1/3 to 2/3. It is observed that the required power and bandwidth decrease quickly when relay moves from source to destination. This is an intuitive result as the relay transmission not only enjoys a better channel, but also has less information from the source. It should be noted that the maximum capacity, SE, and EE change with varying relay location. When both the licensed and cognitive RRs are fixed, the optimal relay location can be easily calculated [47].

Fig. 7 illustrates the EE-SE tradeoff curve. The tradeoff lower bounds are obtained by three different approaches and shown to be consistent. The first approach, denoted as 'numerical',

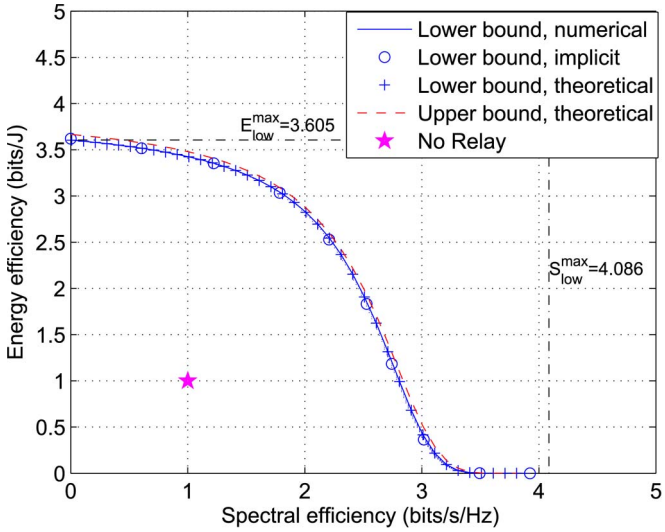

Fig. 7. Upper and lower bounds of the EE- SE trade-off ($r_{sr} = 1/2$).

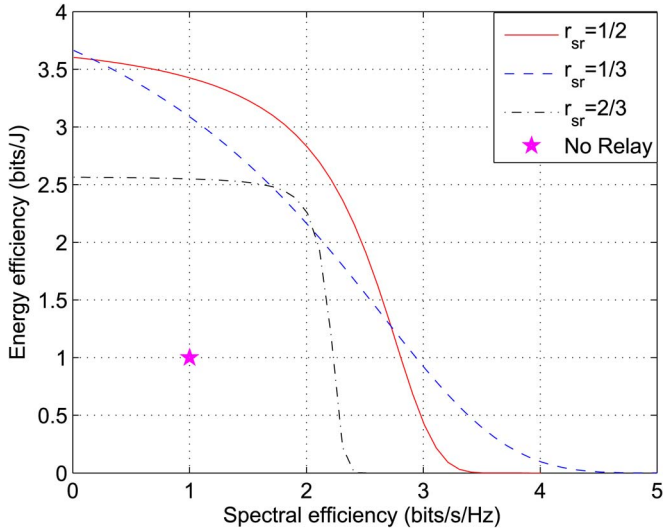

Fig. 8. Lower bounds of the EE-SE trade-off with varying relay locations and unconstrained θ and φ.

is obtained by a brutal force search for the EE-SE boundary from paired (θ, φ) values, each ranging from -15 dB to 15 dB. In other words, the Pareto front of optimization problem $\max_{\theta,\phi}(SE, EE)$ is approximated by a brutal force sampling of the (θ, φ) space. The second approach, denoted as 'implicit', is obtained from Corollary 2, i.e., the optimal (θ, φ) given by (12) is substituted into (8) and (25) to give SE and EE, respectively. This approach expresses EE and SE as implicit functions of the optimal θ and φ. The third approach is to directly calculate EE as a function of SE based on Proposition 5. The EE-SE tradeoff curves obtained by these three approaches are shown to be identical, thereby validating Proposition 5 and Corollary 2. Furthermore, based on Corollary 1, the maximum values of lower-bound EE and SE are calculated as $E_{\max} = 3.605$ bits/J and $S_{\max} = 4.086$ bits/s/Hz, respectively. These maximum values are shown to agree well with the tradeoff curves in Fig. 7. Apart from the lower bound, the upper bound of EE-SE tradeoff is also calculated based on Proposition 5 and shown to differ from the lower bound only by a small margin. Finally, the EE and SE without cognitive relay can be easily calculated as 1 bits/J and 1 bits/s/Hz, respectively. This is shown as the star sign on Fig. 7 as a reference point to indicate the benefits of deploying cognitive relay.

Fig. 8 aims to illustrate the impact of system parameters on the EE-SE tradeoff. The normalized source-relay distances are set to vary at 1/3, 1/2, and 2/3. It is observed that when the distance increases from 1/3 to 1/2, the tradeoff improves for smaller values of SE (SE < 2.7 bits/Hz/s) but degrades for larger values of SE (SE $>$ 2.7 bits/Hz/s). This is because for a lower targeted SE, the given licensed RR is abundant and the required cognitive RR can be reduced by moving the relay toward the destination. However, for a higher targeted SE, the licensed RR becomes the limiting factor and moving the relay toward the destination will reduce the performance limit. When the distance increases from 1/2 to 2/3, the tradeoff degrades significantly. This is because in this distance range the licence RR is always the limiting factor that caps the overall performance.

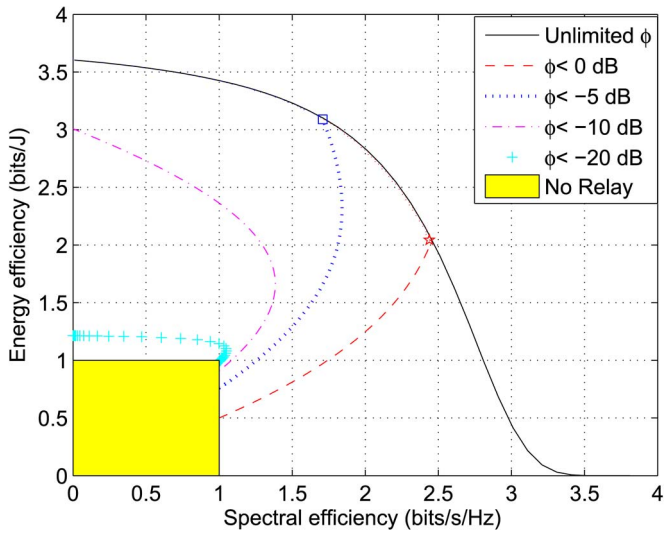

Fig. 9. Lower bounds of the EE-SE trade-off with varying power constraints ($\bar{\varepsilon} = 1, r_{sr} = 1/2$).

These observations imply the following guidelines: In general, it is better to deploy relays around the middle of source and destination. However, if the targeted SE is high, it is then better to deploy relays closer to the source.

In practice, both θ and φ should be limited. In the context of CR, it is likely to have a relatively large θ but small φ. Fig. 9 illustrates the impact of constrained φ on the EE-SE tradeoff. For a given range of φ, a "boundary curve" can be numerically evaluated. The area between a boundary curve and the Y-axis represents the feasible space of EE-SE pair. We see that when the maximum φ decreases, the EE-SE feasible space gradually reduces toward the Y-axis and finally collapses into the single line of EE $= 1$, $0 <$ SE < 1. From Fig. 9, we can easily infer that when the constraints on θ and φ gradually reduce, the EE-SE feasible space will shrink toward the point (EE $= 1$, SE $= 1$) from both axis.

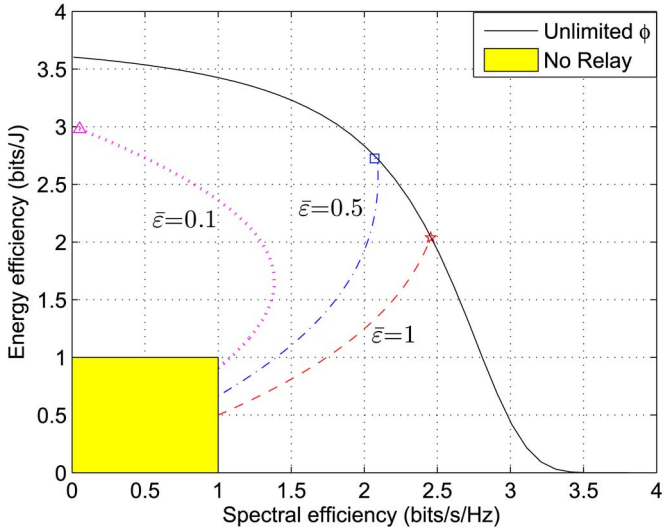

Fig. 10. Lower bounds of the EE-SE trade-off with varying reliability constraints ($\varphi < 0$ dB, $r_{sr} = 1/2$).

Finally, we would like to understand the impact of reliability parameter $\bar{\varepsilon}$ on the EE-SE tradeoff. When θ and φ are unconstrained, the EE-SE tradeoff is not affected by $\bar{\varepsilon}$. This may seem counter-intuitive at the first glance, but is evident from (34) and (35). This is because any reductions on $\bar{\varepsilon}$ can ultimately be compensated by increased θ and φ. Furthermore, because we consider the effective bandwidth $(\bar{\varepsilon}\theta W_1)$ and effective power $(\bar{\varepsilon}\varphi P_1)$ in the definitions of SE and EE (see (18) and (27)), the changes on $\bar{\varepsilon}$, θ, and φ tend to balance out. On the other hand, when θ and φ are limited, reliability does have a significant impact on the EE-SE tradeoff. Such an impact is illustrated in Fig. 10, which shows that the EE-SE feasible space reduces with decreasing reliability.

VI. Conclusion

This paper studies the multi-objective power-bandwidth allocation problem in HCGRC. The optimization objectives are to maximize the capacity, SE, and EE with respect to their upper or lower bounds. We have first derived the optimal power-bandwidth allocation strategies with respect to each single objective. The Pareto-optimal EE-SE tradeoff curve has been characterized analytically. The impact of relay location on resource allocation and EE-SE tradeoff has been studied. Our main observation is that the multi-objective power-bandwidth allocation problem can be conveniently characterized by five regions. Each region represents a unique tradeoff relationship among capacity, SE, and EE. Moreover, given unconstrained cognitive RR, reliability has no impact on the EE-SE tradeoff. However, once bandwidth or power is limited, reliability becomes an important factor that limits the feasible ranges of EE and SE. Our results are useful in providing basic guidelines for the design of hybrid cooperative cognitive radio systems. Future work can seek to extend the result from this paper to Rayleigh fading channels [43], multi-antenna scenarios, and multi-user multi-relay scenarios.

Appendix A

This appendix aims to derive the information-theoretic capacity lower bound of the HCGRC (Proposition 1). For the three mutual information terms in (4), $I(X;Y_1)$ and $I(X;\tilde{Y})$ can be easily evaluated. The key is to derive the middle term $I(\tilde{X};Y_2)$ because another source of randomness is introduced to Y_2 by ε. Recall that ε varies at a much slower rate than the transmit symbols $\tilde{X}$ and the channel capacity is evaluated over a sufficiently long time period. This allows us to calculate $I(\tilde{X};Y_2)$ in two steps: 1) calculate the conditional instantaneous mutual information given ε, and 2) calculate the long-term average mutual information by further taking the expectation over ε. Such a two-step calculation is a standard procedure in evaluating the ergodic capacity (i.e., long-term average capacity) of fading channels [44]. It follows that

$$\begin{aligned} I(\tilde{X};Y_2) &= \mathop{\mathbf{E}}_{\varepsilon}\left[I(\tilde{X};Y_2|\varepsilon)\right] \\ &= (1-\bar{\varepsilon})I(\tilde{X};Y_2|\varepsilon=0) + \bar{\varepsilon}I(\tilde{X};Y_2|\varepsilon=1) \\ &= \bar{\varepsilon}\left[\mathbf{H}(Y_2|\varepsilon=1) - \mathbf{H}(Y_2|\tilde{X},\varepsilon=1)\right] \\ &= \bar{\varepsilon}\left[\mathbf{H}(\sqrt{P_2 h_{rd}}\tilde{X}+Z_2) - \mathbf{H}(Z_2)\right] \\ &\leqslant W_1\theta\bar{\varepsilon}\log\left(\frac{P_1\varphi h_{rd}+N_0W_1\theta}{N_0W_1\theta}\right) \\ &= W_1\theta\bar{\varepsilon}\log\left(1+\frac{\rho_1\varphi h_{rd}}{\theta}\right) \end{aligned} \tag{38}$$

where operator $\mathbf{E}(.)$ returns expectation and $\mathbf{H}(.)$ returns entropy. The preceding inequality follows from the well-known theorem of band-limited Gaussian channel capacity [45]. The upper bound can be derived following similar procedures.

Appendix B

This appendix aims to derive the optimal power-bandwidth allocation for capacity maximization (Proposition 2). As shown in (6) and (7), the difference between the lower and upper bounds lies on the second item $C_{2,low}$ and $C_{2,up}$, both of which have non-zero values and are unrelated to θ and φ. On the other hand, $C_{1,low}$ and $C_{1,up}$ are monotonically increasing functions of both θ and φ. Therefore, (12) and (13) can be derived as the solutions of $C_{1,low}(\theta,\varphi) = C_{2,low}$ and $C_{1,up}(\theta,\varphi) = C_{2,up}$, respectively.

Appendix C

This appendix aims to derive the optimal power-bandwidth allocation for SE maximization (Proposition 3). First, analyzing the derivatives of (18) and (19) with respect to θ and φ, it can be easily shown that $S_{1,low}$ is a monotonically increasing function of φ and a convex function of θ. In addition, $S_{2,low}$ is a monotonically decreasing function of θ but unrelated to φ. Further observing (18) and (19), we can get $S_{1,low}(0,\varphi) < S_{2,low}(0,\varphi)$. Moreover, both $S_{1,low}$ and $S_{2,low}$ approach zero as θ goes to infinity. Consequently, for any given φ, there exists

a unique $\theta^*_{s(1)}$ that maximizes $S_{1,low}$ and one solution $\theta^*_{s(2)}$ that satisfies $S_{1,low}(\theta^*_{s(2)}) = S_{2,low}(\theta^*_{s(2)})$. The optimal θ^*_s that achieves the highest SE is either $\theta^*_{s(1)}$ or $\theta^*_{s(2)}$. It can be further drawn that there is a unique threshold point φ^*_{th} that gives $\theta^*_{s(1)} = \theta^*_{s(2)}$.

For a given power ratio φ, if $\varphi < \varphi^*_{th}$, the optimal resource allocation is $\theta^*_s = \theta^*_{s(1)}$, and the corresponding spectral efficiency is $S_{lower} = S_{1,low}\left(\theta^*_{s(1)}\right)$; otherwise, the optimal resource allocation is $\theta^*_s = \theta^*_{s(2)}$, and the corresponding spectral efficiency is $S_{lower} = S_{1,low}\left(\theta^*_{s(2)}\right) = S_{2,low}\left(\theta^*_{s(2)}\right)$.

Recall that $\theta^*_{s(1)}$ is the maximum point of $S_{1,low}$. Since $S_{1,low}$ is a continuous and differentiable function, we have $\frac{\partial S_{1,low}}{\partial \theta}\left(\theta^*_{s(1)}\right) = 0$, i.e.,

$$\log\left(1 + \frac{\varphi\rho_1 h_{rd}}{\theta^*_{s(1)}}\right) - \log(1+\rho_1 h_{sd}) - \frac{\varphi\rho_1 h_{rd}\left(\theta^*_{s(1)}\bar{\varepsilon}+1\right)}{\ln 2\left(\theta^*_{s(1)} + \varphi\rho_1 h_{rd}\right)} = 0. \quad (39)$$

Let $a = \varphi\rho_1 h_{rd}, b = 1+\rho_1 h_{sd}$,$x = (\theta+a)/(\theta b)$, (39) can be rewritten as

$$\ln(x) = 1 + \frac{a\bar{\varepsilon}-1}{bx}. \quad (40)$$

Let β=$(a\bar{\varepsilon}-1)/b$, (40) can be rearranged into the common form of the Lambert W function as $\frac{\beta}{e} = \frac{\beta}{x}e^{\frac{\beta}{x}}$. Consequently, the solution of (40) is $x = \beta/W_0(\beta/e)$. It follows that $\theta^*_{s(1)}$ can be calculated by (22). Recall that $\theta^*_{s(2)}$ is defined as the intersection of two curves, i.e., $S_{1,low}\left(\theta^*_{s(2)}\right) = S_{2,low}\left(\theta^*_{s(2)}\right)$. It follows that $\theta^*_{s(2)}$ should lay on the curve of $S_{2,low}$ and it is easy to show that $\theta^*_{s(2)} = \Psi^{-1}(\varphi)$.

APPENDIX D

This appendix aims to derive the optimal power-bandwidth allocation for SE maximization (Proposition 4). The analysis of EE is very similar to that of SE and there exists a strong duality between Propositions 3 and 4. The derivatives of $E_{1,low}$ and $E_{2,low}$ indicate that $E_{1,low}$ is a monotonically increasing function of θ and a convex function of φ. In addition, $E_{2,low}$ is a monotonically decreasing function of φ but unrelated to θ. Define $\varphi^*_{e(1)}$ as the maximum point of $E_{1,low}$. The calculation of $\varphi^*_{e(1)}$ can follow the same procedure as the derivation of $\theta^*_{s(1)}$ in Proposition 3. Define $\varphi^*_{e(2)}$ as the intersection of $E_{1,low}(\varphi)$ and $E_{2,low}(\varphi)$, it can be calculated as the inverse function of (8). Furthermore, the threshold θ^*_{th} is the solution of $\varphi^*_{e(1)}(\theta) = \Psi(\theta)$. It can be shown that the lower bound and upper bound of EE overlap when $\theta < \theta^*_{th}$.

APPENDIX E

This appendix aims to derive the EE-SE trade-off curve by expressing the maximum EE as a function of SE (Proposition 5). In other words, given S, we aim to derive $\max_{\theta,\phi}(E(S))$. Our proof takes three steps. The first step is to express EE as functions of S and θ. The second step is to prove that EE is a monotonically increasing function of θ. The final step is to obtain the maximum θ given S and substitute it into $E(\theta, S)$.

Step 1: Recall that EE and SE are both functions of θ and φ, from (18) and (27), we can express the two EE terms as functions of S and θ

$$E_{1,low}(\theta, S_{1,low}) = \frac{(\theta\bar{\varepsilon}+1)W_1 S_{1,low}}{P_1\left\{1 + \frac{\bar{\varepsilon}\theta}{\rho_1 h_{rd}}\left[\left(\frac{2^{(\theta\bar{\varepsilon}+1)S_{1,low}}}{1+\rho_1 h_{sd}}\right)^{\frac{1}{\theta\bar{\varepsilon}}} - 1\right]\right\}} \quad (41)$$

$$E_{2,low}(\theta, S_{1,low}) = \frac{W_1\log(1+\rho_3 h_{sr})}{P_1\left\{1 + \frac{\bar{\varepsilon}\theta}{\rho_1 h_{rd}}\left[\left(\frac{2^{(\theta\bar{\varepsilon}+1)S_{1,low}}}{1+\rho_1 h_{sd}}\right)^{\frac{1}{\theta\bar{\varepsilon}}} - 1\right]\right\}} \quad (42)$$

respectively.

Step 2: From (27), the partial derivative of $E_{1,low}$ over θ can be obtained as

$$\frac{\partial E_{1,low}}{\partial\theta} = \frac{W_1\bar{\varepsilon}}{P_1(1+\varphi\bar{\varepsilon})\ln(2)}\left[\ln\left(1+\frac{\varphi d}{\theta}\right) - \frac{\varphi d}{\theta+\varphi d}\right]. \quad (43)$$

We aim to prove that the above partial derivative is always greater than zero. Let $x = \frac{\varphi d}{\theta}$, $f(x) = \ln(1+x)$, $g(x) = \frac{x}{1+x}$, it is easy to see that our proof is consistent with proving $f(x) > g(x)$. To this end, the derivative of $f(x)$ and $g(x)$ can be derived as $f'(x) = \frac{1}{1+x}$ and $g'(x) = \frac{1}{(1+x)^2}$, respectively. Because $f'(x) > g'(x)$ and $f(0) = g(0)$, we have $f(x) > g(x)$. It follows that $\ln\left(1+\frac{\varphi d}{\theta}\right) > \frac{\varphi d}{\theta+\varphi d}$ and then $\frac{\partial E_1}{\partial\theta} > 0$. As a result, $E_{1,low}$ is monotonically increasing with bandwidth ratio θ. On the other hand, $E_{2,low}$ in (28) does not change with θ. Therefore, $E = \min(E_1, E_2)$ increases monotonically with θ before it reaches a maximum value.

Step 3: Given S, there is a maximum value of θ. This value can be calculated from (19) as

$$\theta^{\max}_{low} = \frac{\frac{\log(1+\rho_3 h_{sr})}{S_{low}} - 1}{\bar{\varepsilon}}. \quad (44)$$

With this $\theta^{\max}_{low}$, it is easy to show that (41) and (42) become identical, it follows that

$$E_{low}(\theta, S_{low}) = \frac{(\theta\bar{\varepsilon}+1)W_1 S_{low}}{P_1\left\{1 + \frac{\bar{\varepsilon}\theta}{\rho_1 h_{rd}}\left[\left(\frac{2^{(\theta\bar{\varepsilon}+1)S_{low}}}{1+\rho_1 h_{sd}}\right)^{\frac{1}{\theta\bar{\varepsilon}}} - 1\right]\right\}}, \quad \theta < \theta^{\max}_{low}. \quad (45)$$

Substituting (44) into (45), we obtain (34).

APPENDIX F

This appendix aims to derive the maximum values of SE and EE given unlimited power and bandwidth (Proposition 5). Let S

approach zero in (34) and taking the first order Taylor expansion we get

$$K_4^{\frac{S}{K_5-S}} = e^{\frac{S}{K_5-S}\ln(K_4)} \approx 1 + \frac{S}{K_5-S}\ln(K_4). \quad (46)$$

It follows that the maximum value of EE in lower bound is given by

$$\begin{aligned} E_{low}^{\max} &= \lim_{S\to 0} E_{low}(S) \\ &= \frac{K_1}{K_2 - \frac{K_3}{S} + K_3\left(\frac{1}{S} - \frac{1}{K_5}\right)K_4^{\frac{S}{K_5-S}}} \\ &\approx K_1 S\left[K_2 S + K_3 \cdot \frac{S}{K_5-S}\cdot\ln(K_4)\right. \\ &\qquad \left. -K_3\cdot\frac{S}{K_5}\cdot\left(1+\frac{S}{K_5-S}\cdot\ln(K_4)\right)\right]^{-1} \\ &= \frac{W_1\log(1+\rho_3 h_{sr})}{P_1\left[1+\frac{1}{\rho_1 h_{rd}}\ln\left(\frac{1+\rho_3 h_{sr}}{1+\rho_1 h_{sd}}\right)\right]}. \end{aligned} \quad (47)$$

Substituting (47) into (28), the minimum power ratio in lower bound can be obtained as (36).

The maximum value of SE is obtained when the bandwidth ratio approaches zero in (19). It follows that

$$S_{low}^{\max} = \lim_{\theta\to 0} S_{2,low} = \log(1+\rho_3 h_{sr}). \quad (48)$$

Similar proof applies to the upper bound.

REFERENCES

[1] "Facilitating opportunities for flexible, efficient, and reliable spectrum use employing cognitive radio technologies," Fed. Commun. Comm., Washington, DC, USA, FCC Rep. 03-322, Dec. 2003.

[2] J. Mitola III and G. Q. Maguire Jr., "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

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

[4] Z. Fei *et al.*, "Power allocation for OFDM-based cognitive heterogeneous networks," *Sci. China Inf. Sci.*, vol. 56, no. 4, Apr. 2013, Art. ID. 042310.

[5] X. Li *et al.*, "Resource allocation for MIMO-OFDMA downlink based cognitive radio systems with imperfect channel learning," *Sci. China Inf. Sci.*, vol. 56, no. 10, Oct. 2013, Art. ID. 102307.

[6] G. Bansal, M. J. Hossain, and V. K. Bhargava, "Optimal and suboptimal power allocation schemes for OFDM based cognitive radio systems," *IEEE Trans. Commun.*, vol. 7, no. 11, pp. 4710–4718, Nov. 2008.

[7] Y. Zhang and C. Leung, "Resource allocation in an OFDM-based cognitive radio system," *IEEE Trans. Commun.*, vol. 57, no. 7, pp. 1928–1931, Jul. 2009.

[8] Y. Zhang and C. Leung, "Resource allocation for non-real-time services in OFDM-based cognitive radio systems," *IEEE Commun. Lett.*, vol. 13, no. 1, pp. 16–18, Jan. 2009.

[9] S. Wang, "Efficient resource allocation algorithm for cognitive OFDM systems," *IEEE Commun. Lett.*, vol. 14, no. 8, pp. 725–727, Aug. 2010.

[10] H. Ding, J. Ge, D. B. da Costa, and Z. Jiang, "Energy-efficient and low-complexity schemes for uplink cognitive cellular networks," *IEEE Commun. Lett.*, vol. 14, no. 12, pp. 1101–1103, Dec. 2010.

[11] S. Wang, F. Huang, and Z. Zhou, "Fast power allocation algorithm for cognitive radio networks," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 845–847, Aug. 2011.

[12] W. Yao, Y. Wang, and T. Wang, "Joint optimization for downlink resource allocation in cognitive radio cellular networks," in *Proc. IEEE Consumer Commun. Netw. Conf.*, Las Vegas, NV, USA, Jan. 2011, pp. 664–668.

[13] Y. J. Zhang and A. M. So, "Optimal spectrum sharing in MIMO cognitive radio networks via semidefinite programming," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 362–373, Feb. 2011.

[14] M. Ge and S. Wang, "Fast optimal resource allocation is possible for multiuser OFDM-based cognitive radio networks with heterogeneous services," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1500–1509, Apr. 2012.

[15] S. Wang, F. Huang, and C. Wang, "Adaptive proportional fairness resource allocation for OFDM-based cognitive radio networks," *Wireless Netw.*, vol. 19, no. 4, pp. 273–284, Mar. 2013.

[16] A. Lam, V. Li, and J. Yu, "Power-controlled cognitive radio spectrum allocation with chemical reaction optimization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3180–3190, Jul. 2013.

[17] C. Zhai and W. Zhang, "Adaptive spectrum leasing with secondary user scheduling in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3388–3398, Jul. 2013.

[18] M. Naeem, A. Anpalagan, M. Jaseemuddin, and D. C. Lee, "Resource allocation techniques in cooperative cognitive radio networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 729–744, Jun. 2014.

[19] C.-X. Wang, X. Hong, H.-H. Chen, and J. S. Thompson, "On capacity of cognitive radio networks under average interference power constraints," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1620–1625, Apr. 2009.

[20] X. Hong, C.-X. Wang, M. Uysal, X. Ge, and S. Ouyang, "Capacity of hybrid cognitive radio networks with distributed VAAs," *IEEE Trans. Veh. Technol.*, vol. 59, no. 7, pp. 351–3523, Sep. 2010.

[21] M. Z. A. Khan and U. B. Desai, "User allocation for tri-band cognitive cellular networks," in *Proc. 2nd U.K.-India-IDRC Int. Workshop Cognitive Wireless Syst.*, New Delhi, India, Dec. 2010, pp. 1–5.

[22] M. D. Mueck, M. Di Renzo, and M. Debbah, "Opportunistic relaying for cognitive radio enhanced cellular networks: Infrastructure and initial results," in *Proc. 5th IEEE Int. Symp. Wireless Pervasive Comput.*, Modena, Italy, May 2010, pp. 556–561.

[23] Y.-J. Choi and K. G. Shin, "Opportunistic access of TV spectrum using cognitive-radio-enabled cellular networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3853–3864, Oct. 2011.

[24] T. Zhang, Y. Wu, K. Lang, and D. H.-K. Tsang, "Joint spectrum allocation and relay selection in cellular cognitive radio networks," *Mobile Netw. Appl.*, vol. 16, no. 6, pp. 748–759, Dec. 2011.

[25] K. Xue, X. Hong, L. Chen, J. Shi, and C.-X. Wang, "Performance analysis and resource allocation of heterogeneous cognitive Gaussian relay channels," in *Proc. GLOBECOM*, Atlanta, GA, USA, Dec. 2013, pp. 1167–1172.

[26] M. Gastpar, "On capacity under receive and spatial spectrum-sharing constraints," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 471–487, Feb. 2007.

[27] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1813–1827, May 2006.

[28] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4391–4399, Nov. 2007.

[29] S. Srinivasa and S. A. Jafar, "The throughput potential of cognitive radio: A theoretical perspective," *IEEE Commun. Mag.*, vol. 45, no. 5, pp. 73–79, May 2007.

[30] S. A. Jafar and S. Srinivasa, "Capacity limits of cognitive radio with distributed and dynamic spectral activity," *IEEE. J. Sel. Areas Commun.*, vol. 25, no. 3, pp. 529–537, Apr. 2007.

[31] X. Kang, Y.-C. Liang, A. Nallanathan, H. K. Garg, and R. Zhang, "Optimal power allocation for fading channels in cognitive radio, networks—Ergodic and outage capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 940–950, Feb. 2009.

[32] L. Musavian and S. Aissa, "Capacity and power allocation for spectrum-sharing communications in fading channels," *IEEE. Trans. Wireless Commun.*, vol. 8, no. 1, pp. 148–156, Jan. 2009.

[33] M. F. Hanif and P. J. Smith, "On the statistic of cognitive radio capacity in shadowing and fast fading environments," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 844–852, Feb. 2010.

[34] L. Musavian and S. Aissa, "Effective capacity of delay-constrained cognitive radio in Nakagami fading channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1054–1062, Mar. 2010.

[35] S. Sridharan and S. Vishwanath, "On the capacity of a class of MIMO cognitive radios," *IEEE J. Sel. Signal Process.*, vol. 2, no. 1, pp. 103–117, Feb. 2008.

[36] R. Zhang and Y.-C. Liang, "Exploiting multi-antennas for opportunistic spectrum sharing in cognitive radio networks," *IEEE. J. Sel. Signal Process.*, vol. 2, no. 1, pp. 88–102, Feb. 2008.

[37] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.
[38] Y. Liang and V. V. Veeravalli, "Gaussian orthogonal relay channels: Optimal resource allocation and capacity," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 328–3289, Sep. 2005.
[39] J. Vilardebo, I. Ana, and M. Najar, "Energy efficient communications over the AWGN relay channel," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, pp. 32–37, Jan. 2010.
[40] D. Feng *et al.*, "A survey of energy-efficient wireless communications," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 167–178, Feb. 2013.
[41] X. Hong, J. Yu, C.-X. Wang, J. Shi, and C.-X. Ge "Energy-spectral efficiency trade-off in virtual MIMO cellular systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2128–2140, Oct. 2013.
[42] S. Huang, H. Chen, J. Cai, and F. Zhao, "Energy efficiency and spectral-efficiency trade-off in amplify-and-forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4366–4378, Nov. 2013.
[43] X. Hong, J. Wang, C.-X. Wang, and J. Shi, "Cognitive radio in 5G: A perspective on energy-spectral efficiency trade-off," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 46–53, Jul. 2014.
[44] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley Inter-Science, 2006.
[46] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, no. 1, pp. 329–359, Dec. 1996.
[47] L. Xu, H.-W. Zhang, X.-H. Li, and X.-L. Wu, "Optimum relay location in cooperative communication networks with single AF relay," *Int. J. Commun. Netw. Syst. Sci.*, vol. 4, no. 3, p. 147, Mar. 2011.

Xuemin Hong (S'05–M'12) received the Ph.D. degree from Heriot-Watt University, U.K., in 2008. Since 2011, he has been an Associate Professor at Xiamen University, China. From 2009 to 2011, he was a Post-doc Research Fellow at the University of Waterloo, Canada and Heriot-Watt University, U.K. He has published one book chapter and over 40 papers in refereed journals and conference proceedings. His research interests include cognitive radio networks, wireless channel modeling, and fifth-generation wireless communications.

Chao Zheng received the B.Eng. degree from Xiamen University, China, in 2013. He is currently a postgraduate student in the School of Information Science and Engineering, Xiamen University. His current research interests include cognitive radio networks and fifth-generation wireless communications.

Jing Wang received the B.Eng. degree from Henan Normal University, China, in 2012. She is currently a postgraduate student in the School of Information Science and Engineering, Xiamen University. Her research interests include cognitive radio networks and fifth-generation wireless communications.

Jianghong Shi received the Ph.D. degree from Xiamen University, China, in 2002. He is currently a Professor in the School of Information Science and Engineering, Xiamen University. He is also the Director of the West Straits Communications Engineering Center, Fujian Province, China. His research interests include wireless communication networks and satellite navigation systems.

Cheng-Xiang Wang (S'01–M'05–SM'08) received the B.Sc. and M.Eng. degrees in communication and information systems from Shandong University, Jinan, China, in 1997 and 2000, respectively, and the Ph.D. degree in wireless communications from Aalborg University, Aalborg, Denmark, in 2004.

Since 2005, he has been with Heriot-Watt University, Edinburgh, U.K., where he was promoted to Professor in 2011. He is also an Honorary Fellow of the University of Edinburgh, Edinburgh, and a Chair/Guest Professor with Shandong University, and with Southeast University, Nanjing, China. From 2001 to 2005, he was a Research Fellow with the University of Agder, Grimstad, Norway. In 2004, he was a Visiting Researcher with Siemens AG-Mobile Phones, Munich, Germany. From 2000 to 2001, he was a Research Assistant with the Hamburg University of Technology, Hamburg, Germany. He is the Editor of one book. He has published one book chapter and over 210 papers in refereed journals and conference proceedings. His research interests include wireless channel modeling and simulation, green communications, cognitive radio networks, vehicular communication networks, massive multiple-input multiple-output systems, and fifth-generation wireless communications.

Dr. Wang is a Fellow of the Institution of Engineering and Technology and the HEA and a member of the Engineering and Physical Research Council Peer Review College. He has served as an Editor for eight international journals, including the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (since 2011) and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–2009). He was the lead Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, Special Issue on Vehicular Communications and Networks. He has served as a Technical Program Committee (TPC) Member, TPC Chair, and General Chair for more than 70 international conferences. He received the Best Paper Awards from IEEE Globecom 2010, IEEE ICCT 2011, ITST 2012, and IEEE VTC 2013-Spring.