Cognitive Radio in 5G: A Perspective on Energy-Spectral Efficiency Trade-off

Xuemin Hong, Jing Wang, Cheng-Xiang Wang, and Jianghong Shi

ABSTRACT

A cognitive cellular network, which integrates conventional licensed cellular radio and cognitive radio into a holistic system, is a promising paradigm for the fifth generation mobile communication systems. Understanding the trade-off between energy efficiency, EE, and spectral efficiency, SE, in cognitive cellular networks is of fundamental importance for system design and optimization. This article presents recent research progress on the EE-SE trade-off of cognitive cellular networks. We show how EE-SE trade-off studies can be performed systematically with respect to different architectures, levels of analysis, and capacity metrics. Three representative examples are given to illustrate how EE-SE trade-off analysis can lead to important insights and useful design guidelines for future cognitive cellular networks.

Xuemin Hong, Jing Wang, and Jianghong Shi are with Xiamen University.

Cheng-Xiang Wang (corresponding author) is with Shandong University and Heriot-Watt University.

The authors acknowledge the support from the Natural Science Foundation of China (Grant No. 61201195), Fujian Natural Science Foundation (Grant No. 2012J01292), the 863 project in 5G wireless networking, Ministry of Science and Technology (Grant No. SS2014AA012101), and the Opening Project of the Key Laboratory of Cognitive Radio and Information Processing (Guilin University of Electronic Technology), Ministry of Education (No. 2013KF01).

INTRODUCTION

The quick proliferation of media-rich mobile devices has caused users to demand ubiquitous mobile broadband that can provide services comparable to the fixed broadband Internet. It is reported that the actual mobile traffic in 2010 is more than five times greater than an official forecast made by the International Telecommunication Union (ITU) in 2005. Recent forecasts suggest that the mobile data demand may continue to increase on the order of hundreds of times in the next decade. To accommodate the exploding traffic at low cost is critical to ensure the competitiveness of cellular networks in the future. To this end, research initiatives on the fifth generation (5G) cellular communication systems have gained accelerating momentum globally [1].

The challenges faced by 5G cellular networks are multifold. Besides the exploding traffic volume, mobile data is becoming increasingly random and diverse. First, mobile traffic is distributed unevenly across space and time. In fixed Internet, it was observed that the peak-to-mean traffic ratio can reach over 100:1. For future data-oriented mobile broadband networks, an even greater ratio is expected due to strong correlations in people clustering, device functions, and local events. Such a drastic traffic variation causes an obvious dilemma in planning the network infrastructure: the capacity is either insufficient for peak traffic demands or overabundant (and therefore cost-ineffective) for average traffic loads. Second, the mobile traffic is becoming increasingly diverse. Much of the mobile data demand is driven by new applications like web browsing, gaming, social media, and multimedia download, all of which have different quality of service (QoS) requirements than traditional voice services. Conventional cellular networks built on expensive licensed bands and reliable core networks are typically optimized to deliver low-volume delay-sensitive services such as voice. Various high-volume delay-insensitive data services such as multimedia downloads are still too cost-ineffective to gain massive popularity over current cellular networks.

Cognitive radio [2] is a promising technology to tackle, or at least partly address, the above challenges in 5G cellular networks. Under the current spectrum regulation policy, a cellular system can only operate on the licensed band with fixed limited bandwidth. While the cellular bands are heavily capitalized and exploited, measurements have revealed that some frequency bands licensed to other incumbents are significantly underutilized. Cognitive radio allows a cellular network to dynamically lease the underutilized frequency bands without causing harmful interference to the incumbents. As the leased spectrum is fundamentally opportunistic and unreliable, the cost of leasing the spectrum is expected to be much lower than the cost of purchasing a licensed band. Therefore, cognitive radio allows a cellular network to expand its spectrum on demand at a relatively low cost, thereby offering a natural solution to cope with exploding, random, and diverse mobile data traffic. For example, the leased bands can be used to cope with overload traffic in peak hours or provide opportunistic multimedia downloads at low cost.

Cognitive cellular networks, as defined in this article, are cellular networks that employ cognitive radio to lease addition spectrum outside the licensed cellular bands [3]. The radio resource (RR) at a particular band can be characterized by the bandwidth, maximum transmit power, and reliability (or availability). The RRs in a cognitive cellular network include licensed (cellular band) RR and cognitive RR. The licensed RR

 0163-6804/14/$25.00

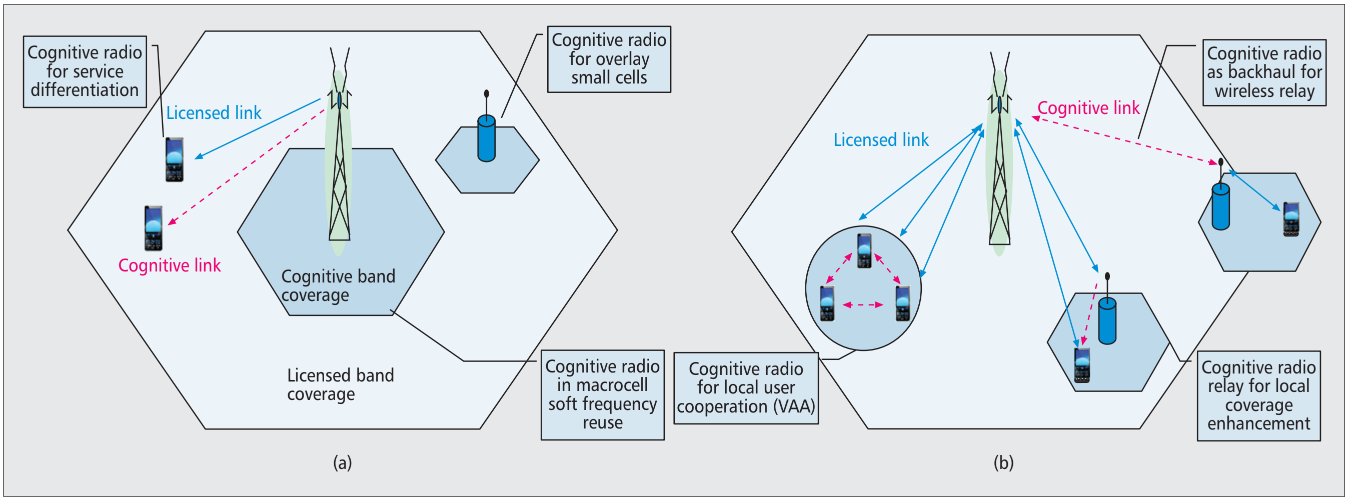

Figure 1. Architecture and usage scenarios of cognitive cellular networks: a) non-cooperative architecture; b) cooperative architecture.

has relatively small bandwidth, high transmit power, and high reliability. On the contrary, the cognitive RR is typically characterized by its potentially broad bandwidth, low transmit power, and low reliability. It is obvious that these two types of RR are complementary in nature and demand different system designs for effective utilization.

The unique research challenge of cognitive cellular networks is to jointly utilize licensed and cognitive RRs to optimize the overall system performance. To this end, it is important to understand the energy efficiency (EE)/spectral efficiency (SE) trade-off in cognitive cellular networks. The SE indicates how efficiently the system bandwidth is utilized, while the EE measures how efficiently the energy is consumed. It is well known that maximizing EE and SE are conflicting objectives, and there is a fundamental trade-off between them [4]. Analysis of the EE-SE trade-off offers a balanced view of the nature of a communication system and has gained popularity in recent years, providing guidelines for wireless system design and optimization. Apart from its theoretical importance, EE-SE trade-off study has a unique practical value for cognitive cellular networks. It is envisioned that the operational expenditure (OPEX) of cognitive cellular networks will consist of spectrum leasing bills (related to bandwidth and SE) and electricity bills (related to EE) as two significant parts. Therefore, understanding the EE-SE trade-off can provide direct guidelines to the OPEX management of cognitive cellular networks.

Research on cognitive cellular networks is still in an early stage. This article aims to provide an overview on recent advances in this field from the perspective of EE-SE trade-off. In the remainder of the article, we introduce different architectures and usage scenarios of cognitive cellular networks. We give an overview on various capacity definitions applicable to the study of the EE-SE trade-off. Three examples are subsequently presented to illustrate how EE-SE trade-off analysis can give interesting insights into system design and optimization.

Architecture of Cognitive Cellular Networks

An architecture enables a cognitive cellular network to effectively integrate conventional licensed radio and cognitive radio into a holistic system. The architectures of cognitive cellular networks can be categorized broadly into two types: non-cooperative and cooperative [5].

Non-Cooperative Architecture

As illustrated in Fig. 1a, in a non-cooperative network, there are two separate radio interfaces operating at the licensed and cognitive RRs, respectively. In other words, the cognitive RR is used to build a new network (i.e., a standalone cognitive network) that overlays the existing licensed cellular network. The two networks are separated in the physical layer, but can be integrated in upper layers to perform joint scheduling. A base station (BS) may have both or one of the radio interfaces. In a broader context, a non-cooperative network falls into the category of multi-radio access technology (RAT) systems.

The system capacity of a non-cooperative cognitive cellular network is simply the sum of the two networks. In many cognitive bands, tight transmit power constraints are imposed to protect the incumbents. Consequently, the capacity of the standalone cognitive network usually diminishes quickly with increasing communication range [5], making it only worthwhile to deploy the network for short- to-medium range communications.

As illustrated in Fig. 1a, there are multiple usage scenarios for non-cooperative cognitive cellular networks. Similar to the philosophy of soft frequency reuse, one usage scenario is to use the power-limited cognitive RR for users near a macrocell BS, while the licensed RR is reserved to serve users further away. Another possible usage scenario highlights service differentiation. For example, services with strict QoS requirements can be scheduled to the (higher cost and more reliable) licensed radio interface, whereas services with relaxed QoS requirements

It is worth noting that non-cooperative cognitive small cells are being actively pursued by the industry, with several white papers recently published by major industrial players to advocate the joint deployment of cognitive radio and small cells.

can be delivered over the (lower cost and less reliable) cognitive radio interface. Another promising case is to deploy cognitive small cells (or femtocells) that use cognitive RR to cover traffic hotspots or coverage gaps. Compared to licensed-band-based small cells, cognitive small cells can offer potentially higher capacity and better interference protection to the macrocell. Finally, it is worth noting that non-cooperative cognitive small cells are being actively pursued by the industry, with several white papers recently published by major industrial players to advocate the joint deployment of cognitive radio and small cells.

Cooperative Architecture

Cooperative cognitive cellular networks were proposed in [3] to use both licensed and cognitive RRs to design a single integrated physical layer using principles of cooperative communications. Cooperative communications allow distributed users to process and relay information in a coordinated fashion to achieve significant performance gains. By breaking a point-to-point communication process into multiple phases among multiple entities, cooperative communications typically create heterogeneous wireless channels in the system. For example, consider a group of nearby users cooperating to deliver information to a distant BS; the wireless channels among users will be very different from the channels between the users and the BS. Recall that a cognitive cellular network has heterogeneous RRs in the licensed and cognitive bands. The rationale of cooperative cognitive cellular networks is to carefully match the heterogeneous RRs to the heterogeneous channels. For example, licensed RRs are better used for long-range communications, while cognitive RRs are better used for short-range communications to facilitate local cooperation.

The system capacity of a cooperative cognitive cellular network was investigated in [3, 5]. It was found that compared to non-cooperative systems, cooperative cognitive cellular networks have major advantages in being able to enhance the capacity of longer-range communications. Moreover, the capacity gains are more stable against fluctuations of the cognitive RRs.

Two typical usage scenarios of cooperative cognitive cellular networks are illustrated in Fig. 1b. In one scenario, a cognitive relay is deployed for coverage or capacity enhancement. A cognitive relay can communicate with a BS using licensed RRs and provide local coverage using cognitive RRs. Such a cognitive relay is able to work in a duplex fashion to outperform conventional relays. Alternatively, cognitive RRs can be used for backhaul and licensed RRs for local coverage. This alternative setting is promising when the cognitive band is located in super high frequency (SHF) (i.e., 3–30 GHz) and is suitable for fixed microwave access. A unique merit of this alternative configuration is that no modification is needed for conventional user devices that only operate in the licensed band. Considering another scenario, nearby mobile devices can use cognitive RRs to form a virtual antenna array (VAA) [5]. The VAA can then establish a virtual multiple-input multiple-output (MIMO) link in the licensed band to harvest performance gains comparable to a MIMO system. Finally, from an industrial perspective, the recent introduction of carrier aggregation (CA), advanced relay, and device-to-device (D2D) technologies into the Long-Term Evolution-Advanced (LTE-A) standards paves the way for cooperative cognitive cellular networks to enter major 5G standards.

Capacity of Cognitive Cellular Networks

Spectrum (or bandwidth) and energy (or power) are two types of natural resources underpinning all wireless communication systems. SE and EE measure, respectively, how effectively the spectrum and energy is utilized. While bits per second per Hertz is universally used as the metric for SE, EE metrics can take various forms such as energy per bit to noise power spectral density ratio (i.e., E_b/N_0), bit per Joule capacity, rate per energy, or Joule per bit [4]. These EE metrics are essentially the same and mutually convertible. The energy consumed by a cellular system has multiple types, including the energy radiated into the air, energy consumed by the circuit (in signal processing and power amplification), energy consumed by supporting facilities (e.g., a cooling system), and energy embodied in the infrastructures [6]. Our discussion in this article is restricted to the radiated energy as it relates to the fundamentals of a communication system.

As both EE and SE metrics are directly related to the "capacity" measured in bits per second, it is necessary to first introduce the various capacity definitions in wireless communication systems. For ad hoc networks, the commonly used capacity metrics are transport capacity and transmission capacity. For centralized networks such as cellular systems, the capacity studies can be roughly classified into three levels: link level, cell level, and system level.

The link-level study is information-theoretic-oriented, and focuses on the performance of a single source and destination pair. For additive white Gaussian noise (AWGN) channels, the classic Shannon capacity is used to characterize the maximum error-free data rate the channel can support. For fading channels with varying channel gains, the instantaneous capacity becomes a random variable and is subject to statistical measures. To this end, two types of capacity definitions can be distinguished. One is ergodic capacity [7, 8], which is the expectation of the random capacity over the distribution of fading channel states. The capacity-achieving code must be sufficiently long so that a received codeword is affected by all possible fading states. The other capacity definition is outage capacity [9, 10], which is defined as the largest rate of reliable communication at a certain outage probability p. In other words, given an outage probability p, the outage capacity is the largest rate that the channel can support with probability $1 - p$.

Capacity with outage allows messages sent over a given transmission burst to be decoded with some probability. Clearly, the definition of outage capacity is closely related to the trade-off between capacity and reliability.

Cell- and system-level capacity studies are network-engineering-oriented and focus on the performance of large-scale networks. Cell-level study focuses on a single cell with one BS and multiple users. It extends the link-level study to the multi-user regime by considering various cell-level factors such as multiple access, scheduling, user distribution, and path loss. A commonly used capacity metric at the cell level is the aggregate/sum capacity measured at the BS. Typically, the aggregated capacity is also a random variable and can be characterized by its mean or distribution/outage.

System-level capacity study further extends cell-level study to include multiple BSs. It addresses system-level considerations such as spatial distribution of BSs and intercell interference. Traditionally, the cellular network is modeled by the hexagon model, in which the coverage of a cell is assumed to be a hexagon, and the BSs are placed in regular grids. In recent years, stochastic geometry models [11] have gained increasing popularity to describe the locations of BSs as random spatial point patterns. This is because in practice, BS location is usually randomized due to terrain features, site availability, and local coverage requirements. In future cellular networks with heterogeneous and smaller cells, the randomness of BS locations is expected to increase. The stochastic geometry model of cellular networks enables a wide range of analytical capacity studies. Under this framework, the capacity of a "typical user" at an arbitrary location is of concern. The capacity of a typical user is random due to random locations, random interference, and random channel gains. Similar to link-level capacity concepts, system-level capacity can be evaluated by its mean or outage. It is worth noting that the outage capacity at the system level has two interpretations. For example, given an outage capacity C with outage probability p, one interpretation is that for an arbitrary user, the probability that it can communicate with the BS with a capacity greater than C is $1 - p$. Another interpretation is related to coverage, which states that the fraction of areas that provide coverage with capacities greater than C is $1 - p$. Both interpretations are equivalent as long as the underlying stochastic geometry models are spatially ergodic, which is a widely used assumption in the literature.

EE-SE Trade-off of Cognitive Cellular Networks

As mentioned above, cognitive cellular networks have various architectures. For each architecture, the capacity can be studied at three levels. For each level, at least two types of capacity definitions can be applied: ergodic capacity and outage capacity. Consequently, a series of EE-SE trade-off studies can be conducted with respect to each combination of the architectures, levels, and capacity definitions. We note that this is still largely an open research field. In what follows, we give three representative examples to show how EE-SE trade-off study can lead to interesting insights about cognitive cellular networks. These examples are carefully chosen as they not only cover all the architectures, levels, and capacity types, but also reflect the latest research progress and what we consider as the most promising application scenarios. In particular, the three examples are selected to address the following combinations:

- Example 1: Non-cooperative cognitive small cell + system level + outage capacity
- Example 2: Cooperative cognitive VAA + cell level + ergodic capacity
- Example 3: Cooperative cognitive relay + link level + ergodic capacity

In practice, cognitive RRs vary significantly in different frequency bands. Such variation is caused by different characteristics of the incumbents (e.g., transmitter behavior and receiver protection requirements) and different coexisting mechanisms (e.g., underlay, interweave, and overlay) [2]. To avoid unnecessary details, it is useful to consider an abstraction of the cognitive RR characterized by three parameters: cognitive bandwidth W_c, cognitive power P_c, and reliability parameter k. Here $0 < k < 1$ indicates the probability that the cognitive band is available at a given moment. Let W_o and P_o denote the bandwidth and power of the licensed RR; it is usually useful to consider the bandwidth ratio $\theta = W_c/W_o$ and power ratio $\varphi = P_c/P_o$.

Cognitive cellular networks have various architectures. For each architecture, the capacity can be studied at three levels. For each level, at least two types of capacity definitions can be applied: ergodic capacity and outage capacity.

Example 1

In this example, we consider the system-level outage capacity of non-cooperative cognitive small cells. Because the physical layers in the licensed and cognitive bands are independent, our interest is in understanding the EE-SE trade-off of a standalone cognitive small cell network. Consider a scenario of massively deployed cognitive small cells in a 2D plane. At the center of each cell sits a cognitive BS transmitting with identical power. The cognitive BSs are randomly deployed as a stationary Poisson point process with density λ. All channels are subject to Rayleigh fading and path loss with an exponent of 4. Consider the downlink capacity for an arbitrary user; a closed-form formula of the outage capacity can be obtained following the analytical framework introduced in [11].

In Fig. 2, we show the EE-SE trade-off derived from outage capacity. The EE metric we used is bits per Joule. To gain better insight, the EE is normalized over $\pi^2\lambda^2$. There are three factors that can influence the EE-SE trade-off: BS density λ, outage probability p, and cognitive RR reliability k. For simplicity, $k = 1$ is assumed in Fig. 2. The following interesting observations are made. First, the density of BSs has little impact on the normalized EE. This means that the actual EE is proportional to λ^2. Second, both SE and EE increase with increasing outage probability. This indicates that reliability or coverage can be traded for better SE and EE performance. Third, a nearly linear trade-off is observed between SE and normalized EE. Fourth, for a given outage, there is a maximum value for both SE and EE. This is different from the classic result derived from the Shannon capacity formula, where SE can go to infinity. The reason is that intercell interference is considered at the system level, and hence the performance is interference-limited. Finally, we can see that both EE and SE increase with acceleration with increasing outage probability.

Now let us turn our attention to the reliability parameter k. A simple mapping exists between k and p as they are both measures of reliability. Specifically, for any cognitive radio system with k and outage requirement p, its EE-SE trade-off is equivalent to another system with $k = 1$ and outage requirement $1 + (p - 1)/k$. For example, for $k = 0.6$ and $p = 0.7$, the EE-SE trade-off is the same as $k = 1$ and $p = 0.5$. The significant impact of k on the EE-SE trade-off can be observed in Fig. 2. Given $p = 0.7$ and k changing from 1 to 0.6, the EE-SE trade-off will degrade from the upmost curve to the middle one.

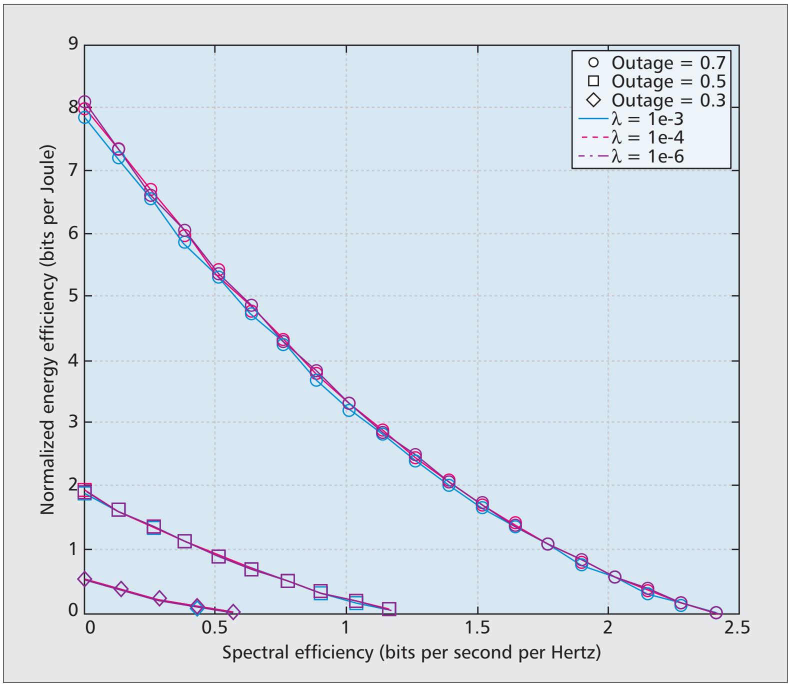

Figure 2. EE-SE trade-off of distributed cognitive small cells at the system level.

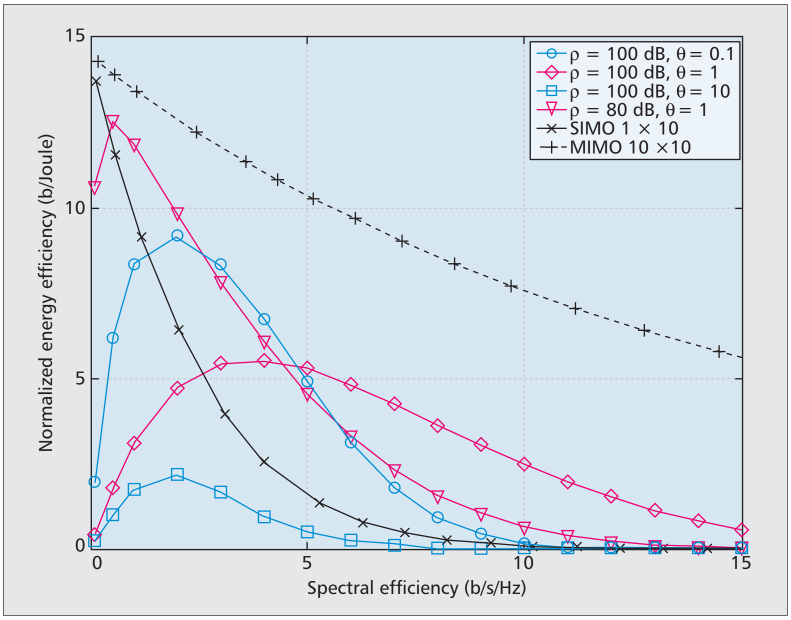

Figure 3. EE-SE trade-off of cognitive cooperative VAA systems at the cell level ($\lambda_b = 1/\text{km}^2$, $\lambda_v = 20/\text{km}^2$, $\lambda_u = 200/\text{km}^2$).

EXAMPLE 2

In this example, we consider cell-level ergodic capacity of cooperative cognitive VAA systems. As far as ergodic capacity is concerned, the impact of the reliability parameter k becomes trivial as it only causes a linear scaling on the cognitive radio capacity. Therefore, we subsequently assume $k = 1$ for simplicity. The scenario of interest is the uplink of a large cellular virtual MIMO network with multiple BSs and mobile users. A fraction of the users are active users (sources) that have messages to transmit to the BS. Other users are cooperative users (relays) that are willing to form VAAs with active users to facilitate virtual MIMO transmission. A virtual MIMO transmission consists of two phases: local broadcasting (phase I) and distributed MIMO access (phase II). We assume that phases I and II operate in the cognitive and licensed bands, respectively, in a frequency-division fashion. Each BS is equipped with r antennas, while each mobile user has only one antenna. Rayleigh fading and path loss with an exponent of 4 is assumed. The spatial distributions of BSs, mobile users, and active users are assumed to be stationary Poisson point processes with densities λ_b, λ_u, and λ_v, respectively.

We assume that an active user only communicates with the nearest BS. Similarly, an inactive user only cooperates with the nearest active user. In the first phase, each active user broadcasts its own message with a transmit power ρ (normalized to the noise power). An inactive user becomes a participant of the VAA only when it can successfully decode the message transmitted from the nearest active user. Let θ be the bandwidth ratio. We are interested in how the EE-SE trade-off varies with ρ and θ. Following similar steps in [12], the EE-SE trade-off can be evaluated numerically. We note that our calculation does not take into account the intercell interference at the BS; therefore, it is regarded as a cell-level study.

Figure 3 compares the EE-SE trade-off of the above cooperative VAA system with corresponding MIMO and single-input multiple-output (SIMO) systems. The number of antennas at the BS is assumed to be 10. The EE is normalized over $\pi^2\lambda_b^2$ for better clarity. The following insights can be obtained from Fig. 3. First, compared to the SIMO system (i.e., no VAA), VAA is only beneficial for high SE values. This is because the philosophy of VAA is to invest extra resources in phase I in exchange for some multiplexing gains in phase II. The benefits of multiplexing gains become significant only at high SE values. Second, the EE and SE are not necessarily in a trade-off relationship. When both EE and SE are relatively small, they can both increase simultaneously. Third, initial investments in cognitive RR (i.e., increase either ρ or θ) can improve the performance for higher SE values at the cost of degrading the performance at lower SE values. However, if the cognitive RR continues to increase, the capacity benefit of VAA is eventually saturated, and both the EE and SE suffer from the overprovision of cognitive RR. Overprovisioning of cognitive RR occurs when the capacity does not increase at the same proportion as the increase of power or bandwidth (e.g., due to the well-known log-shape capacity-bandwidth curve or self-interference). Such an overprovisioning phenomenon can be observed clearly in Fig. 3. For example, comparing the three curves with $\rho = 100$ dB, an initial increase of bandwidth ratio θ from 0.1 to 1 improves EE at the higher SE regime (SE > 5 b/s), while a further increase of θ from 1 to 10 only serves to degrade the EE-SE performance.

EXAMPLE 3

In this example, we consider link-level ergodic capacity of cooperative cognitive relay systems. In particular, we consider a simple relay system with three nodes, where a source broadcasts to a relay and a destination using licensed RRs, while the relay forwards information to the destination using cognitive RRs in a duplex fashion. In a traditional relay channel, the source and relay are subject to a total resource constraint. The cognitive relay channel we consider is fundamentally different from the conventional relay channel [13] in that the source and relay are subject to separate resource constraints. Rayleigh fading and path loss with an exponent of 4 are assumed.

Following the derivations in [14], the lower and upper bounds of the link-level capacity for the cognitive relay channel can be calculated. As the gaps between lower and upper bounds are small, the lower bound capacity is used for the evaluation of EE and SE. Without loss of generality, the bandwidth and power of the licensed band is set to 1, and the relay is located halfway between the source and destination. We are interested in how the EE-SE trade-off varies with different values of cognitive bandwidth ratio θ and power ratio φ. We let $k = 1$ because its impact on the ergodic capacity is a trivial issue.

Figure 4 shows the EE-SE trade-off curve of the cognitive relay channel. Each curve is calculated by fixing θ or φ and varying the other. Two key observations are made in Fig. 4. First, for a given θ or φ, there is a maximum SE and EE. The reason is that the capacity of the relay channel is ultimately limited by the (predetermined) licensed RR according to the cut-set bounds, which state that the overall channel capacity is the minimum of two capacities related to the licensed and cognitive RRs, respectively [14]. Second, EE and SE are not necessarily in a trade-off relationship; there are cases where both EE and SE can be improved simultaneously. This happens when the channel capacity is severely limited by the cognitive RR (i.e., when θ and φ are relatively small), so the licensed and cognitive RRs are highly imbalanced. In this case, a small linear increase of θ or φ will lead to nearly linear increase of the overall capacity, while the total resource input $1 + \theta$ or $1 + \varphi$ will scale at a much slower rate.

For a cognitive relay channel, we are particularly interested in the following question: given

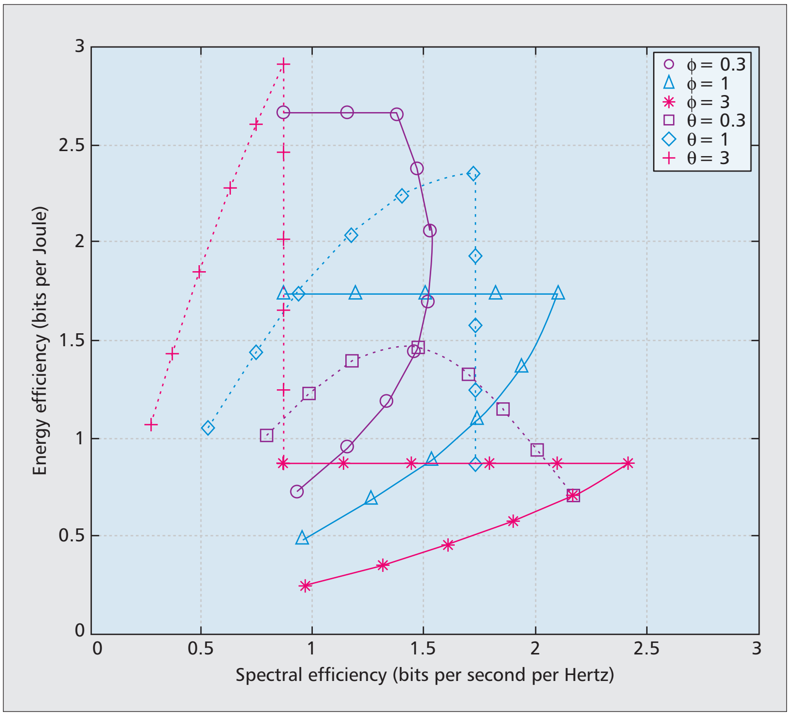

Figure 4. EE-SE trade-off of cognitive Rayleigh relay channel at the link level.

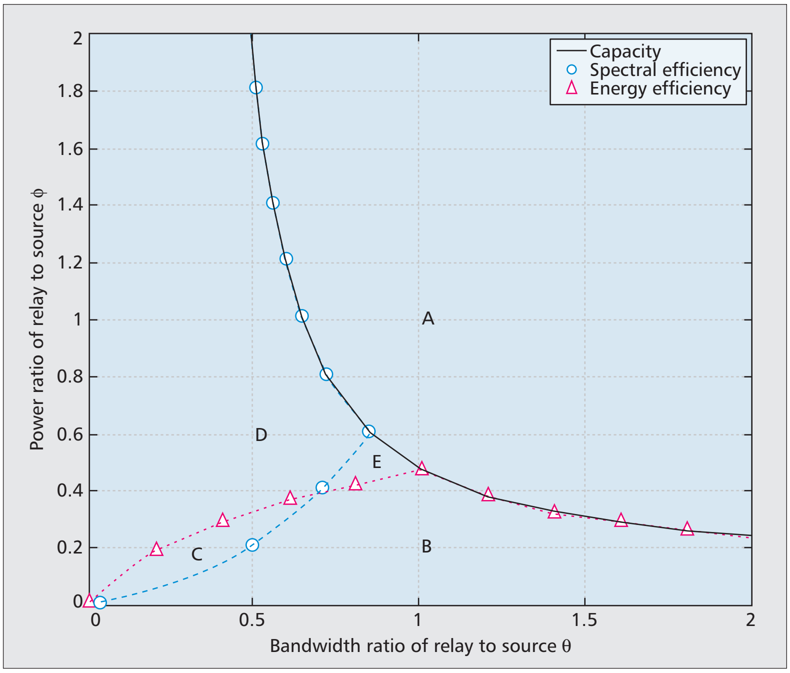

Figure 5. Regimes of bandwidth and power allocation for a link level cognitive relay.

It is envisioned that non-cooperative cognitive cellular networks would contribute to enormous increases in data rate and cell throughput, while cooperative cognitive cellular networks will make similar levels of improvements in spectral and energy efficiency.

licensed RR, how much cognitive RR should be used by the cognitive relay to achieve optimal performance in terms of either capacity, SE, or EE. Following the work in [14], the optimal power-bandwidth trade-off curve with respect to each metric can be calculated. The three curves divide the power-bandwidth plane into five areas, as shown in Fig. 5. These five areas have interesting implications as follows [14]:

- Regime A: Resource excess regime, where power and bandwidth are overprovisioned and cause negative impacts on SE and EE
- Regime B: Power limited regime, where increasing power improves all three metrics, while increasing bandwidth improves capacity and EE but degrades SE
- Regime C: Power and bandwidth limited regime, where increasing either power or bandwidth improves all three metrics
- Regime D: Bandwidth limited regime, where increasing bandwidth improves all three metrics, while increasing power improves capacity and SE but degrades EE
- Regime E: Trade-off regime, where increasing power improves capacity and SE but degrades EE, while increasing bandwidth improves capacity and EE but degrades SE

From the above three examples, we can see that EE-SE trade-off study is a powerful analytical tool that not only provides interesting theoretical insights into the fundamental limits of cognitive cellular networks, but also offers useful guidelines for radio resource management and optimization in practice. A cross-example observation highlights the fact that EE and SE do not follow a classic trade-off relationship under the cooperative architecture, implicating the importance of cognitive radio resource management in cooperative cognitive cellular networks. Although the specific performance targets of 5G systems have not yet been officially released, there is a building consensus that 5G networks will achieve 1000 times the system capacity, 10 times the spectral efficiency, energy efficiency, and data rate, and 25 times the average cell throughput of 4G systems [1]. To this end, it is envisioned that non-cooperative cognitive cellular networks would contribute to enormous increases in data rate and cell throughput, while cooperative cognitive cellular networks will make similar levels of improvements in spectral and energy efficiency.

Summary and Future Research Challenges

In this article, we have introduced the concept of the cognitive cellular network and illustrated its promising applications in 5G mobile communication systems. We have systematically explained the two system architectures, various usage scenarios, three levels of capacity analysis, and two types of capacity metrics that enables a wide range of EE-SE trade-off analysis. Examples have demonstrated that the EE-SE relationship in cognitive cellular networks can be rather different from conventional systems. In particular, the insights gained from EE-SE analysis are shown to be useful in providing guidelines for cognitive radio resource management.

Future research can seek to conduct a thorough investigation on the EE-SE trade-off with respect to various combinations of usage scenarios, levels of analysis, and capacity metrics. For non-cooperative cognitive networks, the system-level EE-SE trade-off deserves further study when practical issues such as BS power control, interference in heterogeneous networks, and multi-user uplink access are taken into account. For cooperative cognitive networks, the link-level and cell-level EE-SE trade-off concerning multiple antennas and multiple users has yet to be fully studied, while the system-level analysis is still by far an open issue. Taking a broader scope, greater research challenges lie in understanding not just the EE-SE trade-off, but how EE and SE are related to the reliability constraint, delay constraint [15], and overhead constraint. Finally, apart from the above theory-oriented studies, practical investigations can be pursued regarding radio resource management and joint scheduling of licensed and cognitive RRs under diverse QoS requirements.

References

[1] C.-X. Wang *et al.*, "Cellular Architecture and Key Technologies for 5G Wireless Communication Networks," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 122–30.

[2] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE JSAC*, vol. 23, no. 2, Feb. 2005, pp. 201–20.

[3] C.-X. Wang *et al.*, "On Capacity of Cognitive Radio Networks with Average Interference Power Constraints," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, Apr. 2009, pp. 1620–25.

[4] F. Heliot, M. A. Imran, and R. Tafazolli, "On the Energy-Efficiency Spectral-Efficiency Trade-Off over the MIMO Rayleigh Fading Channel," *IEEE Trans. Commun.*, vol. 60, no. 5, May 2012, pp. 1345–56.

[5] X. Hong *et al.*, "Capacity of Hybrid Cognitive Radio Networks with Distributed VAAs," *IEEE Trans. Vehic. Tech.*, vol. 59, no. 7, Sept. 2010, pp. 4201–13.

[6] I. Humar *et al.*, "Rethinking Energy-Efficiency Models of Cellular Networks with Embodied Energy," *IEEE Network*, vol. 25, no. 3, Mar. 2011, pp. 40–49.

[7] M. Gastpar, "On Capacity Under Receive and Spatial Spectrum-Sharing Constraints," *IEEE Trans. Info. Theory*, vol. 53, no. 2, Feb. 2007, pp. 471–87.

[8] S. Sridharan and S. Vishwanath, "On the Capacity of A Class of MIMO Cognitive Radios," *IEEE J. Sel. Signal Proc.*, vol. 2, no. 1, Feb. 2008, pp. 103–17.

[9] X. Kang *et al.*, "Optimal Power Allocation for Fading Channels in Cognitive Radio Networks — Ergodic and Outage Capacity," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, Feb. 2009, pp. 940–50.

[10] 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, Feb. 2010, pp. 844–52.

[11] J. G. Andrews, F. Bacccelli, and R. K. Ganti, "A Tractable Approach to Coverage and Rate in Cellular Networks," *IEEE Trans. Commun.*, vol. 59, no. 11, Nov. 2011, pp. 3122–34.

[12] X. Hong *et al.*, "Energy-Spectral Efficiency Trade-Off in Virtual MIMO Cellular Systems," *IEEE JSAC*, vol. 31, no. 10, Oct. 2013, pp. 2128–40.

[13] I. Ku, C.-X. Wang, and J. S. Thompson, "Spectral-Energy Efficiency Trade-Off in Relay-Aided Cellular Networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 10, Oct. 2013, pp. 4970–82.

[14] K. Xue *et al.*, "Performance Analysis and Resource Allocation of Heterogeneous Cognitive Gaussian Relay Channels," *Proc. IEEE GLOBECOM '13*, Atalanta, GA, Dec. 2013.

[15] L. Musavian and S. Aissa, "Effective Capacity of Delay-Constrained Cognitive Radio in Nakagami Fading Channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, Mar. 2010, pp. 1054–62.

BIOGRAPHIES

XUEMIN HONG [S'05, M'12] (xuemin.hong@xmu.edu.cn) received his Ph.D. degree from Heriot-Watt University, United Kingdom, in 2008. Since 2011, he has been an associate professor at Xiamen University, China. His research interests include cognitive radio networks, wireless channel modeling, and wireless ad hoc networking. He has published more than 30 technical papers in major international journals and conferences.

JING WANG (jingw@xmu.edu.cn) received her B.Sc. degree from Henan Normal University, China, in 2012. She is currently a postgraduate student at the School of Information Science and Technology, Xiamen University. Her research interest is cognitive radio networks.

CHENG-XIANG WANG [S'01, M'05, SM'08] (cheng-xiang.wang@hw.ac.uk) received his Ph.D. degree from Aalborg University, Denmark, in 2004. He joined Heriot-Watt University as a lecturer in 2005 and became a professor in August 2011. He is also a Chair Professor at Shandong University, China. His research interests include wireless channel modeling and 5G wireless communication networks. He has served or is serving as an Editor or Guest Editor for 11 international journals, including *IEEE Transactions on Vehicular Technology* (2011–), *IEEE Transactions on Wireless Communications* (2007–2009), and *IEEE Journal on Selected Areas in Communications*. He has edited one book, and published one book chapter and about 200 papers in journals and at conferences.

JIANGHONG SHI (shijh@xmu.edu.cn) received his Ph.D. from Xiamen University, China, in 2002. He is currently a professor in the School of Information Science and Technology, 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.