Abstract—This paper discusses the implications of facilitating higher energy efficiency in Cognitive Radio Networks (CRN) from the perspective of fundamental trade-offs, i.e. what needs to be sacrificed to be energy efficient? These trade-offs are identified as QoS, fairness, PU interference, network architecture, and security, which are also essential network design dimensions. We analyze each dimension considering their interactions with each other and with energy efficiency. Furthermore, future research directions which are related to the integration of CRN with other networking paradigms regarding the energy efficiency are introduced and discussed.

I. INTRODUCTION

Energy efficiency (EE) is now at the interest of cognitive radio (CR) research, as networks become more and more energy-demanding. This demand has been under spotlight due to environmental concerns and rising energy costs. However, ensuring high EE in a CRN setting is formidable due to the difficulty in satisfying the competing demands of different stakeholders such as Primary Users (PU), CRs, and the CR operator. For example, the PUs put strict requirements on interference and channel usage of CRs while CRs expect high quality-of-service (QoS) from the operator, and operator desires low operating and management costs. This multifaceted challenge constitutes the essence of our paper: how to provide EE in CRNs while meeting the expectations of different actors and elements in the system? Taking this question as our motivation, we focus on five fundamental trade-offs which are paramount since they affect all the constituents of CRN design and implementation: QoS, fairness, PU interference, network architecture, and security. Although we elaborate on each trade-off separately, we should note that the relation between these trade-offs is inextricably intertwined. For example, potential solutions such as relaying for balancing EE vs. PU interference trade-off lead us to EE vs. network architecture trade-off, e.g. complexity and deployment cost.

The investigated trade-offs are depicted in Fig. 1 from a cognitive map perspective. Deployment and network level factors can be decomposed into various sub items such as offloading, heterogeneous network architecture, and relaying. Cooperation mechanisms in that domain affect all system-level items in that figure. Security establishes another trade-off which affects cooperation item due to trust mechanisms. It is a system-wide attribute which might override EE concerns due to criticality. PU interference is a factor that is CR-native and critical for CR feasibility. It interacts with QoS factor, which in turn complicates the fairness trade-off. There are also general issues such as learning, complexity, and dynamicity, which emerge when elaborating on EE trade-offs. Therefore, the distinction among trade-offs is not that clear-cut. In this work, we hope to provide an insight on the EE trade-offs and highlight some research endeavors.

II. EE vs. QoS

The explosive growth in the use of real-time applications on mobile devices and proliferation of multimedia traffic have resulted in stricter QoS guarantees in terms of sustainable data-rates, packet drop limits, and delay bounds. CRNs have to address the relevant operational environment under these circumstances. This situation complicates EE concerns since QoS requirements become harder to implement when EE requirements are also applied. The interaction between QoS and EE is depicted in Fig. 2. It is apparent that the QoS
improvement mechanisms may be contradictory to the EE requirements. Moreover, there are also inherent complicating factors such as interference limitations, the power budget of the CR system, and imperfect channel sensing. Hence, this problem configuration is typically reflected in resource allocation and assignment optimization works for CRNs [1].

QoS for CRNs or CR embedded networks have been examined typically from the dynamic spectrum access (DSA) perspective. Disruptions from fundamental operations involved in DSA protocols render deployment of QoS mechanisms challenging. In that regard, QoS can be considered in three directions, the first being the *PU-centric approach* where the primary focus is to protect QoS of PUs while facilitating DSA. For this purpose, probability of misdetection is important. The sensing and medium access mechanisms should be highly conservative and SU-insensitive. Thus, the main constraint is not to disturb while maintaining QoS and EE. The EE dimension is not substantially critical for this case. The second approach is to prioritize SUs without harming PUs to have a *SU-centric* QoS environment. For this setting, interference limitations are relaxed and the problem is more flexible. In this case, the aim is to reduce probability of false alarm as much as possible. For both of these approaches, the specific probability criterion can be reduced by either increasing signal-to-noise (SNR) ratio and/or by increasing sensing time and sampling frequency. As the SNR is beyond the control of the CRN and the sampling frequency is device related, increasing sensing time is the only viable solution. However, this also results in more energy consumption for the network, especially considering the periodic nature of sensing. Another dimension is to differentiate among SUs especially when their QoS requirements cannot be met. It is desirable to have the spectrum access opportunity related to the user priority if they belong to different priority classes [2]. A natural extension of these two former approaches is to have a *hybrid* setting where QoS of PUs and SUs are not differentiated categorically but evaluated in a more flexible manner.

In centralized CRNs, once the list of available opportunities are determined, the Cognitive Base Station (CBS) assigns the spectrum opportunities using one of the above-mentioned approaches. The CBS can exploit various diversities to attain the highest trade-off among EE and QoS goals. Conceptually, these diversity techniques can be categorized into four main groups: link diversity, spatial diversity, channel diversity, and CR diversity [3]. CBS can consider the time-varying channel conditions and CR diversity in assigning the most favorable channels. However, CRs are to operate in a wide range of frequencies, which may be spectrally distant from each other. Overhead of channel switching have to be accounted especially for fragmented spectrum (i.e., spectrum is non-contiguous) as well as each channel’s throughput promise and energy demand. Since a CR spends nonnegligible time in radio reconfiguration for tuning to the new channel, time spent during channel switching reflects as throughput loss. Moreover, CR consumes power in channel reconfiguration, which translates into energy expenditure. Therefore, a channel switching should be performed only if the new channel can provide net gains in EE of this CR [4]. To the best of our knowledge, the literature is not mature enough regarding experimental results from the CRN testbeds on the cost of channel switching.

In multi-hop CRNs, ensuring QoS becomes more challenging due to routing since the paths among network nodes are highly dependent on channel availability [5]. A typical case is cognitive ad hoc networks where a routing algorithm can establish QoS paths with reserved bandwidth on a per flow basis in a multi-hop transmission. For routing, the fundamental QoS mechanism is to establish bandwidth guaranteed routes while considering EE. However, dynamicity of CRNs complicates this class of solutions. For CRs, as the number of hops increases finding a stable/reliable path between the sender and the receiver becomes more of an issue as the channel occupancy may change frequently between hops.

III. EE VS. FAIRNESS

Fairness for a communication system refers to the degree at which users utilize a fair share of the system resources [6]. Since CRs allow SUs to share the spectrum with PUs in a dynamic manner, the fairness between SUs is crucial. Spectrum access and allocation methods should let each user get certain amount of spectrum regardless of its spectral environment, location, or neighbor distribution. The typical trade-off in EE setting is that being unfair in certain settings can be beneficial for EE. The basic underlying requirements are to allocate spectrum as fairly as possible while using spectrum resource maximally and maintaining EE [7].

For fairness on the downlink of centralized systems, EE is usually not mandatory but desired in order to adopt less complex hardware and decrease operational costs. However, for the uplink, this is usually required due to mobility and battery-powered operation of network end devices. The fairness lends itself to a multiobjective optimization problem since it is not usually considered as the sole objective for CRN design and operation. Thus, the fairness trade-off is typically embedded in QoS mechanism implementation and optimization settings. For instance, [4] introduces a *satisfaction ratio* for each SU in order to make the scheduler fairness-aware and incorporates this term as a multiplicative term in the resource allocation optimization problem. However, the trade-off between EE and fairness is yet to be explored adequately. The infrastructure sharing concept is also interesting to explore this trade-off. That case introduces another layer of complexity where a fair
allocation is desired among different CRNs while satisfying the inherent “PU-biased unfairness”.

IV. EE vs. PU INTERFERENCE

The fundamental restriction on CR operation is that CRs must not harm the PU communications. In other words, the resulting interference due to CR transmission at the nearby PUs must be below the tolerable interference limits for underlay CRNs, and the simultaneous transmission time with the PUs must be considerably short for overlay CRNs. The interference arises under two cases: PU misdetection and PU reappearance. To cope with the first case, CRs must sense with high detection accuracy, i.e. \( P_d \), so that they rarely transmit with a PU simultaneously at the same channel. This calls for high \( P_d \), which might be achieved by various techniques: cooperative sensing, longer sensing duration, and higher sampling frequency to name a few. On the other hand, these high \( P_d \) promising solutions may be costly in terms of energy consumption compared with a solution demanding less reliability (lower \( P_d \) which is yet higher than minimum reliability \( P_d^{\text{min}} \) required by the PU regulations). Furthermore, for the second case no matter how high the achieved \( P_d \) is, CRs may still result in PU interference due to the nature of periodic sensing.

CRNs typically operate on a frame by frame basis in which certain duration of the frame is dedicated for sensing and the rest for transmission. The duration between two consecutive sensing periods determines the performance of spectrum opportunity discovery (thus the throughput) and resulting PU interference. In periodic sensing, a CR does not notice a reappearing PU until the next sensing period. Frequent sensing results in increased energy consumption while improving the sensing performance, which directly affects throughput. Hence, tuning the sensing and transmission durations as well as the period [8] is of major concern for playing with the EE vs. PU interference trade-off.

To account for these two cases, a CR may select to be conservative at the sensing step and/or at the transmission step of the cognitive cycle. Solutions at the sensing step include period adaptation (considering PU traffic pattern [8]) and playing with the sensing accuracy. Alternatively, a CR can meet \( P_d^{\text{min}} \) restriction and can control the interference via regulating its transmission power \( (P_{tx}) \). Given that perceived interference at the victim node is a function of \( P_{tx} \), a CR can decrease the PU interference by decreasing \( P_{tx} \). However, as the Shannon’s formula shows, channel capacity also (logarithmically) decreases with \( P_{tx} \). Consequently, both methods result in a trade-off between EE and PU interference. Avoiding PU collision is not only essential to protect the PUs, but also to avoid any retransmissions of CR traffic and to achieve the maximum capacity of the channel. CR transmitter’s traffic colliding with a PU may not be decodable at the CR receiver, which in turn may require retransmission(s) due to the QoS requirements. Hence, from the EE viewpoint, simultaneous transmission with the PU must be kept at minimum.

In order to change the EE-PU interference trade-off in favour of EE, CRs can benefit from relaying [9] and channel aggregation [3]. As depicted in Fig. 3, relaying lets the CR transmit with lower power but via multiple hops while channel aggregation facilitates CRs to transmit simultaneously via multiple channels. In case some intermediate nodes relay the CR traffic, capacity improvement due to shorter transmission distance may compensate the channel capacity loss due to the lower \( P_{tx} \). Similarly, channel capacity loss due to lower \( P_{tx} \) is compensated by higher bandwidth of the aggregated channels. Regarding the cost, relaying may require a change in the network architecture if relays are supposed to be dedicated devices for assisting CRs. The dedicated relay devices evidently add to the energy consumption of the network. We should recall that providing EE adhering to the PU interference restrictions requires us to tune EE vs. network architecture trade-off. Alternatively, each CR may serve as a relay for the others at the expense of increased energy consumption for relaying. On the other hand, channel aggregation demands for more capable hardware at the CRs. Both schemes lead to new challenges that deserve further analysis, such as how to select a relay, how to place the relays, and how to allocate power at each channel for the optimal EE.

V. EE vs. NETWORK ARCHITECTURE

Different types of network architecture are possible for achieving higher EE in CRNs. Almost all of these different architectures try to benefit from adding additional hops or infrastructure layer between the CR and the core network in order to decrease the required transmission energy of CR by decreasing the transmission distance. These architectures can be listed as small cells, relays, ad hoc networks, and clustering.

The goal of deploying cognitive small cells is to offload user traffic from the CBS to small cell access point (SAP), be it a femtocell or a microcell, etc. Interpreting the usage statistics that majority of traffic originates from indoors, small cells deployed either by users or operators can provide high capacity at small localities, e.g. home for femtocells or shopping malls/airports for pico/micro cells. Small cells benefit from spatial diversity to achieve better frequency reuse that leads
to a higher spectral/throughput efficiency. When cooperative sensing is employed, the energy burden for spectrum sensing may be on the CRs served by the SAP. Furthermore, the number of handoffs a CR performs during operation increases drastically, especially for high mobility cases. We should also mention that the handoff procedure is more complicated and more energy consuming in a heterogeneous CRN compared to a classical CRN architecture. However, cognitive small cells can cope with the interference issue arising with the unplanned deployment of small cells to some extent by utilizing the unused PU spectrum opportunities.

Another alternative is to use relays together with amplify-and-forward or decode-and-forward type of cooperative communications to save transmission energy by both decreasing the distance and retransmissions. If CRs are used for relaying packets (which may not be the case as they may not be willing to consume their energy for other users), they will have an additional energy consumption. Moreover, some CRs may become bottlenecks due to their location (it may be the only alternative for relaying) and their battery may drain rapidly. Even if dedicated relays are used, it has been shown that if the traffic load is low or channel conditions are good or the transmitter is close to the receiver, relaying may not be as energy-efficient as it seems [10]. Hence, it would be better to decide whether to relay or not on a case by case basis. In a highly dynamic environment (i.e. channel conditions) or with highly mobile CRs, this decision induces extra overhead to the network. Another problem with relaying is that the time it takes for successful transmission is doubled. This may not be feasible for some applications. The delay also increases if relaying is performed in a multi-hop manner.

Both of the discussed approaches bring an additional layer to the system. The monetary and energy cost of operating additional hardware, like relays or SAPs, is usually neglected. However, it is known that idling (waiting idle for possible packet reception) consumes almost as much energy as reception [11]. Therefore, clever mechanisms are needed for reducing operating costs such as sleep scheduling, which induces additional overhead together with decreased throughput. This should then be considered from an EE vs. QoS trade-off viewpoint. Moreover, the locations and the number of these additional network components should be selected carefully to be effective. In a dynamic cellular network, the solutions for these problems are not trivial.

Another concern about the choice of a specific network architecture is to decide whether to utilize internal sensing or external sensing (i.e., spectrum sensing vs. geolocation databases). For the supporters of the latter, sensing and intelligence can be located at the network (i.e., Radio Environment Maps, a.k.a. REM) instead of individual CRs. Putting the discussion aside that this approach contradicts with the essence of the CR, REMs have to be deployed at various scales (e.g., country-wide or campus-coverage). Therefore, each device contributes to the energy consumption required for processing, cooling, and synchronization. Besides, it is not considered to be green to deploy such machines everywhere. On the other hand, REM can ease the learning process by processing the gathered data by sensors (e.g., CRs or other external entities) and deriving the characteristics of the radio environment from a more complete data. Thus, a CR can improve its environment awareness instead of using its less complete perception of the operating environment. Hence, we have to consider the trade-off between deployment and operating cost introduced by these entities and the performance improvement (e.g., higher EE) by a more enhanced learning scheme.

Regarding learning and CR’s intelligence, we should consider environment monitoring along with the dynamism of the RF/CR environment. In other words, tracking and collecting data on the network is beneficial only if the conclusion about the network state via the inference of the collected data remains valid at the time of the inference. If the cognition requires long-term information keeping and computationally long time compared with the dynamism of the CR or PU network, CR decision using this obsolete inference may even hurt the CRN performance. Consequently, CRNs need computationally efficient learning algorithms that can keep a faster pace with the changing nature of the agents in the operating environment. REMs deployed at small localities and cooperative learning can be explored further for their capabilities of faster cognition.

Fig. 4 presents a rough comparison of various network architectures in terms of PU interference, network architecture complexity, and CR throughput. This analysis may change based on specific equipment, communication protocols and technology. All issues related to traditional ad hoc networks and clustering apply to the cognitive counterparts with the challenge of establishing a reliable common control channel. These architectures are simple but their uncontrolled/distributed operation may degrade CR performance and can have difficulty in efficiently managing PU interference. Assuming that REMs implement efficient learning mechanisms and store up-to-date information, they provide high throughput at the expense of high complexity. Plain cellular network with high power transmitters creates more PU interference whereas small cells achieve high throughput and low interference owing to offloading and close proximity to SAP.
VI. EE vs. Security

Equipping CRs with security protocols requires taking some precautions, which result in additional processing both at the transmitter and the receiver. In secure environments, these precautions may seem to decrease the EE of the network as each entity spends processing power/time and some of the channel capacity for transmitting these authentication and integrity messages. On the other hand, in problematic environments with malicious or misbehaving nodes, security mechanisms, although resulting in overhead, may improve EE by avoiding interactions with malicious users and appropriately detecting the misbehaving nodes. For example, a CR with its security protocols may detect PU emulation attacks and can use the idle spectrum that would be wasted without an attack detection mechanism. Hence, effect of security precautions on CRN may be intricate if assessed from an EE viewpoint.

Security attacks can be generated either by an insider as in spectrum sensing data falsification (known as SSDF attack) in cooperative spectrum sensing, or by an external entity as in PU emulation attacks. In the latter, the attacker emulates the PU signal to block CR transmission and instead transmits itself at the idle spectrum band. All widely-recognized attacks aim to collapse the CRN’s sensing capability. These attacks make the CRN fail at the very early stage of cognitive cycle (i.e., sensing) because of the shortage of transmission opportunities. Malicious users tend to report the existence of PU. Althunibat et. al. determine the optimal number of security bits in a message for attaining the highest trade-off between attained security level and EE for a CRN subject to SSDF attacks [12]. Optimal number of security bits depends on the fusion rule at the fusion center, number of SSDF attackers, and number of legitimate users.

To cope with security threats while not trading off the EE, CRNs can define cooperative protocols that encourage cooperation among trusted CRs and keep track of the trustworthiness of each other. We discuss this issue further in Section VII from a social network perspective.

VII. Future Research Directions

The future directions for CR EE can be broadly divided into two groups:

- **CR-native – endogenous** such as more energy-efficient sensing schemes, learning frameworks or sensory data gathering,
- **Integration with other networking paradigms – exogenous** such as social networks, user behavior, and energy harvesting.

In this work, we focus on the latter which provides new degrees of freedom and opens new directions for CR research.

A. Social Network Analysis (SNA)

A social network views a network as a group of nodes with their interrelations (e.g., physical distance, contact frequencies) to benefit from these structural and social ties for higher efficiency. Taking this definition, a CRN is unquestionably a social network in which CRs may have various ties with others depending on their spatial and social properties. Hence, uncovering the interconnections among CRs and designing protocols accordingly can substantially improve the CRN performance. Fig. 5 illustrates a social network of CRs in which nodes have different characteristics and diverse view of the same network according to their social ties. Social graphs are key to keeping track of both interactions and social ties among users. The former is beneficial for estimating the structure of the network such as connectivity and proximity of two users while the latter may inform us about the trust among nodes and influential nodes in the network. This information can be exploited in designing cooperative communications with higher EE, e.g., cooperative sensing and cooperative learning.

Previous works on cooperative sensing largely overlook the burden of cooperation to the individual CRs (especially in terms of energy consumption) and implicitly assume that each CR is cooperative. However, such a cooperative behaviour may not be applicable in practical CRNs. For instance, a user that does not have any active communication, runs out of battery only because it receives sensing requests frequently from the other CRs and consumes much of its battery on sensing. Instead, we envision a more realistic operation scheme in which CRs’ cooperation willingness depends on the social ties among the users of these CR devices. Initial works show that CRs benefit from such a social-aware cooperative sensing scheme in which each CR selects a cooperation set based on its friendship ties as well as the historical sensing performance of the cooperating nodes [13], and CRs can benefit from other CR’s recommendations on PU channels [14]. Thus, CRs are expected to spend less effort for sensing and learning which in turn lead to higher EE.

Environmental awareness improves the CR performance via letting the CR proactively take the best action at the expense of increased energy consumption due to the constant environment monitoring. Instead, CRs can share their experiences, which reduces this burden. However, this solution raises the following question: To which extent a CR can trust to the other CRs’ reports and what if the recommendations are inaccurate? Although trust can be rather sophisticated in different contexts, we can model trust among CRs based on their social ties (e.g. friendship and community) and can dynamically adapt the trust between CRs based on their interactions and feedbacks. For example, CRs with similar profiles may have high trust towards each other as in human societies, but may decrease if
for this kind of behavior, the battery life is still reduced as it is charged for a longer period of time than required and it is recharged before being fully discharged. Furthermore, the circuitry still consumes a small amount of energy in the time frame between the battery is fully charged and the user plugs the device out in the morning. Since the number of users is on the order of billions, even that small amount sums up to a huge consumption. With user behavior learning, the device can make the necessary adjustments to save energy.

C. Energy Harvesting

Energy harvesting or scavenging is basically the process by which the energy is extracted from external ambient sources such as RF environment, thermal variations, or kinetic energy for improving EE or enabling energy-source free operation. It requires two main functionalities for being practical in wireless systems: energy generation and storage. However, the bursty nature of wireless traffic results in large spatio-temporal variances in system load. Additionally, the inherent randomness in the energy harvesting and thus energy flow prediction leads to the problem of consumption-generation matching and storage planning.

Environment-awareness is a key enabler for optimization of energy harvesting functionality. Considering the fact that CRs are expected to operate in a manner that is aware of their environment, they lend themselves to energy harvesting paradigm with their intrinsic capabilities. For instance, Park et al. explore how a CR with energy harvesting capability can adapt in both spectrum-limited and energy-limited regimes [15]. Moreover, the adaptation and learning capabilities of CRs can be augmented with energy harvesting. The assumed capabilities of advanced sense-decide-act-learn cycle for CRs require fundamental changes in RF, baseband, and power management in wireless devices. These enhancements can also be utilized for energy harvesting. At the single node level, CRs may schedule their delay-tolerant traffic lazily according to the location and mobility pattern of the user in order to utilize upcoming energy harvesting opportunities in an efficient manner. They can also be more aggressive in case that they predict a sluggish user behavior and an increasing energy harvesting potential in time and location dimensions. At the network level, CRs are envisioned to be self-aware agents communicating and cooperating with each other. In that regard, cooperation for network-level EE may rely on the altruistic load redistribution to make energy harvesting sufficient for the energy-source operation of underprivileged network nodes.

B. User Behavior

The most important aspect of any type of communication network is the user. CR is a learning entity that senses and decides based on the environment it operates. However, the interaction between the user and the device is usually ignored. We discuss a couple of examples from state of the art cellular devices about how the device-user interaction can save energy, but the arguments also apply to CR devices.

Almost all of the modern cellular devices or smart phones come together with Bluetooth and Wi-Fi units in addition to 3G/4G. Moreover, these devices have both of these circuits on in their factory setting. On the other hand, an average user scarcely uses these protocols, especially Bluetooth. The critical point is that the user does not care about the energy consumption of these circuits unless he/she has low battery. In addition, some users even do not know how to turn them off to save energy. Thus, both protocols periodically seek some pairing/association all the time. A CR device can learn and analyze the user behavior such that when and/or where the user employs these types of additional communication units, and turn them off when it predicts that they are not needed.

Another example is what we call the “50% problem”. An average user usually plugs his/her device at night just before going to bed for charging in order to start the next day with full battery. The battery charging process is shown in Fig. 6. Although modern devices have developed circuitry...
accordingly. Moreover, the emerging CR capabilities can be utilized in an energy generation perspective where efficiency can be augmented with energy harvesting. We believe that these research directions can enable new solutions that will facilitate higher EE with a good strike of the listed trade-offs.

REFERENCES