IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received May 21, 2018, accepted June 15, 2018, date of current version July 12, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2849446

Cognition-Inspired 5G Cellular Networks: A Review and the Road Ahead

KOK-LIM ALVIN YAU[1], (Member, IEEE), **JUNAID QADIR**[2], (Senior Member, IEEE), **CELIMUGE WU**[3], (Member, IEEE), **MUHAMMAD ALI IMRAN**[4], (Senior Member, IEEE), **AND MEE HONG LING**[1]

[1]Department of Computing and Information Systems, Sunway University, Selangor 47500, Malaysia
[2]Information Technology University, Lahore 54000, Pakistan
[3]Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan
[4]School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K.

Corresponding author: Kok-Lim Alvin Yau (koklimy@sunway.edu.my)

ABSTRACT Despite the evolution of cellular networks, *spectrum scarcity* and *the lack of intelligent and autonomous capabilities* remain a cause for concern. These problems have resulted in low network capacity, high signaling overhead, inefficient data forwarding, and low scalability, which are expected to persist as the stumbling blocks to deploy, support and scale next-generation applications, including smart city and virtual reality. Fifth-generation (5G) cellular networking, along with its salient operational characteristics—including the cognitive and cooperative capabilities, network virtualization, and traffic offload—can address these limitations to cater to future scenarios characterized by highly heterogeneous, ultra-dense, and highly variable environments. Cognitive radio (CR) and cognition cycle (CC) are key enabling technologies for 5G. CR enables nodes to explore and use underutilized licensed channels; while CC has been embedded in CR nodes to learn new knowledge and adapt to network dynamics. CR and CC have brought advantages to a cognition-inspired 5G cellular network, including addressing the spectrum scarcity problem, promoting interoperation among heterogeneous entities, and providing intelligence and autonomous capabilities to support 5G core operations, such as smart beamforming. In this paper, we present the attributes of 5G and existing state of the art focusing on how CR and CC have been adopted in 5G to provide spectral efficiency, energy efficiency, improved quality of service and experience, and cost efficiency. This main contribution of this paper is to complement recent work by focusing on the networking aspect of CR and CC applied to 5G due to the urgent need to investigate, as well as to further enhance, CR and CC as core mechanisms to support 5G. This paper is aspired to establish a foundation and to spark new research interest in this topic. Open research opportunities and platform implementation are also presented to stimulate new research initiatives in this exciting area.

INDEX TERMS Cognitive radio, cognition cycle, software-defined network, 5G.

I. INTRODUCTION

The traditional cellular network is plagued with two main problems: *spectrum scarcity*, and *the lack of intelligent and autonomous capabilities*. These problems if unresolved threaten to hinder the widespread deployment of next-generation applications (such as IP multimedia subsystem (IMS) multimedia services, vehicular networks, public safety and social networks, Internet of things (IoTs), smart city, virtual reality and augmented reality) that will characterize the next-generation cellular networks, namely fifth-generation (5G) networks.

Spectrum scarcity is attributed to the traditional fixed spectrum assignment policy that allocates a large part of the spectrum to licensed users (or primary users, PUs), leaving a smaller part to unlicensed users (or secondary users, SUs). Generally speaking, the licensed channels are largely underutilized, although some of them can be highly utilized, particularly the cellular channels—such as the channels used for global system for mobile communication (GSM) and wide band code division multiple access (WCDMA)—which are the licensed channels allocated to the cellular networks [1]. On the other hand, the unlicensed channels, such as channels in the 2.4 GHz industrial, scientific, and medical (ISM) radio bands, 5 GHz unlicensed national information infrastructure (U-NII) radio bands, and 60 GHz millimeter-wave (mmWave), are highly utilized. Despite being allocated with licensed channels that can be potentially extended to unused frequency bands, particularly those below 3.5 GHz, the fixed and limited channel capacity allocated for the traditional cellular networks is still insufficient for the next-generation cellular networks. Meanwhile, the lack of intelligent and autonomous capabilities causes the networks and nodes to follow a set of static rules or manual instructions that may require a lot of human interaction, hindering the networks and nodes from achieving self-configuration, self-optimization, and self-resilience.

Cognitive radio (CR) has been proposed to enable SUs to explore and use underutilized licensed channels (or white spaces) owned by the PUs. Cognition cycle (CC), which is an intrinsic part of CR, represents the intelligent and autonomous capabilities that enable a node or network to learn new knowledge and adapt to the dynamicity of network condition and application. In addition to improving the efficiency of spectrum utilization (i.e., addressing spectrum scarcity), a CR reduces the interference to PUs' activities (i.e., via intelligent and autonomous capabilities). In the next-generation cellular networks, a cellular network user can operate as a PU that utilizes its licensed channels (or cellular channels), or can operate as a SU that explores and uses the white spaces in other licensed channels (or cognitive channels) and unlicensed channels.

There are three main advantages of a cognition-inspired 5G cellular network. *Firstly*, with the transition from analogue to digital television broadcasting, multiple television (TV) stations can transmit in a single TV channel via multiplexing, which leaves some TV white spaces (TVWS) available for CR. Naturally, CR with CC can be used to explore and use the white spaces. *Secondly*, the coexistence of heterogeneous entities—e.g., operating channels with different transmission ranges, and radio access technologies (RATs) such as IEEE 802.11, third-generation (3G) and 4G—is an intrinsic characteristic of 5G. CR with CC allows the heterogeneous entities to explore and use each other's resources, promoting interoperation. *Thirdly*, the use of CC introduces intelligent and autonomous capabilities to 5G, further enhancing the 5G core operations—e.g., smart beamforming supported by massive multiple-input and multiple-output (MIMO) and high frequency mmWave communication—and the overall 5G efficiency.

This article presents CR and CC as enabling technologies for 5G, particularly for mmWave communication due to: a) the need for dynamic channel access, b) the need for directive narrow-beam links, and c) the need for learning and predicting the regular mobility patterns and behaviors of nodes for beam tracking due to the mobility of either transmitter or receiver. 5G is aspired to provide substantial network performance enhancement, including high data rate (i.e., up to 50 Gbps and 5 Gbps for fixed and mobile networks, respectively [2]), low round-trip time (i.e., less than 1 ms for mission critical tactile applications) [3], and low energy consumption (i.e., 10 to 1000 times lower). Our focus is the networking aspect of CR and CC applied to 5G. The rest of this section presents a background of CR and CC, our contributions, and the organization of this paper.

A. COGNITIVE RADIO

Using CR, the cellular networks must reduce interference to the PUs' activities in the cognitive channels, which are the licensed channels with white spaces, such as TVWS, and the existing systems' activities in the unlicensed channels, such as the unlicensed WiFi networks. There are two main CR operations, namely *channel sensing* and *dynamic channel access*. Channel sensing enables a cellular network to sense for the real-time occupancy of PUs' and existing systems' activities in the cognitive and unlicensed channels, respectively, in order to identify white spaces. Dynamic channel access enables a cellular network to select and use white spaces in the cognitive and unlicensed channels dynamically via two approaches, namely *opportunistic channel access* and *channel trading or leasing*. Opportunistic channel access, particularly long-term evolution in unlicensed spectrum (LTE-U) (or license-assisted access, LAA) [4]–[6], allows nodes—referred to as user equipment and devices for simplicity henceforth—to select and share white spaces in the cognitive and unlicensed channels in an opportunistic manner with the condition that the nodes must achieve a fair coexistence among different networks, and shift their operations to other cognitive and unlicensed channels when PUs' or existing systems' activities reappear in the operating cognitive and unlicensed channels. Channels with higher amount of white spaces and higher signal-to-interference-plus-noise ratio (SINR) [7] are favorably selected for operation. In addition, listen-before-talk (LBT) (or clear channel assessment, CCA), which has been applied in traditional wireless networks (e.g., IEEE 802.11), has also been adopted in 5G for an enhanced opportunistic channel access. Channel trading or leasing allows a node to negotiate with PUs and existing systems, potentially via a spectrum broker, to purchase white spaces in terms of time duration in exchange of monetary gain or technical contribution for the PUs and existing systems. For instance, a node must help to relay a PUs' packets as a form of payback in cooperative communication. Apart from the two main CR operations, the CC embedded in CRs can be used to observe and analyze the operating environment, learn, and subsequently plan and select actions that meet the policy constraints, network objectives and goals to improve network performance over time. For instance, using CC, cognitive channels can be selected in a predictive and proactive manner rather than reacting to the current and instantaneous network dynamics.

B. COGNITION CYCLE

CC is a concept for learning the optimal actions from past experiences as time goes by. Artificial intelligence

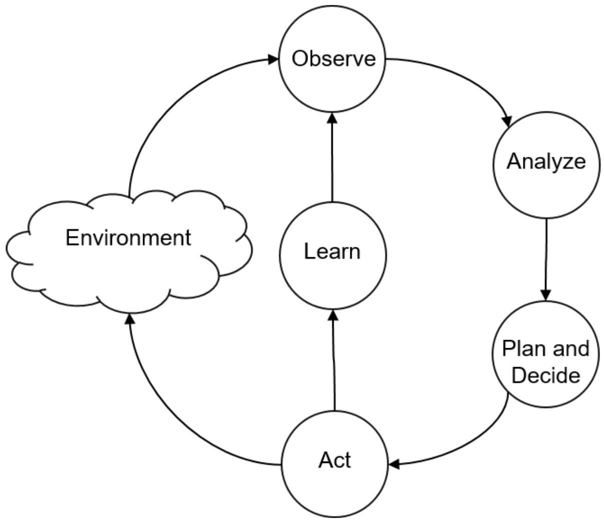

FIGURE 1. A cognition cycle adapted from [11], [12].

approaches [8]–[10], such as reinforcement learning, swarm intelligence, multilayer perceptron, and genetic algorithm, have been widely applied to implement CC in CR networks, although the investigation of CC in the context of 5G has been limited in the literature. There are five main steps in a CC as shown in Figure 1 [11], [12]. *Firstly*, in the *observe* stage, data (e.g., users' preference and experience, and network performance) are collected from the network and applications (e.g., social network), or observations (e.g., channel availability, radio signals, traffic behaviors, traffic load, and node mobility) are gathered from the operating environment through sensing. The data are stored in a common data repository. *Secondly*, in the *analyze* stage, the data are analyzed to discover and identify data patterns and system behavior. *Thirdly*, in the *plan and decide* stage, various alternatives or options are considered and an action (e.g., reconfiguration of an operating parameter) is selected. *Fourthly*, in the *act* stage, an action is executed in the operating environment. *Fifthly*, in the *learn* stage, new knowledge is learned by monitoring the consequences of the actions.

The CC concept has a close resemblance to the concept of self-organization [13], [14]. In general, self-organization enables nodes to maintain their objectives in the presence of network dynamics without the presence of a centralized controller (or a leader), as seen in a flock of cranes maintaining a delta formation during a group flight [13]. While the self-organization concept has been investigated in a wide range of context including ad hoc networks, wireless sensor networks, and cellular networks [13], the CC concept has been limited to CR, and even more so to the dynamic channel access application. Having said that, the three characteristics that define self-organization can be observed in CC as well: *scalability*, *stability*, and *agility*. Scalability maintains complexity despite increasing number of nodes in distributed networks, stability traverses a finite number of states within an acceptable finite time period without oscillation, and agility adapts to the changes in the operating environment acutely [13]. While the focus of self-organization has been largely on distributed networks, the intelligent and autonomous capabilities, in the context of 5G, must appear in both centralized entities, particularly the core network (or cloud) (see Figure 2), and distributed entities, such as the base stations of the macrocells and small cells, as well as the nodes. Without leveraging the CC concept from CR, there would be a large void, and so this article serves the purpose of understanding the achievement of intelligent and autonomous capabilities from the CR perspective.

Traditionally, the base station of a cellular network serves as a central controller managing its nodes' operations. Incorporating CC into 5G helps to solve complex networking problems via learning with minimal human interaction. As an example, a virtual resource allocation scheme equipped with CC can adapt to a diverse range of network conditions and real-time application requirements, particularly quality of service (QoS) and quality of experience (QoE), and subsequently select the best possible actions to maximize network performance. Examples of network conditions are intercell interference and congestion levels. QoS represents network-oriented performance such as throughput, end-to-end delay, jitter and packet loss rate, while QoE represents user-oriented performance such as the subjective user experience and user satisfaction of the quality and timeliness of a service. The virtual resource allocation scheme can be embedded at a cloud, a base station or a node, so that resource allocation can be performed both at the global and the local levels.

C. OUR CONTRIBUTIONS

General review of 5G [15], [16], as well as reviews on specific topics including energy-efficient techniques [17], network slicing [18] and user association [19], and dynamic spectrum access [20], [21], along with cloud support [22], have been presented in the literature. In addition, there have been investigations on various aspects of 5G, including network security [23], [24], information theory [25], and physical layer covering transmitter and receiver designs [26], [27], antenna designs (e.g., MIMO and beamforming [28], [29]), software-defined network frameworks [30], [31], network coding [32], transmission power control [33], modulation [34], and synchronization [35]. This article complements the highlighted works. Our contribution is to provide a review on the networking aspect of CR and CC applied to 5G, with a focus on recent work and open research opportunities. In addition, to facilitate the understanding of this topic, important terms and concepts, as well as the attributes of 5G, are presented. To the best of our knowledge, this article serves as the first review on this topic, and it is timely due to the urgent need to investigate, as well as to further enhance, CR and CC as core mechanisms to support 5G. This article is aspired to establish a foundation and to spark new research interest in this topic.

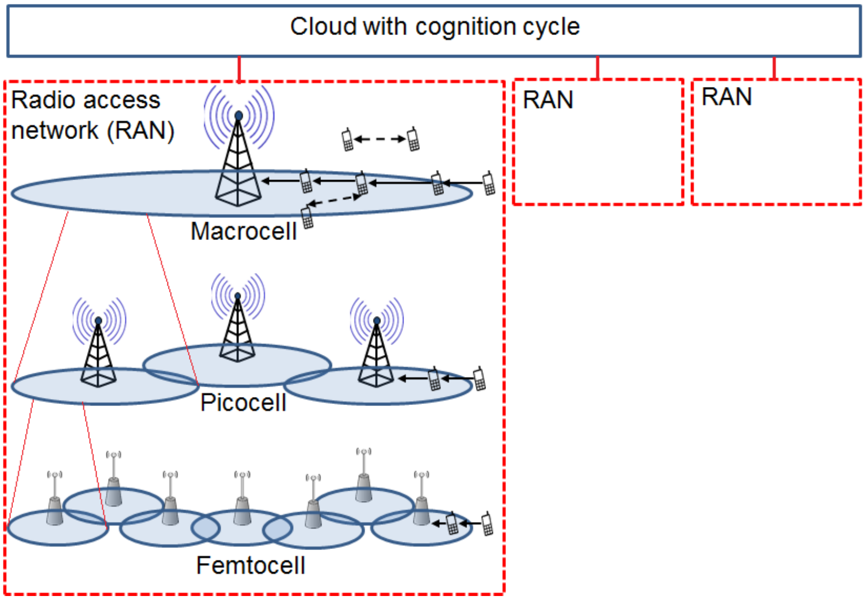

FIGURE 2. A 5G architecture in which a radio access network (RAN) consists of macrocells, picocells and femtocells that overlap among themselves. The architecture can be separated into control and data planes. Indicated in the figure is a femtocell with the smallest coverage that can be completely overlapped by a picocell, and a picocell with a larger coverage that can be completely overlapped by a macrocell. Each RAN has a connection to the cloud. This architecture has been proposed in the literature, in which cloud with CC is a new addition to 5G. Solid line with arrow represents communication with base station, while dotted line with arrow represents communication between two nodes. For simplicity, existing systems, such as the unlicensed WiFi networks, are not shown in the Figure.

D. ORGANIZATION OF THIS PAPER

The rest of this article is organized as follows. Section II presents a proposed architecture. Important terms (e.g., core networks, network cells, and so on) and concepts (e.g., control and data planes, and channel access in 5G) are presented. Subsequently, this section presents and explains the attributes of 5G (e.g., limitations of traditional cellular networks, characteristics of next-generation cellular network scenarios, characteristics of 5G operations, and improvement achieved by 5G). Section III presents a review of some recent work on CR and CC as core mechanisms enabling 5G from the networking perspective. Examples of the recent work are the application of CC to 5G applications, channel sensing, spectrum sharing and intercell interference coordination, D2D communication, and so on. Section IV presents possible future work directions. Finally, Section V presents conclusion.

II. PROPOSED ARCHITECTURE AND ATTRIBUTES OF 5G

This section presents a proposed architecture and the attributes of 5G.

A. PROPOSED COGNITION ENHANCED 5G ARCHITECTURE

5G is a holistic multi-tier cellular network evolved from existing cellular networks. Figure 2 shows a 5G architecture proposed in the literature [2], [36]–[39]. In addition to macrocells that provide larger coverage but lower data rate and small cells (i.e., picocells and femtocells) that provide higher data rate but smaller coverage, the 5G achitecture consists of a core network comprising clouds with CCs that provides consistent and efficient centralized (or global) management. The cloud with CC is a new addition that allows dynamic and real-time reconfiguration of the operating parameters in 5G. For simplicity, existing systems, such as the unlicensed WiFi networks, are not shown in the Figure. The rest of this subsection explains this 5G architecture.

1) CONTROL PLANE AND DATA PLANE

Traditionally, both network control functions (e.g., routing and scheduling) and data packet forwarding are incorporated into a single specialized hardware unit (e.g., routers and switches). In 5G, software-defined network (SDN) technology provides programmability that allows dynamic and real-time reconfiguration of the operating parameters (e.g., operating channel) of the underlying radio frequency (RF) front end. This transfers network control functions (e.g., routing and scheduling) from traditional specialized hardware units (e.g., routers and switches) to software-based controllers running on standard hardware units. The standard hardware units can be the base station of a macrocell for a macrocell-level control, as well as servers and data centers in the core network for a global-level control. Naturally, this segregates the network into a: a) *control plane* that provides centralized management and control at the macrocell- or global-level, and b) *data plane* that provides distributed management and control.

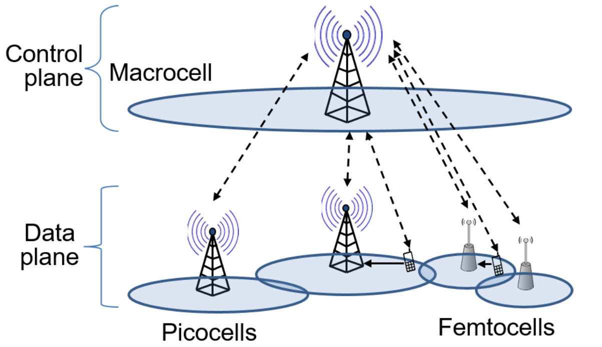

FIGURE 3. A control/ data separation architecture [42] that offloads control messages from picocells and femtocells to a macrocell in order to provide higher channel capacity for data transmission. Dotted line with arrow represents control message exchange, while solid line with arrow represents data transmission.

Specifically, the *control plane* possesses centralized information (e.g., channel availability and network topology), manages a pool of resources and services available, and performs centralized, complex yet delay tolerant functions; while the *data plane* possesses local information, manages local resources and services in the neighborhood, and performs distributed, simple yet delay sensitive functions [40], [41]. Three examples are presented. *Firstly*, as shown in a control/ data separation architecture [42] (see Figure 3), the control plane provides control message exchange between a macrocell and picocells/ femtocells, while the data plane provides data transmission between picocells/ femtocells and nodes. By offloading control messages from picocells/ femtocells to macrocells, higher channel capacity is available to provide higher data rate. *Secondly*, the control plane performs routing based on centralized information (e.g., network topology) at the macrocell or global level, and updates the flow tables (or forwarding tables) of the routers and switches in the data plane. The flow tables provide the optimal links and operating channels to a next-hop router and switch. On the other hand, the data plane forwards packets following the flow tables. *Thirdly*, the control plane manages operational policy (e.g., constaints on intercell interference and operating channels) at the macrocell or global level, while the routers and switches in the data plane make instantaneous decisions (e.g., operating parameters including operating channel and transmission power) at the local level. The control plane and data plane communicate among themselves (e.g., data plane reports local network condition to the control plane) using network protocols such as OpenFlow [43]. By updating the software in the control plane, new network protocols, functions, features and policies can be flexibly adopted and incorporated into the network, providing updates to a large number of routers and switches.

2) CORE NETWORKS, NETWORK CELLS AND NODES

The core network (called evolved packet core, EPC in 4G long term evolution (LTE) networks) incorporates a centralized software-based controller that manages and gathers information about clouds, and makes global decisions. Each cloud pools together a certain type of network information, resources and services for sharing among networks and nodes. Examples of network information are operating parameters, traffic behaviors, congestion level, intercell interference level and network performance. Examples of resources are computing and memory. Examples of services are QoS control (e.g., medium access control, scheduling, resource allocation, admission control and route selection [44]), mobility management (e.g., connection setup and handover), intercell interference coordination (ICIC), and those offered by applications (e.g., file transfer and video conferencing). The clouds can be formed in an on-demand manner, and each of them is recognized by the types of resources and services it offers, and its characteristics (e.g., link quality and residual energy). Each network or node can join more than a single cloud to share and use different kinds of resources and services.

In general, lower frequency bands (e.g., less than 2 GHz), which have better propagation characteristics, provide a larger coverage but a lower data rate; while higher frequency bands (e.g., 60 GHz or mmWave) provide a higher data rate but a smaller coverage. Naturally, the macrocells transmit with higher transmission power in lower frequency bands, while the small cells (i.e., picocells and femtocells [45]) transmit with lower transmission power in higher frequency bands. Nevertheless, both macrocells and small cells can use any of the frequency bands [45]. For instance, lower frequency bands can be used in an operating environment that requires better propagation characteristics, such as stairwells and basements in indoor environment. The transmission in higher frequency bands can be highly directional, contributing to higher beamforming gains and lower interference levels, hence improving the signal quality. However, this requires more complex, dynamic and CC-aware mobility management, particularly learning and predicting users for beam tracking.

Cellular, cognitive and unlicensed channels can be used, and the use of cognitive and unlicensed channels in 5G distinguishes itself from the traditional cellular networks that use cellular channels only. The base stations of the macrocells and small cells may support multiple RATs, and may communicate among themselves using optical backhaul. Some important functions of macrocells and small cells are to: a) control transmission power to reduce intercell interference among neighboring or overlapping macrocells and small cells in order to provide efficient channel sharing, b) select the right RATs to serve nodes with different QoS and QoE requirements, c) perform mobility management, and d) achieve a fair coexistence among heterogeneous networks (e.g., licensed cellular and unlicensed WiFi networks must cooperate to achieve fairness).

In addition to the traditional direct message exchange between a node and the base station of a macrocell or a small cell, device-to-device (D2D) communication [46] enables

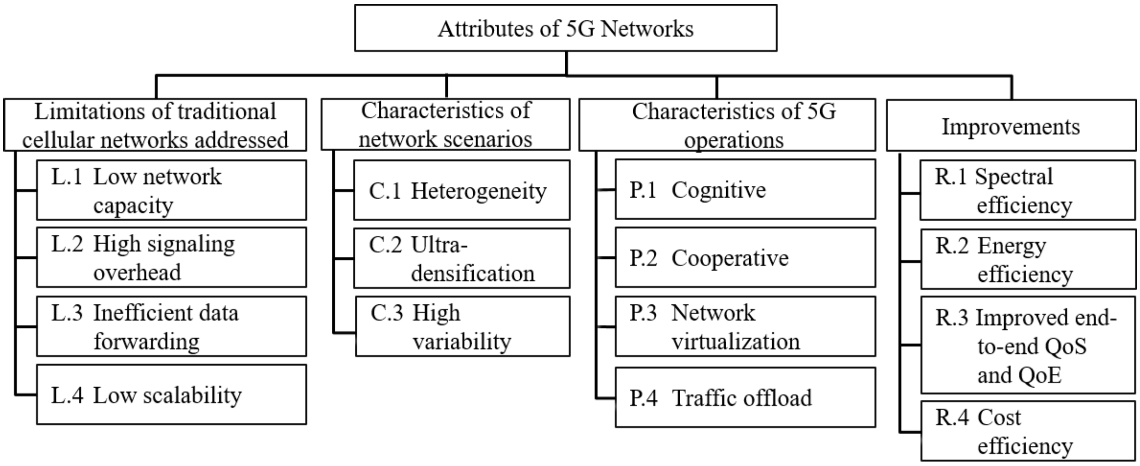

FIGURE 4. **Attributes of 5G networks.**

direct message exchange between a pair of neighboring nodes in an ad hoc manner without going through a base station. There are three main advantages of D2D, which are to: a) address inefficient data forwarding and reduces signaling overhead. This is because nodes can handover from one network cell to another without accessing the base station, and can access to the Internet without going through a base station or the core network [47], b) extend coverage, and c) improve reliability as nodes can communicate among themselves even when the base station has become overloaded or partially unavailable (e.g., due to disaster). Nevertheless, the main disadvantage of D2D is the lack of a feasible and well established business model for mobile network operators (MNOs).

CC can be embedded at different levels to learn new knowledge and adapt to network dynamics: in the clouds for global-level intelligence, in the base stations of the macrocells for macrocell-level intelligence, and in the base stations of the small cells for small cell-level intelligence. CC can also be embedded in nodes for node-level intelligence, although this can increase technical requirements on individual nodes.

3) CHANNEL ACCESS IN 5G

In 5G networks, each node or base station can access a cellular, cognitive or unlicensed channel. Cellular channels are the licensed channels allocated to the cellular networks, cognitive channels are the licensed channels with white spaces (e.g., TVWS), and unlicensed channels (e.g., ISM, U-NII, and mmWave) are those that can be accessed by both licensed and unlicensed users. In unlicensed channels, a node or base station must achieve a fair coexistence among heterogeneous networks. For instance, listen-before-talk (LBT) requires cellular nodes to perform channel sensing prior to packet transmission, which has been practiced by unlicensed user (e.g., WiFi node). Due to the heterogeneity of the channels, white spaces are matched with the right applications to achieve network objectives. Four examples are presented. *Firstly*, macrocells transmit with higher transmission power in cellular channels for long-range transmission, while small cells transmit with lower transmission power in cognitive and unlicensed channels for short-range transmission. *Secondly*, nodes or base stations access cellular channels for better QoS and QoE performance, and cognitive and unlicensed channels for relaxed performance, in order to achieve service differentiation [48]. *Thirdly*, nodes access cellular channels for long-range transmission, and cognitive and unlicensed channels for short-range transmission. This helps to facilitate multi-hop and cooperative communication with the objective of improving the QoS performance. *Fourthly*, nodes access cellular channels to communicate with base station, and cognitive channels to communicate with neighboring nodes, in order to offload traffic from highly utilized cellular channels.

Due to the ultra-dense nature of 5G networks, particularly in IoTs and machine-to-machine (M2M) communication, non-orthogonal multiple access (NOMA) [49]–[51] has been proposed to provide higher channel capacity and spectral efficiency. Using NOMA, a base station uses different power allocation coefficients to multiplex (or superpose) signals for different receivers in the power domain, and subsequently broadcasts the multiplexed signals. Higher power allocation coefficient is used for receivers located further away from the base station. Upon receiving the multiplexed signals, a receiver separates the signals (e.g., using codeword-level successive interference cancellation) and either regards the signal for the other node as noise or cancels it.

B. ATTRIBUTES OF 5G

5G addresses the limitations of the traditional cellular networks and caters to the needs of the next-generation cellular network scenarios with its distinguishing operational characteristics in order to provide network-wide performance enhancement. Figure 4 presents the attributes of 5G described in the rest of this section.

1) LIMITATIONS OF TRADITIONAL CELLULAR NETWORKS ADDRESSED BY 5G

The cellular network has evolved over the last few decades from first-generation (1G) to fourth-generation (4G). To deploy, support and scale next-generation applications that require high data rate for high-speed transmission of media-rich traffic, traditional cellular networks are plagued with *four* main limitations as follows:

- L.1 *Low network capacity.* The traditional cellular network has low network capacity, mainly attributed to the low bandwidth requirement of voice connections that it caters for. This is compounded by the spectrum scarcity problem since only cellular channels are used. The lack of a unified scheduler to coordinate the RATs at a base station is another factor as it can cause a RAT to become either highly utilized or underutilized. The RATs can access cellular, cognitive and unlicensed channels. In addition, the requirement of the traditional cellular networks for a node to establish and maintain at least a single packet data network (PDN) connection can cause low utilization by some next-generation applications, particularly IoTs that may have low data rate traffic with small-sized packets.
- L.2 *High signaling overhead.* The traditional cellular network has high signaling overhead, mainly attributed to the centralized management by the base stations. Nodes must exchange control messages with their respective base stations to establish and maintain at least a single PDN connection, and this must be performed whenever a node handovers to a network from another network. Some next-generation applications, particularly IoTs, may cause high signaling overhead as nodes may access the network simultaneously to report a newly detected event in a real-time manner.
- L.3 *Inefficient data forwarding.* The traditional cellular network has inefficient data forwarding, mainly attributed to the need of the packets of a sender node to go through the base station although the destination node is within its transmission range. This inefficiency can increase the network congestion level (i.e., reduced bandwidth availability at the base station) and reduce network performance (e.g., increased delay and energy consumption). In addition, a mobile node must maintain a PDN connection with its operating base station throughout an IP multimedia subsystem (IMS) voice session although it has moved and become physically closer to another base station.
- L.4 *Low scalability.* The traditional cellular network uses specialized hardware units, which can be inflexible to incorporate next-generation applications and extends the network size due to higher replacement investment, including capital and operational expenses.

2) CHARACTERISTICS OF NEXT-GENERATION CELLULAR NETWORK SCENARIOS SUPPORTED BY 5G

The next-generation cellular network scenarios possess the following *three* main characteristics in which 5G must cater for:

1) *Heterogeneity.* Heterogeneity stems from the presence of a diverse range of network entities and characteristics, including: a) heterogeneous network (HetNet) cells (e.g., macrocells, picocells and femtocells), b) network operators, c) RATs (e.g., IEEE 802.11, 3G and 4G), d) channels (e.g., cellular, cognitive and unlicensed), e) wireless equipment and devices (e.g., laptops, smartphones and tablets), f) network characteristics (e.g., indoor or outdoor environment, as well as with or without cognitive capability), and g) application characteristics (e.g., QoS requirements and the number of required connections). The challenge is to enable interoperability among heterogeneous networks and nodes so that nodes can handover from one network to another different network.
2) *Ultra-densification.* The increasing number of base stations and nodes has caused an exponential growth of data traffic, leading to increased contention level among nodes for channel access. The total number of mobile devices is expected to increase to 11.6 billion by 2020, contributing to an increase in annual traffic load of up to 360 exabytes in mobile networks [52], and an increase in traffic density of up to 900 Gbps/km^2 in ultra-dense area [53]. The challenge is to increase the data rate in the order of multi-Gbps in order to support the ultra-dense applications, as well as to support the large number of base stations and nodes via the intelligent and autonomous capabilities offered by CC contributing to reduced total cost of ownership (TCO) incurred by network operators.
3) *High variability.* Due to the diversity of next-generation applications, the user behavior can vary and the traffic can be bursty, causing the traffic to distribute in a random manner across space and time with a peak-to-mean ratio of up to 100:1. As an example, in IoT, a large number of nodes that send packets to a single base station simultaneously can cause network congestion in parts of the networks. The challenge is to address the insufficiency (or overabundance) of channel capacity caused by high variation in user and traffic behaviors.

3) CHARACTERISTICS OF 5G OPERATIONS

There are *four* main characteristics of 5G operations contributing to network performance enhancement as follows:

- P.1 *Cognitive.* Networks and nodes can achieve intelligent and autonomous capabilities using CC to learn new knowledge and adapt to the dynamicity of network condition and application. This can be achieved

TABLE 1. Summary of attributes for schemes based on CR and CC in 5G.

Schemes	Limitations addressed				Characteristics of network scenarios			Characteristics of 5G operations				Improvements			
	(L.1) Low network capacity	(L.2) High signaling overhead	(L.3) Inefficient data forwarding	(L.4) Low scalability	(C.1) Heterogeneity	(C.2) Ultra-densification	(C.3) High variability	(P.1) Cognitive	(P.2) Cooperative	(P3) Network virtualization	(P.4) Traffic offload	(R.1) Spectral efficiency	(R.2) Energy efficiency	(R.3) Improved QoS and QoE	(R.4) Cost efficiency
Cognition cycle [55]				*		*	*	*			*			*	
Cognition cycle with knowledge transfer [56]	*					*		*			*		*	*	
Channel sensing [57]	*				*		*	*			*	*		*	
Channel sensing [39]	*	*			*	*		*	*			*	*	*	
Spectrum sharing and ICIC [2]	*				*	*		*				*		*	*
Spectrum sharing and ICIC [59]	*				*	*			*		*	*	*		
Spectrum sharing and ICIC [4]	*				*				*			*		*	
Spectrum sharing and ICIC [5]	*				*			*				*		*	
D2D communication [74]	*		*		*						*	*		*	
Directional antenna and beamforming [77]	*					*		*			*		*	*	
Directional antenna and beamforming [78]	*				*						*			*	
Multiple RATs [78]	*				*						*	*		*	
Scheduling [79]		*			*			*		*				*	
Scheduling [44]		*			*					*		*			
Energy harvesting [38]					*							*	*		

through five main steps as shown in Figure 1: a) the *observe* stage collects data and observations from the operating environment, b) the *analyze* stage analyzes data and observations, c) the *plan and decide* stage considers alternatives and selects an action, d) the *act* stage executes the action that maximizes network performance, and e) the *learn* stage learns new knowledge. This enables networks and nodes to solve complex networking problems with minimal human interaction.

P.2 *Cooperative.* Networks and nodes, which may be heterogeneous (e.g., licensed cellular and unlicensed WiFi networks) in nature, cooperate among themselves [54]. A common data repository or a cloud allows networks and nodes to share information and knowledge, and subsequently coordinate actions among themselves. This ensures that actions selected by networks and nodes are not in contradiction among themselves, leading to improved network-wide performance.

P.3 *Network virtualization.* Centralized entities, such as clouds in the core network, pool together network information, resources and services, to provide advanced shared services (e.g., routing and scheduling).

P.4 *Traffic offload.* Traffic is offloaded from a network entity to another, which can be from a macrocell to a small cell via D2D communication, from a congested channel to a less congested channel, and others.

4) IMPROVEMENTS ACHIEVED BY 5G

5G is aspired to achieve the following *four* main improvements:

R.1 *Spectral efficiency* aims to improve the efficiency of spectrum utilization in order to increase channel capacity.

R.2 *Energy efficiency* aims to reduce energy consumption at different levels: over a link, within a network cell, or across multiple network cells.

R.3 *Improved end-to-end QoS and QoE* aims to fulfill a diverse range of QoS and QoE requirements of different networks and nodes.

R.4 *Cost efficiency* aims to increase the monetary gain of network operators.

III. COGNITIVE RADIO AND COGNITION CYCLE FOR 5G NETWORKS: A REVIEW OF SOME RECENT WORK

This section presents a review of some recent works on CR and CC applied to 5G.

Section III-A presents CC applied to 5G. Sections III-B and III-C present two main functions of CR, namely channel sensing as well as spectrum sharing and intercell interference coordination, applied to 5G, respectively. Sections III-D–III-H present the application of CR and CC to various schemes in 5G, including D2D communication, directional antenna and beamforming, multiple radio access technologies, scheduling, as well as energy harvesting. A summary of the attributes for each scheme is presented in Table 1. In addition, platform implementation of a CR-based 5G network is presented in Section III-I.

A. APPLICATION OF COGNITION CYCLE TO 5G APPLICATIONS

Cognition cycle, which represents intelligent and autonomous capabilities, enables an agent, or a decision maker, to learn new knowledge and adapt to the dynamicity of network condition and application. Knowledge transfer allows the knowledge of an agent's CC shared with agents with or without CC. In the context of 5G, the investigation of CC has been limited, and an artificial intelligence approach called reinforcement learning has been applied as presented in the rest of this subsection.

In [55], a CC is embedded in the base station of a picocell to expand or shrink its coverage in order to appropriately offload traffic from the base station of a macrocell in a CC-based 5G network. Larger coverage of the picocell increases traffic offload from the base station of a macrocell to the base station of the picocell; however, the coverage of the picocell is constrained by the SINR of the nodes at the edge of the picocell as lower SINR causes unsuccessful packet transmission. The main objective is to achieve optimal load balancing among macrocells and picocells, while ensuring successful packet transmission. The proposed approach addresses low scalability (L.4) under ultra-dense (C.2) and highly variable (C.3) networks using the cognitive (P.1) and traffic offload (P.4) operational characteristics in order to improve QoS performance (i.e., improved throughput) (R.3). The CC is based on Q-learning, which is a popular temporal difference learning approach of reinforcement learning. The Q-learning model, which is embedded in the base station of a picocell, has three main representations: a) *state* s_t represents the traffic loads of the base stations of the macrocells and picocells, and the SINR of the nodes at the edge of the picocell, b) *action* a_t represents one of the possible coverages of the picocell, and c) *reward* $r_t(s_t, a_t)$ represents the average throughput of each node in the picocell. At time instant t, the base station of the picocell, as an agent, observes its state s_t and chooses an action a_t. At the next time instant $t+1$, it receives a reward $r_{t+1}(s_t, a_t)$, and updates its knowledge represented by Q-value $Q_t(s_t, a_t)$, which is the long-term reward of the state-action pair. Next, the agent observes the next state $s_t \leftarrow s_{t+1}$ and selects the optimal action $a_t^* = \text{argmax}_a[Q_t(s_t, a)]$ that maximizes the Q-value based on the state s_t. The optimal action a_t^* is the coverage with the highest Q-value that provides the best possible throughput performance contributed by load balancing and successful packet transmission rate. This approach has shown to increase throughput performance.

In [56], knowledge—in particular, the best possible channel for each of its nodes—learnt from channel assignment by a base station is shared with its nodes for user association in a CR-based 5G network. In channel assignment, a base station selects and assigns an operating channel to each of its nodes in the presence of network dynamics (i.e., the intercell interference levels and the amount of white spaces). In user association, a node selects and associates with one of the base stations that provides the best possible successful packet transmission rate. The main objective is to enable the base station of a network cell to handover its nodes to neighboring or overlapping network cells whenever the intercell interference levels of the nodes increase. The proposed approach addresses low network capacity (L.1) under ultra-dense (C.2) networks using the cognitive (P.1) and traffic offload (P.4) operational characteristics in order to improve energy efficiency (R.2) and QoS performance (i.e., reduced delay and the number of retransmissions) (R.3). The CC is based on Q-learning. The Q-learning model, which is embedded in a base station, selects the operating channel with the best possible successful packet transmission rate, and it has two main representations: a) *action* a_t represents one of the possible operating channels, and b) *reward* $r_t(a_t)$ represents a positive (negative) constant value for a successful (unsuccessful) packet transmission. The base station, as an agent, updates Q-value $Q_t(a_t)$, selects the optimal action a_t^*, and shares the Q-values with its nodes for user association. The optimal action a_t^* is the operating channel with the highest Q-value that provides the best possible successful packet transmission rate. Each node receives Q-values from its neighboring base stations. For each base station, the node aggregates the Q-values of all the available channels at the base station. Subsequently, the node associates with the base station that provides the highest possible Q-value while ensuring that SINR is acceptable. The enhanced user association scheme has shown to reduce unsuccessful packet transmission, contributing to reduced delay and number of retransmissions, caused by inappropriate user association. In addition, base stations without nodes (e.g., due to high intercell interference level or low amount of white spaces) switch to sleep mode. This approach has shown to reduce energy consumption without significantly jeopardizing QoS performance caused by retransmission.

B. CHANNEL SENSING

Channel sensing enables SUs to sense for the real-time occupancy of PUs' activities in the cognitive channels in order to identify white spaces.

In [57], base stations perform channel sensing and estimate the probability of a channel switch caused by PUs' activities and poor channel quality in a CR-based 5G networks. Channel quality is represented using a hidden Markov model based on channel statistical information, including the detection probability, false alarm probability, and channel idle duration. The main objective is to estimate the probability of a channel switch in a time slot based on the estimation on the presence of PUs' activities (i.e., the arrival time and the duration of PUs' activities in a time slot). The proposed approach addresses low network capacity (L.1) under heterogeneous (C.1) and highly variable (C.3) networks using the cognitive (P.1) and traffic offload (P.4) operational characteristics in order to improve spectral efficiency (R.1) and QoS performance (i.e., reduced delay) (R.3). A channel switch can cause non-negligible delay due to the time needed to identify the next high quality channels in channel sensing and establish high quality links in channel access. The estimated probability of a channel switch can be used for two purposes. *Firstly*, the base station can pre-fetch a lower quality version of a multimedia application (e.g., video) to the receiver prior to a channel switch in order to provide seamless multimedia service with minimal interruption. *Secondly*, the base station can allocate the appropriate channels to its traffic with different traffic patterns (e.g., traffic rate fluctuations) and QoS requirements (e.g., class priorities). In other words, higher quality channels are allocated to higher priority traffic classes. This approach has shown to minimize the number of interruptions and delay caused by channel switches, as well as interference to PUs' activities.

Collaborative channel sensing enables SUs to perform channel sensing individually and send their respective sensing outcomes to a fusion center (e.g., a base station) that makes final decisions on channel availability. Collaborative channel sensing addresses the effects of multipath channel fading and shadowing experienced in channel sensing performed individually. In [39], spectrum agents, which are embedded with CC, are embedded in base station or access point to shift the cognitive capability from SUs to the spectrum agents in a CR-based 5G network. Hence, each SU is not required to perform channel sensing and analysis. The main objective is to address the shortcomings of collaborative channel sensing that requires every single SU to be embedded with CC. The proposed approach addresses low network capacity (L.1) and high signaling overhead (L.2) under heterogeneous (C.1) and ultra-dense (C.2) networks using the cognitive (P.1) and cooperative (P.2) operational characteristics in order to improve spectral efficiency (R.1), energy efficiency (R.2) and QoS performance (R.3). The proposed approach: a) reduces energy consumption and overheads compared to the traditional channel sensing approach that requires sensing outcomes, despite being redundant, to be sent from SUs in the same region to a base station, b) improves the efficiency of channel utilization compared to the traditional channel sensing approach that requires SUs to continuously alternate between channel sensing and channel access (or data transmission), and c) reduces hardware cost compared to the traditional channel sensing approach that requires each SU to be embedded with CC to sense a wide range of spectrum.

C. SPECTRUM SHARING AND INTERCELL INTERFERENCE COORDINATION

Spectrum sharing and ICIC enable nodes to share radio resources among themselves with reduced co-channel and intercell interference. Long-term evolution in unlicensed spectrum, also known as licensed-assisted access, has been investigated in [4]–[6] to allow nodes to access either licensed networks (i.e., cellular networks) or unlicensed networks (i.e., WiFi networks) dynamically. This can be achieved by three main approaches [6]. Firstly, preventing similar operating channels selected by neighboring nodes or network cells, or even overlapping network cells. Secondly, undergoing backoff so that non-contention-based cellular nodes do not interfere with contention-based WiFi nodes, causing unfairness to WiFi networks. Thirdly, addressing the dynamic WiFi activities, which causes the performance of cellular nodes to fluctuate, resulting in QoS provisioning to cellular nodes that varies with WiFi activities. The main objective is to achieve a fair coexistence among different networks. The rest of this subsection considers an omnidirectional antenna, and the use of directional antenna and beamforming, which helps to reduce co-channel and intercell interference as well, is treated separately in Section III-E.

In [2], small cells, as the SUs, select operating channels with higher channel capacity and lower intercell interference level to reduce intercell interference among neighboring and overlapping network cells in order to maximize network-wide channel capacity (or throughput) in a CR-based 5G network. The main objective is to select operating channels with higher SINR and amount of white spaces. The proposed approach addresses low network capacity (L.1) under heterogeneous (C.1) and ultra-dense (C.2) networks using the cognitive (P.1) operational characteristic in order to improve spectral efficiency (R.1), QoS performance (i.e., improved throughput) (R.3) and cost efficiency (R.4). CC has been embedded in the base stations of small cells, and it is based on genetic algorithm. Using genetic algorithm, there are two separate objective functions: one maximizes throughput and SINR, and another one minimizes intercell interference and hardware cost. The physical location of a node is represented by a chromosome. Crossover is performed to increase the diversity of each chromosome, and mutation is performed to prevent from local optimal solutions. Performance achieved by genetic algorithm is compared with that of graph theory. Using graph theory, a network is represented by a set of nodes and links, and subsequently a depth-first search algorithm [58] is used to achieve the objectives while satisfying constraints in terms of data rate and intercell interference. Genetic algorithm has shown to outperform graph theory with lower number of macrocells and higher number of small cells deployed, which helps to reduce intercell interference among neighboring and overlapping network cells, contributing to higher throughput and SINR while reducing hardware cost.

In [59], the main objective is to associate a node with the base station (or network cell) that provides the best possible successful packet transmission rate in a user association scheme in a CR-based 5G network. The proposed approach addresses low network capacity (L.1) under heterogeneous (C.1) and ultra-dense (C.2) networks using the cooperative (P.2) and traffic offload (P.4) operational characteristics in order to improve spectral efficiency (R.1) and energy efficiency (R.2). Consider a node from a macrocell that communicates with the base station of the macrocell. Suppose, it is physically closer to the base station of a small cell. Consequently, it experiences high intercell interference resulting in low SINR. Subsequently, the macrocell associates the node of the base station of the small cell, which causes the highest interference level to the node, and leases some of its white spaces to the small cell through spectrum leasing. Due to shorter-range transmissions between the node and the base station of the small cell, energy consumption and intercell interference among macrocells and small cells are reduced.

In [4], a backoff mechanism is introduced to LBT so that cellular nodes perform channel sensing and backoff prior to packet transmission in unlicensed channels. The proposed approach addresses low network capacity (L.1) under heterogeneous (C.1) networks using the cooperative (P.2) operational characteristic in order to improve spectral efficiency (R.1) and QoS performance (R.3). Each node performs channel sensing based on energy detection. Before a packet transmission, a node must backoff for a random duration at a millisecond scale (i.e., a random number of time slots drawn from a contention window, whose size can double due to failed packet transmission). The node undergoes backoff whenever a channel is sensed idle, and freezes the backoff whenever the channel is sensed busy. Once the backoff completes, the node senses the channel and transmits packet if it is sensed idle for a specified period [6]. For concurrent transmission in multiple channels, a node undergoes similar backoff mechanism in one of the channels (i.e., a primary channel), and once the backoff completes, the node performs channel sensing and transmits packets in the primary channel, as well as the rest of the channels (i.e., secondary channels) whenever they are sensed idle. This approach has shown to increase throughput performance and to reduce the outage probability of application (e.g., voice over Internet protocol). LBT is also investigated in [6], and it has shown to increase throughput performance.

In [5], a centralized controller is introduced to manage radio resources of unlicensed channels used by cellular and WiFi nodes. The proposed approach addresses low network capacity (L.1) under heterogeneous (C.1) networks using the cognitive (P.1) operational characteristic in order to improve spectral efficiency (R.1) and QoS performance (R.3). The time period of an unlicensed channel is segregated into contention free period (CFP) and contention period (CP). During CFP, the controller allows cellular nodes to access the unlicensed channel, and allocates resources among them based on orthogonal time-frequency allocation and channel state information to minimize collision and interference. During CP, WiFi nodes access the unlicensed channel. Since channel access for cellular nodes and WiFi nodes are separated, there is no interference and contention among them. The controller adjusts the durations of CP and CFP dynamically based on node density, traffic load and service requirement. This approach has shown to increase throughput performance.

D. D2D COMMUNICATION

D2D communication enables direct message exchange between a pair of neighboring nodes in an ad hoc manner without going through a base station in 5G. D2D communication has already been introduced into 4G LTE, and it has been approved by the Third-generation Partnership Project (3GPP) standardization community since 2014 [60]. Hence, there have been investigations on D2D in the context of 4G as presented extensively in several review papers, including [46], [61]–[65]. In the context of 5G, a discussion of the D2D concept is presented in [66]and [67]. D2D communication accesses cellular channels using either overlay or underlay approaches. In the overlay D2D approach, separate non-overlapping orthogonal channels are allocated for D2D and cellular communications (i.e., going through a base station), respectively; so the main focus is to minimize underutilization in cellular channels. This includes allocating/ scheduling radio resources for D2D communications by base station [68], [69]. In the underlay D2D approach, similar channels are used for both D2D and cellular communications, and so the main focus is to minimize interference among D2D and cellular nodes. This includes: a) reducing the transmission power of a D2D node based on path loss estimation between a D2D node and a base station [70], and b) allocating/ scheduling radio resources among D2D and cellular nodes based on interference levels [71]–[73]. Nevertheless, the investigation of D2D communications in the context of 5G can be taken further.

In [74], vehicular nodes, as the SUs, explore and use white spaces in a CR-based 5G vehicular network using the underlay approach. There are three main objectives with the goal of supporting vehicular applications such as vehicular safety application: a) to increase link connectivity (or reliability), which is more important than throughput performance, among vehicular nodes with high mobility, b) to reduce interference among vehicular nodes and cellular nodes (or PUs), and c) to reduce delay by offloading traffic from base stations. The proposed approach addresses low network capacity (L.1) and inefficient data forwarding (L.3) under heterogeneous (C.1) networks using the traffic offload (P.4) operational characteristic in order to improve spectral efficiency (R.1) and QoS performance (i.e., reduced delay) (R.3). The vehicular nodes explore underutilized frequency-division duplexing (FDD) uplink (UL) bands allocated for long term evolution-advanced (LTE-A). FDD UL bands provide more reliable transmission and larger coverage as the interference level can be monitored and controlled in these licensed bands, providing better network performance. There are two main steps. *Firstly*, vehicular nodes sense the path loss level between itself and its base station. *Secondly*, the base station calculates the maximum transmission power of a vehicular node so that it does not interfere with existing PUs in FDD UL bands. Subsequently, the vehicular nodes communicate among themselves via D2D. D2D has shown to maintain delay despite increasing number of vehicular nodes as messages are exchanged directly among neighboring vehicular nodes in an ad hoc manner without going through a base station. D2D has also shown to reduce interference to PUs' activities for using the FDD UL bands.

E. DIRECTIONAL ANTENNA AND BEAMFORMING

Traditionally, omnidirectional antenna radiates transmission power in all directions, causing increased interference among neighboring or overlapping nodes and network cells. Software-defined directional antenna with beamforming, supported by massive MIMO that contains an array of small-sized antennas with narrow beams (e.g., up to 16 antennas in each sector) [75], [76], can steer a beam in arbitrary shapes and directions (or pointing angles) to form a sector through self- or remote regulation in order to provide highly directive transmission [28]. This increases channel capacity, SINR, and coverage, as well as reduces energy consumption for the same transmission power used by omnidirectional antenna.

In [77], the base station of a macrocell, which contains multiple antennas, selects a link, represented by its beam and operating channel, connecting to a neighboring base station based on traffic load and network interference in a CC-based 5G network. The main objective is to divert network traffic to a lower number of beams so that more beams can be switched off (or sleep) to reduce energy consumption. The proposed approach addresses low network capacity (L.1) under ultra-dense (C.2) networks using the cognitive (P.1) and traffic offload (P.4) operational characteristics in order to improve energy efficiency (R.2) and QoS performance (R.3). CC has been embedded in the base station of a macrocell, and it is based on RL. There are three main steps. *Firstly*, the base station identifies the beams whose number of active nodes is lower than a threshold, and informs the nodes of the respective beams. *Secondly*, upon notification, a node identifies an alternative link with signal strength greater than a threshold to maintain its QoS, and sends a request to the base station of the alternative link. *Thirdly*, the base station of the alternative link accepts the request if the respective beam has a sufficient number of active nodes. This approach has shown to reduce energy consumption without significantly jeopardizing QoS performance, including throughput, delay and blocking probability, in networks with low traffic load. Nevertheless, performance enhancement is minimal in networks with high traffic load as most beams are switched on.

In [78], a node, which contains multiple antennas, selects a link, represented by its beam and operating channel, connecting to a neighboring mobile node based on its next approximated location (or the direction of arrival) in a CC-based mobile network. The main objective is to maintain the link established using narrow beams for data transmission. The proposed approach addresses low network capacity (L.1) under heterogeneous (C.1) networks using the traffic offload (P.4) operational characteristic in order to improve QoS performance (i.e., improved reliability with reduced number of re-beamforming instances) (R.3). A node uses sensors (e.g., motion sensors embedded in most modern devices) to track and approximate the next location of its neighboring mobile nodes, and subsequently selects an appropriate narrow beam to maintain the link. This helps to reduce the number of re-beamforming instances caused by link breakage as a result of beam misalignment due to mobility, and the overhead incurred during narrow beam setup by reducing the search space of exhaustive beam search. This approach has shown to reduce delay and overhead caused by frequent re-beamforming.

F. MULTIPLE RADIO ACCESS TECHNOLOGIES

With multiple RATs, nodes and network cells can access both control plane and data plane simultaneously. As is commonly seen in CR networks, two different channels can be used at the same time, specifically a common control channel for control message exchange and data channels for data transmission. In [78], the data plane of a picocell uses a higher frequency 60 GHz mmWave radio access network for data transmission, and the control plane of the picocell uses a lower frequency 2.4/ 5 GHz radio access network for control message exchange. The proposed approach addresses low network capacity (L.1) under heterogeneous (C.1) networks using the traffic offload (P.4) operational characteristic in order to improve spectral efficiency (R.1) and QoS performance (i.e., improved throughput and reduced delay) (R.3). The 60 GHz mmWave radio access network provides up to 7 Gbps short-range transmission (e.g., 10 m to 20 m) with poor wall penetration suitable for achieving higher data rate transmission, while the 2.4/ 5 GHz radio access network provides up to 1 Gbps long-range transmission with better wall penetration suitable for achieving large coverage transmission. Using multiple RATs simultaneously, a single 2.4/ 5 GHz control plane can manage several 60 GHz data planes in order to achieve high data transmission rate and large coverage transmission. A node must first sends a request for data transmission via the 2.4/ 5 GHz control plane, and then exchange data via the 60 GHz data plane. This approach has shown to increase transmission probability and reduce delay.

G. SCHEDULING

Scheduling allocates network and radio resources to nodes and network cells. More than a single scheduler can be used to allocate resources at different granularities, such as time and space.

In [79], there are two types of schedulers that allocate transmission opportunities at different time granularities in a CC-based 5G network. *Centralized scheduler*, which is embedded in the control plane in the core networks, has a longer time scale, while *distributed scheduler*, which is embedded in the data plane at the base stations, has a shorter time scale (e.g., a time slot). The main objective is to maximize the transmission rate of the base stations. The proposed approach addresses high signaling overhead (L.2) under heterogeneous (C.1) networks using the cognitive (P.1) and network virtualization (P.3) operational characteristics in order to improve QoS performance (i.e., improved throughput) (R.3). The centralized scheduler in the core networks is embedded with a CC based on a game theoretic approach. A utility function that represents the total transmission rate of the base stations is defined, and a coarse correlated equilibrium is identified to provide mappings among dynamic states (i.e., round-trip time representing the queue size, and channel gain representing the channel quality of the candidate channels) and actions (i.e., transmission power of base stations) that can maximize the transmission rate of the base stations. The distributed scheduler at the base station uses state-action mappings and Lyapunov stochastic optimization to allocate the transmission opportunities, and selects the transmission power for its nodes in order to maximize throughput performance while fulfilling the nodes' QoS requirements. The approach has shown to increase throughput and reduce queue size.

In [44], a small cell with multiple antennas is segregated into interior and exterior areas, and the exterior area is segregated into a number of sectors based on SINR in order to reduce co-channel and intercell interference, particularly interference among nodes at the edge of the network cell. Each sector is served by an antenna. There are two types of schedulers that allocate channels and transmission opportunities at different time granularities. *Spatial scheduler*, which is embedded at the base station of a small cell, has a longer time scale (e.g., 10 ms), while *medium access control* (MAC) scheduler, which is embedded at the antenna of a sector, has a shorter time scale (e.g., 1 ms). The main objective is to maximize the transmission rate of a base station. The proposed approach addresses high signaling overhead (L.2) under heterogeneous (C.1) networks using the network virtualization (P.3) operational characteristic in order to improve spectral efficiency (R.1). The spatial scheduler at the base station selects and schedules channels to each sector such that the operating channels are different from those used in neighboring cells. The MAC scheduler schedules packet transmission within a sector based on SINR and traffic load using the selected channels, and selects the best possible modulation and coding schemes for each packet. Nodes from a sector with high traffic load can be reassigned to another one with lower traffic load. The approach has shown to reduce blocking probability caused by insufficient radio resources.

H. ENERGY HARVESTING

Energy harvesting produces energy from ambient fields, such as solar, wind, and vibration, to improve self-sustainability among nodes and network cells, particularly small cells. In [38], nodes adjust their transmission power levels based on the QoS requirement and SINR in a CR-based 5G networks. Using CR, energy consumption is higher due to two main CR operations, namely channel sensing and dynamic channel access. The main objective is to achieve a balance between energy harvesting and energy consumption. The proposed approach is applied to improve spectral efficiency (R.1) and energy efficiency (R.2) under heterogeneous (C.1) networks. Energy harvesting depends on the spatio-temporal behavior of the ambient field, and the properties of transducers (e.g., photovoltaic) that convert ambient field to power. Energy consumption depends on network architecture and operation, traffic load and the required QoS requirement. Energy harvesting has shown to increase the probability of successful packet transmission, as well as to reduce energy outage probability, which is equivalent to the probability of the total energy consumption exceeding the harvested energy.

I. IMPLEMENTATION OF PLATFORM FOR COGNITION-INSPIRED 5G CELLULAR NETWORKS

The investigation of cognition-inspired 5G cellular networks has been limited to either simulation or theoretical studies. In the literature, there have been platform implementations to investigate the physical layer aspects in 5G networks, including transceivers [80], decoders [81], multiple input multiple output (MIMO) [82]–[84], synchronization [85], channel measuring and modeling [86], [87], modulation [88], channel bonding [89], physical-layer security [90], beamforming [91], blind separation [92], and interference suppression [93]. Nevertheless, platform implementation to investigate the networking aspect is comparatively limited. The rest of this subsection presents investigations of the networking aspect over testbeds.

In [94], a transceiver for 5G is implemented on SDN platform. The transceiver is based on IEEE 802.15.4k direct sequence spreading spectrum (DSSS) physical (PHY) [95] that provides low-data rate long-range low-power transmissions suitable for M2M communication and sensor communication in 5G. CR is adopted to allow SUs to explore and use the white spaces owned by the PUs in channel sharing for spectral efficiency improvement. The platform consists of three universal software radio peripheral (USRP)/ GNU radio units that provide real-time signal processing, including two SUs and a PU, which is the base station of a cellular network. Each USRP/ GNU radio unit performs two main functions: a) *channel sensing* to explore and sense for white spaces, and b) *channel access* to transmit data in the white spaces. The implementation of the transceiver is structured into two chains, namely *transmit chain* and *receive chain*. The transmit chain performs essential tasks, including forward error correction (e.g., using convolutional coding), encoding (e.g., using differential encoder), interleaving (e.g., using pruned bit reversal algorithm) and modulation (e.g., using binary phase-shift keying (BPSK)), as well as specific tasks in IEEE 802.15.4k DSSS PHY (e.g., generating Gold code sequence [95], spreading the input bits to chip rate based on spreading factor, and generating orthogonal variable spreading factor code). The transmit chain produces packets, each consists of a sequence of samples transmitted to the receiver in white spaces. The receive chain performs three main tasks, namely sensing channel (e.g., calculating the power of the received signals using fast Fourier transform), detecting samples (e.g., calculating and identifying the peak of the autocorrelation value of samples in order to detect the presence of a packet with reduced detection time under noisy environment), and decoding samples that reverses the processes of the transmit chain. The platform has demonstrated to provide acceptable packet delivery rate, as well as accurate frame detection and acceptable time required for frame synchronization.

In [96], a MAC protocol is implemented on SDN platform. The MAC protocol enables femtocell base stations to perform transmission power allocation using Q-learning. The platform consists of six USRP/ GNU radio units, including one macrocell base station, one macrocell node, two femtocell base stations, and two femtocell nodes. A USRP/ GNU radio unit performs four main functions: a) MAC functions include node synchronization to minimize misalignment of clock frequencies among nodes, and exchanges of channel state information, b) Q-learning-based power allocation enables a femtocell base station to allocate transmission power based on channel state information. The Q-learning model, which is embedded in each femtocell base station, has three main representations: state represents the interference level at macrocell base stations, action represents its transmission power, and reward is based on its network capacity. The femtocell base stations exchange the Q-values of their current state-action pairs so that a joint action that maximizes the global Q-value can be selected, c) physical-layer functions include packet transmission, and d) SINR estimation, which is implemented using the probe block in GNU radio, calculates the mean and variance of channel sensing outcomes subsequently used to estimate signal and noise powers applied in network capacity calculation. The MAC protocol has demonstrated to provide higher aggregated network capacity among femtocell base stations while reducing intercell interference to macrocells.

IV. LOOKING FORWARD: OPEN RESEARCH OPPORTUNITIES

This article has presented a review of some recent work on cognition-inspired 5G cellular networks. Based on the recent work presented in Section III, some possible future work directions are presented in this section.

A. COGNITION CYCLE

Various artificial intelligence approaches, such as swarm intelligence, multilayer perceptron, and genetic algorithm, can be explored to implement CC. New classes of artificial intelligence approaches, particularly deep learning that enables high-dimensional state space representation with efficient learning [97], can also be explored to address the shortcomings of the traditional artificial intelligence approaches while implementing CC. A holistic framework that facilitates collaboration between CCs at centralized (e.g., the cloud) and distributed entities (e.g., base stations of the macrocells and small cells) can be designed to improve network-wide intelligent and autonomous capabilities. Subsequently, the designed CC can be applied to the key applications of 5G cellular networks, such as smart beamforming supported by massive MIMO, high frequency mmWave communication, and ultra-densification. In order to support the key applications, beam tracking based on past memory (i.e., the regular mobility patterns and behaviors of nodes [98]) can be designed to reduce overheads caused by frequent re-beamforming.

B. CHANNEL SENSING

While the trend has been moving towards collaborative channel sensing in CR networks, independent channel sensing may find its usefulness as the sensing outcomes can be sent to a common data repository in the cloud for analyzing and processing [22]. Nevertheless, an investigation could be pursued to understand tradeoff in terms of performance and cost, such as the round-trip delay incurred in the communication between a base station and the cloud, the delay incurred in control message exchange for collaborative channel sensing, and control message overhead. Meanwhile, novel algorithms can be designed to: a) improve detection probability and reduce false alarm probability, b) improve sensing coverage, and c) select the right nodes in collaborative channel sensing.

C. SPECTRUM SHARING AND INTERCELL INTERFERENCE COORDINATION WITH DIRECTIONAL ANTENNA AND BEAMFORMING

In addition to the intercell interference among the beams of the neighboring and overlapping network cells, special attention should be given to identifying the preferred channel with white spaces for each node and network cell using CR and CC based on past memory (e.g., past satisfaction of using a particular channel) in order to reduce interference to PUs' activities at any geographical locations, particularly the cumulative interference caused by simultaneous transmissions among nodes under heterogeneous networks in the presence of a diverse range of network cells, network operators and RATs. Efforts should be made to cater to the dynamic QoS and QoE requirements of the nodes while reducing overhead incurred during coordination. To better support directional antenna and beamforming, a medium access control (MAC) protocol can be designed to support software-defined directional antenna; and service differentiation can be implemented by giving each sector a constraint on the contention window size and the maximum number of retransmissions. Due to the ultra-dense nature of 5G network, the neighbor list of a node can be highly dynamic, and CC can be used to learn about its neighbors for updating the neighbor list. To better support NOMA, CC can be used to learn and select the optimal power allocation coefficients when multiplexing signals for different receivers, and CR can be used to explore and use white spaces in the cognitive channels.

D. DEVICE-TO-DEVICE COMMUNICATION

D2D communication, being a new feature to cellular networks, enables a pair of neighboring nodes to exchange messages directly without going through a base station. This allows cell extension to support nodes physically located outside a cell or multiple hops away. Using CR, D2D communication can be performed in both cellular and cognitive channels, and by selecting the right channels with the right transmission ranges, co-channel interference reduces.

Using CC, a node can intelligently and autonomously selects one of the possible next-hop neighbor nodes, which caters to the dynamic QoS and QoE requirements of traffic, leading to its destination node which may be multiple hops away. Further investigation can be pursued to investigate the overlay and underlay approaches under the 5G context. Nevertheless, to facilitate the adoption of D2D communication, a feasible and well established business model for MNOs must be defined.

E. SCHEDULING

In 5G, a scheduler must allocate resources under heterogeneous networks in the presence of a diverse range of network cells, network operators and RATs with dynamic QoS and QoE requirements. Five examples are presented. *Firstly*, traffic loads are re-distributed from beams or base stations with heavy traffic loads to those with light traffic loads for load balancing. *Secondly*, traffic from mobile nodes is allocated to macrocells so that handover can be managed by the base stations of the macrocells, rather than small cells, in order to reduce handover instances and delay. *Thirdly*, high-priority traffic is allocated to cellular networks, while low-priority traffic is allocated to cognitive networks. *Fourthly*, a centralized scheduler, which has a longer time scale, manages the longer-term availability of white spaces (e.g., spectrum occupancy stemming from geo-location database), while the distributed scheduler, which has a shorter time scale (e.g., a time slot), manages instantaneous availability of white spaces. *Fifthly*, the duty cycle of nodes and network cells with low residual energy is adjusted to provide shorter data transmission time.

F. ENERGY HARVESTING

Ultra-densification of the base stations of small cells has inevitably increased energy consumption significantly because base stations contribute high energy consumption in communication. Energy harvesting can be adopted by the base stations of small cells to ensure self-sustainability at the small cell level. Other than achieving a balance between energy harvesting and energy consumption [38], future investigations could be pursued to: a) reduce energy outage probability by integrating energy harvesting with smart grid and batteries, b) achieve energy-aware load balancing in which energy burden is shared across the network, and c) improve energy-aware efficiency by adjusting duty cycle of nodes and network cells (e.g., base station with low residual energy has shorter data transmission time).

G. COGNITIVE RADIO AND COGNITION CYCLE SECURITY

5G possesses some essential characteristics, such as cognitive and cooperative, that pose security vulnerabilities to CR and CC operations. As an example, base stations and nodes must cooperate among themselves in collaborative channel sensing (see Section III-B); however, malicious nodes send false sensing outcomes to interfere with PUs, as well as other base stations and nodes. As another example, base stations and nodes must learn knowledge about channel availability in the operating environment; however, malicious nodes launch primary user emulation attacks whereby they impersonate the PUs and transmit jamming signals that share similar characteristics with PUs' signals, resulting in inaccurate knowledge about channel availability. Efforts should be made to address the effects of the characteristics of next-generation cellular network scenarios, including heterogeneity, ultra-densification and high variability, to security vulnerabilities in the use of CR and CC in 5G.

H. REDUCING COMPLEXITY AND ACHIEVING THE RIGHT TRADE-OFF

Due to the massive deployment and ultra-densification of the base stations of small cells, an open cognition-inspired 5G cellular network, which allows multiple network operators to share a common physical infrastructure without setting up own infrastructure to cover the same area, can be implemented. In addition, the right trade-offs, which can be user-specific or application-specific, in both local and global levels can be achieved between: a) spectral efficiency (e.g., revenue from spectrum trading) and energy efficiency (e.g., cost), b) complexity (as well as overheads) and network performance, c) instantaneous and statistical (or longer-term) optimization, and d) cell coverage, capacity, energy consumption (e.g., channel sensing in CR) and hardware cost. CC can be applied by each node and network cell to learn the right trade-off at the right time and geographical location.

In addition, more open research opportunities can be explored using Table 1. As an example, there has been lack of research interest to investigate CC to address high signaling overhead (L.2) and inefficient data forwarding (L.3) under heterogeneous (C.1) networks using the cooperative (P.2) and network virtualization (P.3) operational characteristics in order to improve spectral efficiency (R.1) and cost efficiency (R.4).

V. CONCLUSION

This article presents a review of recent work, as well as open research opportunities, on the networking aspect of cognitive radio (CR) and cognition cycle (CC) applied to 5G, which is the next-generation cellular networks. 5G is envisioned to address the limitations of traditional cellular networks (i.e., low network capacity, high signalling overhead, inefficient data forwarding, and low scalability) and to cater to the characteristics of next-generation network scenarios (i.e., heterogeneity, ultra-densification, and high variability). Various operational characteristics (i.e., cognitive, cooperative, network virtualization, and traffic offload) are being incorporated in 5G to improve spectral efficiency, energy efficiency, end-to-end QoS and QoE, and cost efficiency. This article discusses how both the core mechanisms of 5G, namely CR and CC, can improve 5G networks, and the open research opportunities therein. Future investigations could be pursued to apply CR and CC to improve various schemes in 5G, including spectrum sharing and intercell interference coordination, device-to-device communication, directional antenna and beamforming, multiple radio access technologies, scheduling, and energy harvesting. Certainly, there remains a large amount of future work to investigate the open research opportunities, and this article has laid a solid foundation and opened up new research interests in this area.

REFERENCES

[1] S.-L. Chiu, K. C.-J. Lin, G.-X. Lin, and H.-Y. We, "Empowering device-to-device networks with cross-link interference management," *IEEE Trans. Mobile Comput.*, vol. 16, no. 4, pp. 950–963, Apr. 2017.

[2] F.-H. Tseng, L.-D. Chou, H.-C. Chao, and J. Wang, "Ultra-dense small cell planning using cognitive radio network toward 5G," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 76–83, Dec. 2015.

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

[4] A. Mukherjee *et al.*, "Licensed-assisted access LTE: Coexistence with IEEE 802.11 and the evolution toward 5G," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 50–57, Jun. 2016.

[5] Q. Chen, G. Yu, and Z. Ding, "Optimizing unlicensed spectrum sharing for LTE-U and WiFi network coexistence," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2562–2574, Oct. 2016.

[6] R. Zhang, M. Wang, L. X. Cai, Z. Zheng, X. Shen, and L.-L. Xie, "LTE-unlicensed: The future of spectrum aggregation for cellular networks," *IEEE Wireless Commun.*, vol. 22, no. 3, pp. 150–159, Jun. 2015.

[7] A. Bohli and R. Bouallegue, "A novel cognitive MAC layer protocol towards 5G spectrum and energy efficiency," in *Proc. 24th Int. Conf. Softw., Telecommun. Comput. Netw. (SoftCOM)*, Split, Croatia, Sep. 2016, pp. 1–5.

[8] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[9] R. Li *et al.*, "Intelligent 5G: When cellular networks meet artificial intelligence," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 175–183, Oct. 2017.

[10] X. Wang, X. Li, and V. C. M. Leung, "Artificial intelligence-based techniques for emerging heterogeneous network: State of the arts, opportunities, and challenges," *IEEE Access*, vol. 3, pp. 1379–1391, 2015.

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

[12] J. Qadir, "Artificial intelligence based cognitive routing for cognitive radio networks," *Artif. Intell. Rev.*, vol. 45, no. 1, pp. 25–96, Jan. 2016.

[13] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A survey of self organisation in future cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 336–361, 1st Quart., 2013.

[14] A. Imran, A. Zoha, and A. Abu-Dayya, "Challenges in 5G: How to empower SON with big data for enabling 5G," *IEEE Netw.*, vol. 28, no. 6, pp. 27–33, Nov. 2014.

[15] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.

[16] A. Gupta and E. R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, Jul. 2015.

[17] S. Buzzi, C.-L. I, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, Apr. 2016.

[18] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network slicing in 5G: Survey and challenges," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 94–100, May 2017.

[19] D. Liu *et al.*, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart. 2016.

[20] O. Adigun, M. Pirmoradian, and C. Politis, "Cognitive radio for 5G wireless networks," in *Fundamentals of 5G Mobile Networks*, 1st ed. J. Rodriguez, Ed. U.K.: Wiley, 2015, pp. 149–162.

[21] F. Hu, B. Chen, and K. Zhu, "Full spectrum sharing in cognitive radio networks toward 5G: A survey," *IEEE Access*, vol. 6, pp. 15754–15776, 2018.

[22] I. Kakalou, K. E. Psannis, P. Krawiec, and R. Badea, "Cognitive radio network and network service chaining toward 5G: Challenges and requirements," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 145–151, Nov. 2017.

[23] X. Duan and X. Wang, "Authentication handover and privacy protection in 5G HetNets using software-defined networking," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 28–35, Apr. 2015.

[24] V. Sucasas, A. Radwan, S. Mumtaz, and J. Rodriguez, "Effect of noisy channels in MAC-based SSDF counter-mechanisms for 5G cognitive radio networks," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Brussels, Belgium, Aug. 2015, pp. 736–740.

[25] P. Xu, Z. Ding, X. Dai, and H. V. Poor, "A new evaluation criterion for non-orthogonal multiple access in 5G software defined networks," *IEEE Access*, vol. 3, pp. 1633–1639, 2015.

[26] M. Danneberg, R. Datta, A. Festag, and G. Fettweis, "Experimental testbed for 5G cognitive radio access in 4G LTE cellular systems," in *Proc. 8th Sensor Array Multichannel Signal Process. Workshop (SAM)*, A Coruña, Spain, Jun. 2014, pp. 321–324.

[27] A. Prata, A. S. R. Oliveira, and N. B. Carvalho, "An agile and wideband all-digital SDR receiver for 5G wireless communications," in *Proc. Euromicro Conf. Digit. Syst. Design (DSD)*, Aug. 2015, pp. 146–151.

[28] E. J. Black, A. R. Katko, and B. Deutsch, "Software defined apertures for 5G wireless network communications," in *Proc. IEEE Conf. Standards Commun. Netw. (CSCN)*, Oct. 2015, pp. 124–129.

[29] S. Bhattarai, G. Naik, and L. Hong, "A computationally efficient node-selection scheme for cooperative beamforming in cognitive radio enabled 5G systems," in *Proc. IEEE Conf. Commun. Workshop (INFOCOM WKSHPS)*, San Francisco, CA, USA, Apr. 2016, pp. 1021–1026.

[30] J. Zeng, L. Rong, and X. Su, "NO stack: A software-defined framework for 5G mobile network," in *Proc. 23rd Int. Conf. Telecommun. (ICT)*, Thessaloniki, Greece, Aug. 2016, pp. 1–5.

[31] J. Costa-Requena *et al.*, "Software defined 5G mobile backhaul," in *Proc. 1st Int. Conf. 5G Ubiquitous Connectivity (5GU)*, Nov. 2014, pp. 258–263.

[32] J. Hansen, D. E. Lucani, J. Krigslund, M. Medard, and F. H. P. Fitzek, "Network coded software defined networking: Enabling 5G transmission and storage networks," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 100–107, Sep. 2015.

[33] H. Park and T. Hwang, "Energy-efficient power control of cognitive femto users for 5G communications," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 772–785, Apr. 2016.

[34] O. Font-Bach *et al.*, "When SDR meets a 5G candidate waveform: Agile use of fragmented spectrum and interference protection in PMR networks," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 56–66, Dec. 2015.

[35] V. Jungnickel *et al.*, "Software-defined open architecture for front- and backhaul in 5G mobile networks," in *Proc. 16th Int. Conf. Transparent Opt. Netw. (ICTON)*, Jul. 2014, pp. 1–4.

[36] H. Elsawy, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Virtualized cognitive network architecture for 5G cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 78–85, Jul. 2015.

[37] K. Lin, W. Wang, X. Wang, W. Ji, and J. Wan, "QoE-driven spectrum assignment for 5G wireless networks using SDR," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 48–55, Dec. 2015.

[38] S. A. R. Zaidi, A. Afzal, M. Hafeez, M. Ghogho, D. C. Mclernon, and A. Swami, "Solar energy empowered 5G cognitive metro-cellular networks," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 70–77, Jul. 2015.

[39] Z. Zhang, W. Zhang, S. Zeadally, Y. Wang, and Y. Liu, "Cognitive radio spectrum sensing framework based on multi-agent arc hitecture for 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 34–39, Dec. 2015.

[40] Y. Lu, X. Su, J. Zeng, L. Rong, C. Yang, and X. Xu, "Software defined radio access network in 5G mobile network," in *Proc. 10th Int. Conf. Commun. Netw. China (ChinaCom)*, Shanghai, China, Aug. 2015, pp. 132–136.

[41] C. Niephaus, O. G. Aliu, M. Kretschmer, S. Hadzic, and G. Ghinea, "Wireless Back-haul: A software defined network enabled wireless Back-haul network architecture for future 5G networks," *IET Netw.*, vol. 4, no. 6, pp. 287–295, Nov. 2015.

[42] A. Mohamed, O. Onireti, M. Imran, A. Imrany, and R. Tafazolli, "Correlation-based adaptive pilot pattern in control/data separation architecture," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 2233–2238.

[43] J. Pagé and J.-M. Dricot, "Software-defined Networking for low-latency 5G core network," in *Proc. Int. Conf. Military Commun. Inf. Syst. (ICMCIS)*, Brussels, Belgium, Jun. 2016, pp. 1–7.

[44] A. A. Gebremariam, T. Bao, D. Siracusa, T. Rasveed, F. Granelli, and L. Goratti, "Dynamic strict fractional frequency reuse for software-defined 5G networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[45] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sep. 2008.
[46] Y. Wu *et al.*, "Device-to-device meets LTE-unlicensed," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 154–159, May 2016.
[47] H. Wang, S. Chen, H. Xu, M. Ai, and Y. Shi, "SoftNet: A software defined decentralized mobile network architecture toward 5G," *IEEE Netw.*, vol. 29, no. 2, pp. 16–22, Mar./Apr. 2015.
[48] 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.
[49] X. Xiong, W. Xiang, K. Zheng, H. Shen, and X. Wei, "An open source SDR-based NOMA system for 5G networks," *IEEE Wireless Commun.*, vol. 22, no. 6, pp. 24–32, Dec. 2015.
[50] X. Wei *et al.*, "Software defined radio implementation of a non-orthogonal multiple access system towards 5G," *IEEE Access*, vol. 4, pp. 9604–9613, 2016.
[51] Z. Ma, Z. Zhang, Z. Ding, P. Fan, and H. Li, "Key techniques for 5G wireless communications: Network architecture, physical layer, and MAC layer perspectives," *Sci. China Inf. Sci.*, vol. 58, no. 4, pp. 1–20, Apr. 2015.
[52] Cisco. (2016). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015–2020 White Paper*. Accessed: Mar. 2, 2018. [Online]. Available: http:\\www.cisco.com\c\dam\m\en_in\innovation\enterprise\assets\mobile-white-paper-c11-520862.pdf
[53] *Project: Scenarios, Requirements and KPIs for 5G Mobile and Wireless System*, document ICT-317669-METIS/D1.1, 2013. Accessed: Mar. 2, 2018. [Online]. Available: http:\\www.metis2020.com\wp-content\uploads\deliverables\METIS_D1.1_v1.pdf
[54] S. S. Soliman and B. Song, "Fifth generation (5G)cellular and the network for tomorrow: Cognitive and cooperative approach for energy savings," *J. Netw. Comput. Appl.*, vol. 85, pp. 84–93, May 2017.
[55] C. A. S. Franco, J. R. B. de Marca, and G. L. Siqueira, "A cognitive and cooperative SON framework for 5G mobile radio access networks," in *Proc. IEEE Globecom Workshop (GC Worksh)*, Washington, DC, USA, Dec. 2016, pp. 1–6.
[56] Q. Zhao and D. Grace, "Transfer learning for QoS aware topology management in energy efficient 5G cognitive radio networks," in *Proc. 1st Int. Conf. 5G Ubiquitous Connectivity (5GU)*, Nov. 2014, pp. 152–157.
[57] M. J. Piran, N. H. Tran, D. Y. Suh, J. B. Song, C. S. Hong, and Z. Han, "QoE-driven channel allocation and handoff management for seamless multimedia in cognitive 5G cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6569–6585, Jul. 2017.
[58] W. Ni, I. B. Collings, J. Lipman, X. Wang, M. Tao, and M. Abolhasan, "Graph theory and its applications to future network planning: Software-defined online small cell management," *IEEE Wireless Commun.*, vol. 22, no. 1, pp. 52–60, Feb. 2015.
[59] C. Yang, J. Li, M. Guizani, A. Anpalagan, and M. Elkashlan, "Advanced spectrum sharing in 5G cognitive heterogeneous networks," *IEEE Wireless Commun.*, vol. 23, no. 2, pp. 94–101, Apr. 2016.
[60] *Network Architecture (Release 12)*, document 3GPP TS 23.002 V12.6.0, Technical Specification Services and System Aspects, 2014.
[61] P. Mach, Z. Becvar, and T. Vanek, "In-band device-to-device communication in OFDMA cellular networks: A survey and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1885–1922, 4th Quart., 2015.
[62] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, Nov. 2014.
[63] J. Liu, N. Kato, J. Ma, and N. Kadowaki, "Device-to-device communication in LTE-advanced networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1923–1940, 4th Quart., 2015.
[64] G. A. Safdar, M. Ur-Rehman, M. Muhammad, M. A. Imran, and R. Tafazolli, "Interference mitigation in D2D communication underlaying LTE-A network," *IEEE Access*, vol. 4, pp. 7967–7987, 2016.
[65] S. Sun, Q. Gao, W. Chen, R. Zhao, and Y. Peng, "Recent progress of long-term evolution device-to-device in third-generation partnership project standardisation," *IET Commun.*, vol. 9, no. 3, pp. 412–420, 2015.
[66] H. A. U. Mustafa, M. A. Imran, M. Z. Shakir, A. Imran, and R. Tafazolli, "Separation framework: An enabler for cooperative and D2D communication for future 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 419–445, 1st Quart., 2016.
[67] G. Fodor *et al.*, "An overview of device-to-device communications technology components in METIS," *IEEE Access*, vol. 4, pp. 3288–3299, 2016.
[68] G. Fodor *et al.*, "Design aspects of network assisted device-to-device communications," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 170–177, Mar. 2012.
[69] M. Sheng *et al.*, "On-demand scheduling: Achieving QoS differentiation for D2D communications," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 162–170, Jul. 2015.
[70] B. Kaufman and B. Aazhang, "Cellular networks with an overlaid device to device network," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Oct. 2008, pp. 1537–1541.
[71] T. Peng, Q. Lu, H. Wang, S. Xu, and W. Wang, "Interference avoidance mechanisms in the hybrid cellular and device-to-device systems," in *Proc. IEEE 20th Int. Symp. Pers., Indoor Mobile Radio Commun.*, Sep. 2009, pp. 617–621.
[72] R. Zhang, X. Cheng, L. Yang, and B. Jiao, "Interference-aware graph based resource sharing for device-to-device communications underlaying cellular networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2013, pp. 140–145.
[73] L. Lei, Y. Kuang, X. Shen, C. Lin, and Z. Zhong, "Resource control in network assisted device-to-device communications: Solutions and challenges," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 108–117, Jun. 2014.
[74] S. Mumtaz, K. M. S. Huq, M. I. Ashraf, J. Rodriguez, V. Monteiro, and C. Politis, "Cognitive vehicular communication for 5G," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 109–117, Jul. 2015.
[75] R. Wonil *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
[76] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
[77] J. Lun and D. Grace, "Cognitive green backhaul deployments for future 5G networks," in *Proc. 1st Int. Workshop Cognit. Cellular Syst. (CCS)*, Sep. 2014, pp. 1–5.
[78] K. Chandra, R. V. Prasad, B. Quang, and I. G. M. M. Niemegeers, "CogCell: Cognitive interplay between 60 GHz picocells and 2.4/5 GHz hotspots in the 5G era," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 118–125, Jul. 2015.
[79] C.-F. Liu, S. Samarakoon, and M. Bennis, "Fronthaul-aware software-defined joint resource allocation and user scheduling for 5G networks," in *Proc. IEEE Globecom Workshop (GC Worksh)*, Washington, DC, USA, Dec. 2016, pp. 1–6.
[80] D. A. Wassie, G. Berardinelli, D. Catania, F. M. L. Tavares, T. B. Sorensen, and P. Mogensen, "Experimental evaluation of interference suppression receivers and rank adaptation in 5G small cells," in *Proc. IEEE 82nd Veh. Technol. Conf. (VTC-Fall)*, Boston, MA, USA, Sep. 2015, pp. 1–5.
[81] S. Mhaske, D. Uliana, H. Kee, T. Ly, A. Aziz, and P. Spasojevic, "A 2.48 Gb/s FPGA-based QC-LDPC decoder: An algorithmic compiler implementation," in *Proc. 36th IEEE Sarnoff Symp. (SARNOF)*, Newark, NJ, USA, Sep. 2015, pp. 88–93.
[82] P. Deng and M. Kavehrad, "Real-time software-defined single-carrier QAM MIMO visible light communication system," in *Proc. Integr. Commun. Navigat. Surveill. (ICNS)*, Apr. 2016, pp. 5A3-1–5A3-11.
[83] M. Danneberg *et al.*, "Implementation of a 2 by 2 MIMO-GFDM transceiver for robust 5G networks," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Brussels, Belgium, Aug. 2015, pp. 236–240.
[84] B. Song, Z. Wei, G. Li, X. Zhang, Y. Liu, and D. Yang, "Energy and spectral efficiency analysis of millimeter wave MIMO system based on known-position beamforming," in *Proc. IEEE/ CIC Int. Conf. Commun. China (ICCC)*, Qingdao, China, Oct. 2017, pp. 1–6.
[85] G. Berardinelli *et al.*, "Distributed synchronization of a testbed network with USRP N200 radio boards," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput. (ACSSC)*, Pacific Grove, CA, USA, Nov. 2014, pp. 563–567.
[86] W. Khawaja, O. Ozdemir, and I. Guvenc, "UAV air-to-ground channel characterization for mmWave systems," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Toronto, ON, Canada, Sep. 2017, pp. 1–5.
[87] R. Nilsson and J. van de Beek, "Channel measurements in an open-pit mine using USRPs: 5G—Expect the unexpected," in *Proc. Workshop Commun. Extreme Conditions (ComExCon)*, Doha, Qadar, Apr. 2016, pp. 1–6.
[88] J. J. H. Almeida, P. B. Lopes, and C. Akamine, "A proposal for the next generation of ISDB-TB using FBMC in a SDR implementation on GNU radio environment," *IEEE Latin Amer. Trans.*, vol. 15, no. 7, pp. 1231–1237, 2017.
[89] P. Huang, X. Yang, and L. Xiao, "Dynamic channel bonding: Enabling flexible spectrum aggregation," *IEEE Trans. Mobile Comput.*, vol. 15, no. 12, pp. 3042–3056, Dec. 2016.

[90] N. Gao, X. Jing, S. Lv, J. Mu, and L. Sun, "Wireless physical layer characteristics based random number generator: Hijack attackers," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Toronto, ON, Canada, Sep. 2017, pp. 1–5.

[91] D.-W. Park, G.-H. Gwag, H.-J. Oh, I.-M. Park, and Y.-S. Eo, "28 GHz RF transceiver module for 5G beam-forming system," in *Proc. Asia–Pacific Microw. Conf. (APMC)*, New Delhi, India, May 2017, pp. 1–4.

[92] M. Yu, "The blind separation of wireless communication signals utilizing USRP," in *Proc. 5th Int. Conf. Comput. Sci. Netw. Secur. (ICCSNT)*, Changchun, China, Dec. 2016, pp. 677–680.

[93] D. A. Wassie, G. Berardinelli, F. M. L. Tavares, T. B. Sorensen, and P. Mogensen, "Experimental verification of interference mitigation techniques for 5G small cells," in *Proc. IEEE 82nd Veh. Technol. Conf. (VTC-Fall)*, Boston, MA, USA, Sep. 2015, pp. 1–5.

[94] X. Xiong, T. Wu, H. Long, and K. Zheng, "Implementation and performance evaluation of LECIM for 5G M2M applications with SDR," in *Proc. IEEE Globecom Workshop (GC Worksh)*, Austin, TX, USA, Dec. 2014, pp. 612–617.

[95] *IEEE Standard for Local and Metropolitan Area Networks—Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs)—Amendment 5: Physical Layer Specifications for Low Energy, Critical Infrastructure Monitoring Networks*, IEEE Standard 802.15.4k, 2013.

[96] M. H. M. Elsayed and A. Mohamed, "Distributed interference management using Q-learning in cognitive femtocell networks: New USRP-based implementation," in *Proc. New Technol., Mobility Secur. (NTMS)*, Paris, France, Jul. 2015, pp. 1–5.

[97] P. P. Brahma, D. Wu, and Y. She, "Why deep learning works: A manifold disentanglement perspective," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 10, pp. 1997–2008, Oct. 2016.

[98] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Predictive and core-network efficient RRC signalling for active state handover in RANs with control/data separation," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1423–1436, Mar. 2017.

KOK-LIM ALVIN YAU (M'08) received the B.Eng. degree (Hons.) in electrical and electronics engineering from Universiti Teknologi Petronas, Malaysia, in 2005, the M.Sc. degree in electrical engineering from the National University of Singapore in 2007, and the Ph.D. degree in network engineering from the Victoria University of Wellington, New Zealand, in 2010.

He is currently an Associate Professor with the Department of Computing and Information Systems, Sunway University. He is also a Researcher, a Lecturer, and a Consultant in cognitive radio, wireless networks applied artificial intelligence, and reinforcement learning. He serves as a TPC member and a reviewer for major international conferences, including ICC, VTC, LCN, GLOBECOM, and AINA. He was a recipient of the 2007 Professional Engineer Board of Singapore Gold Medal for being the best graduate of the M.Sc. degree in 2006/2007.

Dr. Yau serves as an Associate Editor for the IEEE Access, an Editor of the *KSII Transactions on Internet and Information Systems*, a Guest Editor of the Special Issues of IEEE Access, *IET Networks*, *IEEE Computational Intelligence Magazine*, and the *Springer Journal of Ambient Intelligence and Humanized Computing*, and a regular reviewer for over 20 journals, including the IEEE journals and magazines, the *Ad Hoc Networks*, the *IET Communications*, and others. He also served as the General Co-Chair of the IET ICFCNA'14 and the Co-Chair of the Organizing Committee of the IET ICWCA'12.

JUNAID QADIR (SM'14) received the B.S. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, and the Ph.D. degree from the University of New South Wales, Australia, in 2008.

He was an Assistant Professor with the School of Electrical Engineering and Computer Sciences (SEECS), National University of Sciences and Technology, Pakistan. At SEECS, he directed the Cognet Laboratory that focused on cognitive networking and the application of computational intelligence techniques in networking. He is currently an Associate Professor with the Information Technology University (ITU),Punjab, Lahore. He is also the Director of the IHSAN Laboratory, ITU, that focuses on deploying ICT for development, and is engaged in systems and networking research. His research interests include the application of algorithmic, machine learning, and optimization techniques in networks. In particular, he is interested in the broad areas of wireless networks, cognitive networking, software-defined networks, and cloud computing.

Dr. Qadir was a recipient of the Highest National Teaching Award in Pakistan—the higher education commission's Best University Teacher Award— from 2012 to 2013. He has been nominated for this award twice (2011, and 2012–2013). He is a regular reviewer for a number of journals and has served in the program committee of a number of international conferences. He serves as an Associate Editor for the IEEE Access, the *IEEE Communication Magazine*, and *Nature Big Data Analytics* (Springer). He was the Lead Guest Editor for the special issue Artificial Intelligence Enabled Networking in IEEE Access and the feature topic Wireless Technologies for Development in the *IEEE Communications Magazine*. He is a member of ACM.

CELIMUGE WU (M'10) received the M.E. degree from the Beijing Institute of Technology, China, in 2006, and the Ph.D. degree from The University of Electro-Communications, Japan, in 2010. He is currently an Associate Professor with the Graduate School of Informatics and Engineering, The University of Electro-Communications. His current research interests include vehicular ad hoc networks, sensor networks, intelligent transport systems, IoT, 5G, and mobile cloud computing. He has been a track co-chair or workshops co-chair for a number of international conferences, including IEEE PIMRC 2016, IEEE CCNC 2017, ISNCC 2017, and WICON 2016.

MUHAMMAD ALI IMRAN (SM'12) received the M.Sc. degree (Hons.) and the Ph.D. degree from Imperial College London, U.K., in 2002 and 2007, respectively. He has over 18 years of combined academic and industry experience, working primarily in the research areas of cellular communication systems. He is currently an Affiliate Professor with The University of Oklahoma, USA, and a Visiting Professor with the 5G Innovation Centre, University of Surrey, U.K. He is also the Vice Dean with the Glasgow College, UESTC, and a Professor of communication systems with the School of Engineering, University of Glasgow. He has 15 patents, authored/co-authored over 300 journal and conference publications, and has been principal/co-principal investigator on over č6 million in sponsored research grants and contracts. He has supervised over 30 successful Ph.D. graduates.

He has an award of excellence in recognition of his academic achievements, conferred by the President of Pakistan. He was a recipient of the IEEE Comsoc's Fred Ellersick Award 2014, the FEPS Learning and Teaching Award 2014, and the Sentinel of Science Award 2016. He was twice nominated for Tony Jean's Inspirational Teaching Award. He was a shortlisted finalist for The Wharton-QS Stars Awards 2014, the QS Stars Reimagine Education Award 2016 for innovative teaching and VC's Learning and Teaching Award in the University of Surrey. He is a Senior Fellow of the Higher Education Academy, U.K.

MEE HONG LING received the bachelor's degree (Hons.) in computer and mathematical studies from Oxford Brookes University, U.K., and the master's degree in data engineering (computer science) from Keele University, U.K. She is currently pursuing the Ph.D. degree in computing. She lectures with the Department of Computing and Information Systems, Sunway University. Her research interests are in the areas of security, cognitive radio networks, and applied reinforcement learning.