Control-Data Separation Architecture for Cellular Radio Access Networks: A Survey and Outlook

Abdelrahim Mohamed, *Student Member, IEEE*, Oluwakayode Onireti, *Member, IEEE*, Muhammad Ali Imran, *Senior Member, IEEE*, Ali Imran, *Member, IEEE*, and Rahim Tafazolli, *Senior Member, IEEE*

***Abstract*—Conventional cellular systems are designed to ensure ubiquitous coverage with an always present wireless channel irrespective of the spatial and temporal demand of service. This approach raises several problems due to the tight coupling between network and data access points, as well as the paradigm shift towards data-oriented services, heterogeneous deployments and network densification. A logical separation between control and data planes is seen as a promising solution that could overcome these issues, by providing data services under the umbrella of a coverage layer. This article presents a holistic survey of existing literature on the control-data separation architecture (CDSA) for cellular radio access networks. As a starting point, we discuss the fundamentals, concept and general structure of the CDSA. Then, we point out limitations of the conventional architecture in futuristic deployment scenarios. In addition, we present and critically discuss the work that has been done to investigate potential benefits of the CDSA, as well as its technical challenges and enabling technologies. Finally, an overview of standardisation proposals related to this research vision is provided.**

***Index Terms*—5G cellular systems; context awareness; control data separation; dual connectivity; energy efficiency; network densification; radio access networks.**

I. Introduction

Nowadays, requirements and performance bounds of fifth generation (5G) cellular systems are becoming of increasing interest in academia and industry fora. According to recent forecasts and worldwide discussions, an incremental advancement of current systems, such as the long term evolution (LTE), may not be sufficient to satisfy the ambitious targets being identified for the 2020 era [1]–[3]. The exponentially increasing traffic demand, heterogeneity of radio access networks (RANs) and new use cases call for the design of efficient, sustainable, scalable, flexible and versatile cellular systems. These requirements are driven by the anticipated capacity and performance targets that need to meet diverse application requirements under cost and energy constraints. This calls for network densification, a short-length wireless link, efficient and minimal control signalling and the ability to switch off the power consuming devices when they are not in use. In this direction, the conventional architecture raises several problems from energy, planning, interference and mobility perspectives.

This work was made possible by NPRP grant No.5-1047-2437 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

A. Mohamed, O. Onireti, M. Imran and R. Tafazolli are with Institute for Communications Systems (ICS), University of Surrey, Guildford, UK. e-mail:{abdelrahim.mohamed; o.s.onireti; m.imran; r.tafazolli} @ surrey.ac.uk

A. Imran is with the School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, USA. e-mail: ali.imran @ ou.edu

In research community, a new RAN architecture with a logical separation between control plane (CP) and data plane (DP) has been proposed. The key concept behind this approach is to separate the signals required for full coverage from those needed to support high data rate transmission. A few macro cells (MCs), also known as control base stations (CBSs), provide the coverage and support efficient radio resource control (RRC) procedures, while dedicated small cells (SCs), known as data base stations (DBSs), provide high rate data transmission within the CBS footprint. In this paper, we provide a survey of existing literature that investigates applications of the control-data separation architecture (CDSA). We identify several areas where the CDSA can overcome limitations of the conventional architecture. In addition, we discuss the technical challenges imposed by the CDSA and survey candidate solutions and enabling technologies. Furthermore, we present some of the ideas already under discussion in standardisation forums.

The remainder of this paper is structured as follows: Section II provides a holistic view of the CDSA concept and basic operation as well as the aspects being considered in international projects. Section III discusses energy efficiency (EE) of cellular systems and surveys energy saving techniques in both the conventional and the separation architectures. Section IV focuses on capacity dimension and discusses superiority of the CDSA over the conventional architecture in dense deployment scenarios. In Section V, we discuss the CDSA benefits related to interference, resource and mobility management by using a centralised CP. Section VI focuses on signalling overhead and identifies techniques to minimise the signalling load. Section VII discusses some of the challenges imposed by the CP/DP separation along with possible solutions and enabling technologies. Section VIII provides an overview of preliminary standardisation proposals related to the CDSA, while Section IX concludes the paper and underlines potential research directions. Table I provides a list for the acronyms used in the paper.

II. CDSA Concept and General Structure

A. Motivation and Basic Operation

The main idea of the CDSA originates from the fact that only a small amount of signalling is required to enable ubiquitous coverage [4]. On the other hand, data transmission

TABLE I: List of Acronyms

Acronym	Full form
3GPP	Third Generation Partnership Project
5G	Fifth Generation
BPSK	Bipolar Phase Shift Keying
BS	Base Station
BW	Bandwidth
CBS	Control Base Station
CDSA	Control-Data Separation Architecture
CoMP	Coordinated MultiPoint
CP	Control Plane
CRS	Cell-specific Reference Signal
CSI-RS	Channel State Information Reference Signal
DBS	Data Base Station
DL	Downlink
DP	Data Plane
DRB	Data Radio Bearer
EC	Energy Consumption
EE	Energy Efficiency
eICIC	enhanced Inter-Cell Interference Coordination
FDD	Frequency Division Duplex
GPS	Global Positioning System
GSM	Global System for Mobile communications
HetNet	Heterogeneous Network
HO	Handover
HOF	Handover Failure
ICI	Inter-Cell Interference
LTE	Long Term Evolution
M2M	Machine-to-Machine
MC	Macro Cell
MIMO	Multiple Input Multiple Output
mm-wave	millimetre wave
OFDM	Orthogonal Frequency Division Multiplexing
OTA	Over The Air
PA	Power Amplifier
PCC	Phantom Cell Concept
QPSK	Quadrature Phase Shift Keying
RAN	Radio Access Network
RF	Radio Frequency
RLF	Radio Link Failure
RME	Resource Management Entity
RRC	Radio Resource Control
RSS	Received Signal Strength
S-GW	Serving Gateway
SC	Small Cell
SDN	Software-Defined Networking
SE	Spectral Efficiency
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SON	Self-Organising Network
TDD	Time Division Duplex
TDM	Time Division Multiplexing
UE	User Equipment
UE-RS	User Equipment specific Reference Signal
UL	Uplink

and its related signalling are needed on demand when there are active user equipment (UE). This calls for a two layer RAN architecture with a logical separation between:

- Network access and data transmission functionalities.
- Idle mode and active mode.
- Cell-specific/broadcast-type and UE-specific/unicast-type signalling.

In the CDSA, a continuous and reliable coverage layer is provided by CBSs at low frequency bands, where the large footprint ensures robust connectivity and mobility. The DP is supported by flexible, adaptive, high capacity and energy efficient DBSs that provide data transmission along with the necessary signalling. As shown conceptually in Fig. 1, all UEs are anchored to the CBS, while active UEs are associated with both the CBS and the DBS in a dual connection mode [5]. With this configuration, the DBS is invisible to both detached and idle UEs. Expressed differently, idle UEs are connected with the CBS only. Thus the DBS carrier can be switched off as long as it is not needed. When the UE becomes active, e.g., starting a data session or receiving a call, the CBS selects the best serving DBS and establishes a high rate DBS-UE connection through backhaul links. This approach comes with a range of benefits which are discussed in Sections III–VI.

Specifying functionalities of each plane is not trivial due to the fact that several functionalities may be needed to support a certain UE activity (e.g., cell reselection requires synchronisation and broadcast functionalities). In addition, a certain signal may be required by more than one network functionality such as the pilot signal, which is needed for synchronisation, paging, etc. [6]. In cellular domain, few separation schemes have been proposed to separate the CP from the DP. Based on network functionalities and a functionality-signal mapping, [6] proposed a separation scheme for the LTE by separating the functionalities required to support connectivity from those needed for data transmission. According to this scheme, the CBS supports synchronisation, broadcast, multicast, paging and RRC functionalities. On the other hand, the DBS supports unicast data transmission and synchronisation functionalities only.

A Similar approach has been followed in [7] for the global system for mobile communication (GSM). The authors of [8] argue that the control and the data channels are logically separated in current standards but they are mixed at the final stage to be transmitted by the same physical node. Thus, [8] proposed a separation scheme for LTE-Railway by mapping all logical control channels to a single physical channel that is transmitted by the CBS, while all logical traffic channels are mapped to a single physical traffic channel handled by the DBS. Table II maps network functionalities, while Table III maps UE states and shows frame allocations of both the CBS and the DBS according to [6], [7] and [8].

It is worth mentioning that the CDSA is a new concept in cellular domain although it has been proposed earlier for other systems, such as sensor networks [9], [10]. Thus, its operation and implementation aspects are currently being studied in several research projects. These include:

- **Beyond Cellular Green Generation (BCG^2)**
  This is a project of the GreenTouch Consortium with a primary target of improving EE of cellular systems. It focuses on benefits of the CDSA from an energy perspective and proposes a cell on-demand approach. In the latter, the DBSs are switched on and off according to traffic variations without affecting the basic connectivity service provided by the CBS. However, such an operation raises several challenges as discussed in Section VII. In particular, BCG^2 tackles the problems of context information detection, serving node selection and management of interaction between the CP and the DP.
- **Toward Green 5G Mobile Networks (5grEEn)**
  As with BCG^2, 5grEEn [11] focuses on designing green 5G cellular networks with a logical separation between

TABLE II: Functionality mapping in CDSA

Functionality	**CBS**	**DBS**	**Reason/Benifit**
Cell search	✓		Network access and connectivity are provided by the umbrella cell (i.e., CBS) only. DBSs can be switched off
System information	✓		
Paging	✓		
Multicast and broadcast	✓		Low-rate/broadcast-type services are supported by the CBS only to maximise the DBS transmission resources and sleep periods
Radio Resource Control	✓		UE is anchored to a cell with a large coverage area, which reduces handover overhead and provides robust mobility performance
Mobility management	✓		
Serving node (DBS) selection	✓		Optimised network driven UE-DBS association based on a wide view of network status
Unicast data transmission	⋆	✓	High data rate services are provided by the DBS in a one-to-one fashion
Link adaptation		✓	These functionalities support data transmission and they require fast adaptation/response, thus they are kept at the DBS
Beam-forming		✓	

⋆ When the CBS provides low rate services, e.g., voice or data transmission to high speed users

TABLE III: UE State and Frame Structure mapping in CDSA

UE state	**CBS**	**DBS**	**Reason/Benifit**
Detached/Idle	✓		Idle UEs maintain a single connection with the CBS only as long as they do not require data transmission. DBSs can be switched off
Active	✓	✓	Active UEs maintain a dual connection with both the CBS and the DBS. CBS: RRC and system information. DBS: high data rate transmission
Signal	**CBS**	**DBS**	**Reason/Benifit**
Synchronisation	✓	✓	Active UEs need to synchronise with both carriers
Pilot	✓	✓	CBS and DBS could have different characteristics such as power, location etc. Thus both frames need to contain pilot signal for channel estimation
Frame control	✓	✓	To specify the allocations within each frame
Paging	✓		These services are provided by the CBS only. Thus the DBS frame does not need allocations for paging, broadcast and multicast bearer signals
Broadcast bearer	✓		
Multicast bearer	✓		
Unicast bearer	⋆	✓	Most of the DBS transmission resources are allocated to the unicast bearer signal

⋆ When the CBS provides low rate services, e.g., voice or data transmission to high speed users

idle mode functions and data transmission services (i.e., CP/DP separation). It investigates the usage of massive reconfigurable antennas with dynamic cell structuring to optimally reshape the DP coverage. Such techniques adapt the network to traffic variations and allow increasing the inter-site distance, thus reducing energy consumption (EC) and improving overall efficiency of the network. In addition, 5grEEn investigates the impact of the CDSA on network deployment strategies and possible backhauling solutions.

- **Millimetre-Wave Evolution for Backhaul and Access (MiWEBA)**
  This is a joint European Japanese research project with a primary target of extending cellular systems capacity by exploiting the millimetre wave (mm-wave) band. MiWEBA integrates mm-wave SCs into conventional cellular systems, and utilises the CDSA to overcome coverage restrictions of the mm-wave link. The network architecture consists of MCs placed on rooftops to provide the basic connectivity service at conventional cellular bands. Data services are provided by mm-wave SCs that are deployed within the MC footprint [12]. Depending on the deployment scenario, the MiWEBA project investigates whether the CP and the DP should be logically and physically separated (i.e., provided by separate physical nodes) or whether it is more feasible to adopt a logical separation only (i.e., control and data interfaces are hosted in the same node). Several key performance indicators are considered in analysing this trade-off such as data channel acquisition delay, data session retainability, EC and signalling overhead [12].
- **Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS)**
  The FP7 research project METIS defines, investigates, characterises and models a potential 5G air interface and considers the CDSA as a candidate interference management technique. Targeting a minimal inter-cell interference (ICI), METIS exploits the wider view of the CBS that controls power and resource allocations of the DBSs under its control, by using centralised interference-aware scheduling mechanisms [13]. Contextual information, such as position and movement history, are used for mobility prediction and handover (HO) optimisation, while signal-to-noise (SNR) databases are used for channel quality prediction. In addition, METIS investigates the usage of carrier aggregation to enable a seamless implementation of the CDSA in current standards [13]. Fig. 1 shows a high level diagram of the CDSA implementation aspects and potential benefits that are investigated in these projects.

B. CDSA and Software-Defined Networking

Software-defined networking (SDN) is an emerging concept that decouples the CP and the DP by separating control decision entities from control action enforcement elements. Although the basic idea of the SDN sounds similar to the cellular CDSA, these two concepts should not be confused with each other. In SDN, CP means the decision makers that determine where and how the traffic should be sent, while DP refers to the system that forwards the packets according to

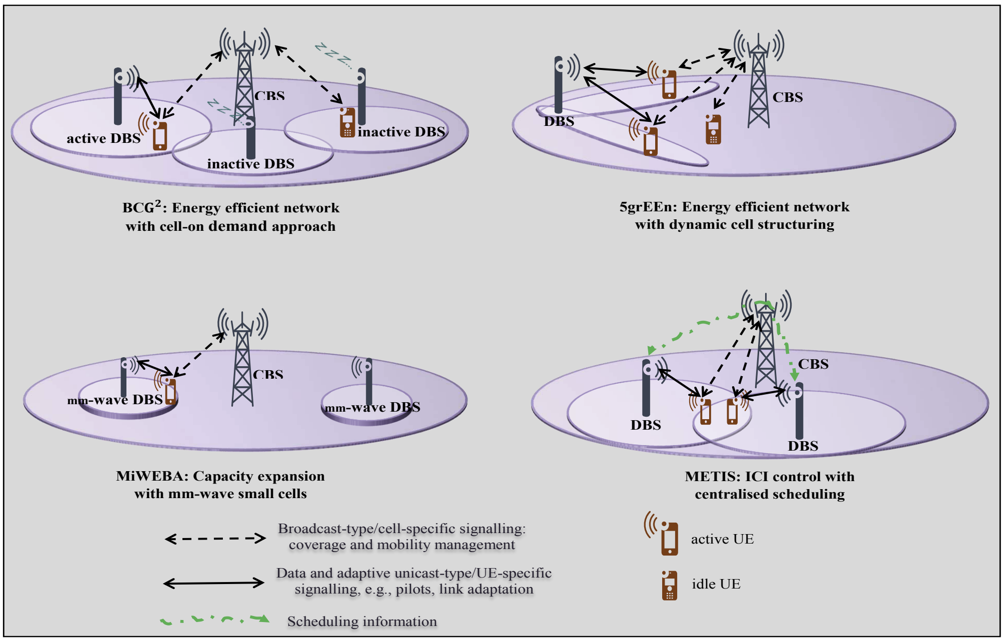

Fig. 1: Perspectives of the CDSA being studied in international projects

the decision taken by the CP. SDN allows cellular networks to be flexible and reconfigurable and it simplifies network management procedures [14]. This is realised by moving the CP to a software application often called controller, resulting into a programmable network. In cellular CDSA, the CP includes decision making entities in addition to (most of) the network-UE signalling that is related to the service being requested/provided by/to the UE. This signalling includes RRC connection establishment and maintenance commands, scheduling information, etc. without which data transmission and seamless connectivity cannot be supported.

As discussed in Section V, the CDSA allows a paradigm shift towards almost centralised control functionalities. Aligned with this trend, the SDN suggests a centralised controller to enable control decisions to be taken based on a wider view of network status and parameters [14]. In other words, the SDN and the CDSA have the same physical realisation: moving towards a centralised CP [15]. SDN and CDSA are closely related concepts in the sense. In both architectures, intelligence is partially or completely removed from most of the nodes in the network to be concentrated in fewer central nodes. This results in cost saving, higher performance and resource efficiency. SDN is manifestation of the above idea in core network, whereas CDSA implements the same idea in RAN. A comparison between the SDN and the CDSA concepts is provided in Table IV. Recent studies have proposed integrating the SDN and the CDSA, [16] refers to such integration as Soft-RAN. In the latter, the SDN concept is adopted to abstract all BSs as a virtual big BS (analogous to the CBS) that hosts a centralised CP for radio elements (analogues to the DBSs). Following a similar approach, [17] proposes a programmable 5G CP where connectivity is provided as a service application running in the controller. In addition, the SDN/CDSA integration has been investigated in [18], where a two layer 5G network architecture has been proposed.

A similar architecture has been proposed in the FP7 CROWD[1] project by combining the CDSA with the SDN. It follows the classical SDN approach of using a centralised controller whilst reducing signalling overhead by terminating some of the control information in local controllers, resulting in a hybrid centralised/distributed control functionalities [19]. The CP is implemented in a software application handled by the local controllers that are hosted in RAN elements and they are used for fast and fine grained control functionalities. Several local controllers are connected to a regional controller hosted in a data centre, which is used for slower, long time scale control operations [20]. The CDSA has been adopted by directing control path of the LTE to the local controllers, while the data path goes to a distributed mobility management entity gateway that provides local mobility support. The reader is referred to [19] and [20] for detailed description of the CROWD architecture.

The authors of [15] integrate the SDN concept with the BCG^2 architecture, and they argue that the CDSA requires redesigning current network hardware components. Thus the SDN is seen as an enabling technology that could allow a feasible and cost-efficient CDSA implementation. In addition, the SDN offers a technology-agnostic CP by allowing the control decisions and commands to be taken at a technology-agnostic level of abstraction [19]. This feature is of great importance when the CBS manages DPs of several operators in infrastructure and/or spectrum sharing scenarios. Furthermore, the SDN enables the CDSA applications related to network-driven resource selection. This can be done by implementing an application that collects information on network status and UE context, and then executes optimisation functions to dynamically associate the UEs with the best serving DBS. The

[1] Connectivity management for eneRgy Optimised Wireless Dense networks.

TABLE IV: Comparison between SDN and CDSA

Comparison criteria	SDN	CDSA
Scope	Core Network	RAN
Network elements	routers	DBS
Central units	Central traffic controllers	CBS
Unique advantages	Software upgrade, technology agonistic, softly defined capacity	Mobility robustness, easy interference management
Unique challenges	Delay	Division of functionality and signalling is not trivial, backhaul/fronthaul networks, new framing structure
Common advantages	Energy saving, cost saving, efficient resource management	
Common challenges	Single point of failure	

optimisation function can have an objective of increasing the EE (e.g., associate the UEs with a small subset of the DBSs and switch off other DBSs), balancing the network load (e.g., offload some UEs from a congested to a low utilised DBS) or to alleviate mobility overhead [16], [19]. These aspects are discussed in details in the following sections.

III. Energy Efficiency

Current cellular systems have been developed and evolved with a primary focus on performance improvement and mobility support without considering energy aspects [21]. Nowadays, the information and communication technology sector contributes 3% to the global EC and generates 2% of the worldwide CO_2 emissions [22], with recent forecasts for doubling this contribution every five years [23]. Thus, neglecting the energy dimension in designing 5G cellular systems will cause them to encounter several environmental and economical problems. In wireless systems, most of the energy is consumed by radio interface components. Precisely, more than 80% of the access network power is consumed by base stations (BSs) in cellular systems [24]. As a result, minimising EC of the access network is the best way to conform to the general trend of sustainable and green communications, as well as to cut the energy bill.

Conventional cellular systems consume high power even in low traffic situations due to the “always-on” service approach adopted in these systems. The results of the EARTH[2] project reported in [25] show that EC of the LTE is almost insensitive to traffic load and is dominated by unnecessary overhead transmission and idle mode signalling, see for example a typical power profile of pico BSs in Fig. 2. The most power consuming component in the BS i.e., the power amplifier (PA) [25], can also be considered as one of the contributors to this load-independent energy profile. Typical PAs operate with a high input power irrespective of the actual traffic load [26]. Although using sophisticated PAs improves the EC profile, such component level optimisation does not overcome the baseline power consumed by active PAs [27].

This EC profile can be justified in high traffic scenarios. However, today’s cellular networks operate in a low load regime. Currently, the average BS utilisation is less than 10% for 45% of the time [28] and [29] estimates that up to 97% of orthogonal frequency division multiplexing (OFDM) wireless resources are not used, with 50% of the traffic being carried by 15% of the deployed BSs [30]. Given the load-insensitive EC profile, it can be concluded that EE of current cellular systems is generally poor.

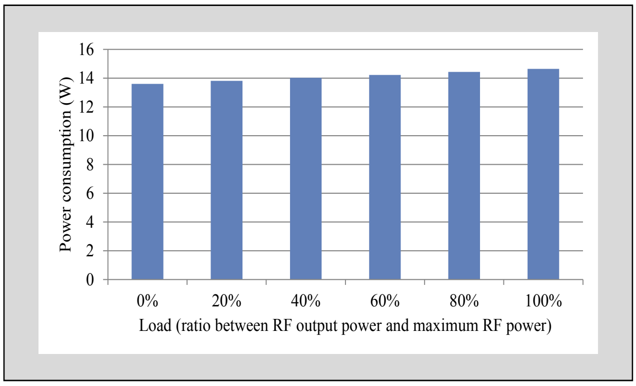

Fig. 2: Typical power consumption of pico BS in LTE system with 10 MHz bandwidth and 2×2 MIMO configuration, based on the EARTH power model [25]

A. Conventional Energy Saving Techniques

Minimising the EC requires exploiting the spare capacity by adapting the network to the actual traffic load. In this context, several energy saving techniques, such as discontinuous transmission, multiple input multiple output (MIMO) muting and cell wilting and blossoming, have been proposed.

- Discontinuous transmission: provides energy saving in time domain by switching off some of the BS components such as the PA during the unoccupied subframes[3]. However, the mandatory transmission of cell-specific reference signals (CRSs) limits the sleep periods in this technique [31].
- MIMO muting: provides energy saving in spatial domain by reducing the number of active antennas [31]. This technique is of great importance since MIMO systems increase the EC significantly due to the large number of PAs as well as the complex processing for multiplexing and diversity gains. Nonetheless, MIMO muting requires fast adaptation to satisfy coverage and performance requirements [32].
- Cell wilting and blossoming: This technique exploits the fact that energy loss is proportional to propagation distance [33] by adapting the cell size to traffic profile. Cell wilting lessens the pilot power in off-peak periods

[2] Energy Aware Radio and neTwork tecHnologies. http://www.ict-earth.eu

[3] In lTE, the radio frame consists of ten subframes. Each subframe is divided into two time slots of 0.5 ms each.

to allow soft reduction of the coverage. When the traffic demand rises, cell blossoming increases the pilot power [34]. However, cell size adaptation may not be timely enough to respond to rapid variations in data traffic. In addition, it requires careful HO management because the reduced overlap area may increase the call drop rates [31].

B. Coverage Restrictions

Achieving a breakthrough in energy saving requires a paradigm shift towards on-demand systems by switching off a subset of the BSs during off-peak periods [35]. Although such wide network adaptation will result in order of magnitude saving in energy, it may not be feasible with the conventional cellular architecture due the tight coupling between coverage and data services. In other words, coverage constraints are considered as the main source of energy inefficiency. In this regard, several techniques have been proposed to preserve the coverage; power control techniques, such as cell zooming, can be employed to increase the power of some BSs when other BSs are switched off [36]. Despite its potential gains, this technique may not guarantee full coverage and provides poor performance to cell edge users due to the increased ICI between active BSs with an extended coverage [36]. Recently, suboptimal sleep mode mechanisms have been proposed for SCs in the third generation partnership project (3GPP). In these mechanisms, either the radio frequency (RF) receiver chain of the BS has to be kept on to receive signalling to switch on the BS, or the RF transmitting chain has to be turned on periodically to transmit beacon signals [37], [38].

Multi-hop relay has been proposed in [39] to allow other terminals to relay the traffic of UEs in the vicinity of a switched off BS. Although this technique does not increase the ICI, finding suitable relays is a challenging task [39], and the received signal at the UE can be very poor depending on location and capabilities of the relays. BSs cooperation techniques, such as coordinated multi-point (CoMP), can also be used to provide coverage to UEs when their nearest BS is switched off. Although the joint transmission of several BSs boosts the received signal at the UE, this technique guarantees neither performance nor full coverage for all affected users. Table V summarises the conventional energy saving techniques and highlights their limitations. Based on this discussion, it can be concluded that the conventional cellular architecture where basic coverage functionalities and data transmission services are provided by the same physical node offers limited opportunities for energy saving. In addition, most of the standardised/proposed techniques are limited by the coverage constraints as well as the mandatory transmission of CRSs.

C. Energy Saving in CDSA

Separating the CP from the DP allows flexible adaptation opportunities without breaking the anywhere/anytime service concept. In the CDSA, the basic coverage is provided by a few CBSs, while data transmission is supported by DBSs as shown in Fig. 1. Hence, adapting the DBSs to traffic load does not affect the coverage provided by the CP. Expressed differently, the CDSA could allow a paradigm shift towards on-demand always-available systems that scale the EC with the traffic load whilst maintaining a full connectivity coverage.

Considering the EARTH 2020 traffic model, [40] shows that the flexible opportunities for DBS on/off operation achieve up to four times higher EE compared with legacy systems. Reference [6] incorporates the DBS sleep opportunities along with the reduction in control signalling, and shows that such an architecture can save up to one third of the energy in urban deployment scenarios whilst scaling the EC with the traffic load. The feasibility study reported in [41] indicates that the potential energy gains of the CDSA will be much higher in low utilisation and dense deployment scenarios. However, this study does not consider the facts that each DBS has a finite capacity and the instantaneous utilisation of the DBS affects its ability to serve other users.

In addition, decoupling the CP from the DP allows flexibility in reshaping the coverage of the DBS (i.e., cell restructuring) without affecting the underlay CP coverage. In contrast to the conventional architecture, the DBS does not transmit CRSs [6]. Thus the DBS can be considered as a UE-specific resource that dynamically transmits the data in directions towards the active UEs only. Considering this feature, [42] proposes a dynamic cell structuring mechanism by using large-scale CoMP. In this technique, a cluster of DBSs is dynamically created around hotspots by controlling beam directions of each DBS. Furthermore, cell wilting and blossoming can be easily realised in the DP with relaxed HO constraints when mobility management is delegated to the CBS. These flexible opportunities for power adjustment and beam-forming result into a high gain which can be translated into an increase in the link level EE [11].

It is well understood from a number of recent studies [43]–[45] that in conventional network operational point for EE and spectral efficiency (SE) are not the same. Network operator has to choose between the two key performance indicators while designing a network. One method to optimize this trade-off dynamically, while taking into account spatio-temporal variation of traffic demand, is to switch on and off the BS. However, conventional cellular networks are not designed for frequent switching on and off. Whereas CDSA, as explained above has all features needed to perform dynamic on and off switching with high agility. A very recent study in [46] has investigated the technical benefits of the CDSA in terms of both SE and EE. An interesting finding from [46] is summarized in plot shown in Fig. 3.

It can be seen from Fig. 3 that the deployment density of DBS that yields maximum EE and deployment density of DBS that yields maximum SE are not the same. This again reinforces the conclusions from EE vs SE studies on conventional networks [43]–[45]. However, unlike conventional architecture, where dynamically changing density of BS is not administratively as well as technically feasible, due to intrinsic decoupled design, in the CDSA effective density of the DBSs can be orchestrated in self-organising fashion with much better administrative and technical ease. This dynamic adaptation of effective DBS density can then allow to choose desired operational point between EE and SE by maintaining optimal effective DBS density, while taking into account

TABLE V: Conventional energy saving techniques and their limits

Technique	Function	Limits
DTX	Switches off the BS components during the empty subframes	Mandatory transmission of CRS limits the sleep period
MIMO Muting	Reduces the number of active antennas and PAs	Requires fast adaptation to satisfy coverage and performance requirements
Cell Wilting and Blossoming	Adapts the cell size to traffic demand (energy loss $\propto$ propagation distance)	Inapt for rapid variations in data traffic and requires careful HO management
Cell Zooming	Increases the power of some BSs when other BSs are switched off	Lack of full coverage and increase in ICI
Suboptimal sleep mode	Either RF receiver or transmitting chain has to be kept "ON", either to receive signal or to transmit beacon signal	
Multi-hop relay	Neighbouring terminals relay the traffic of UE when concern BS is switched off	Finding suitable relays is a challenging task and quality of the received signal can be very poor
CoMP	Joint BSs transmission to provide coverage to UEs when their nearest BS is switched off	Guarantees neither performance nor full coverage for all affected users

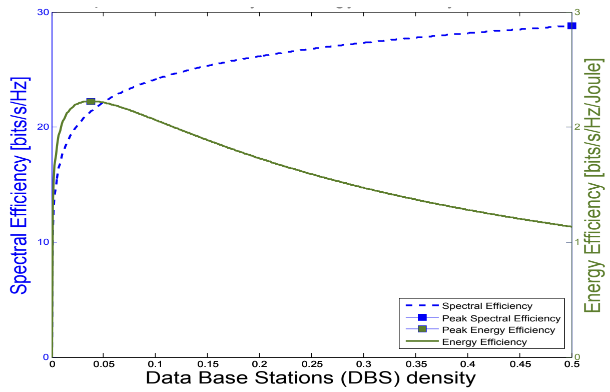

Fig. 3: Energy and spectral efficiencies vs DBS density [46]

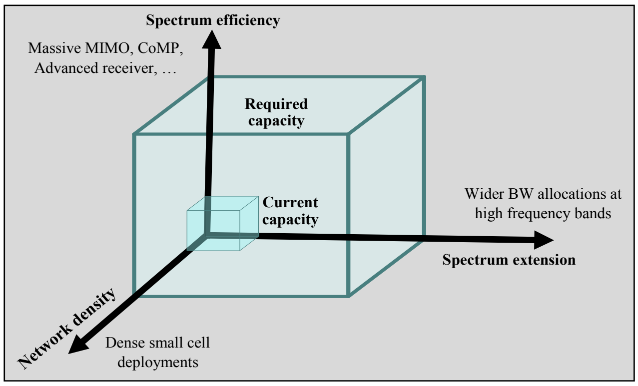

Fig. 4: Capacity evolution cube, based on [49] estimates

spatio-temporal traffic demand.

In addition to the flexibility of trading EE and SE, recent study in [47] has also shown that the CDSA can offer better SE mainly because of selection diversity that stems from large number of DBSs. Opportunities for centralized interference coordination, as discussed in Section V-A, is another feature of the CDSA that can yield better SE as compared to conventional networks. The high SE means the transmission will be done quickly which increases the quiet period of DBSs i.e., more time for the DBS to operate in sleep mode [11]. In addition, better EE in the CDSA mainly comes from low power DBSs, ability to switch off DBSs, lower propagation losses due to smaller distance between DBS and UE, and opportunity for centralized self-organising EE functions to switch off and on suboptimal used DBSs in conjunction with load balancing self-organising networks (SONs). Furthermore, the low-rate/long-range services provided by the CP allow using low order constant envelope modulation, such as BPSK and QPSK, in the CBS. Thus the PA of the CBS can operate at saturation without non-linearity problems, which improves the PA efficiency and hence the EE.

IV. System Capacity

The on-going trend towards Internet of Things and Machine-to-Machine (M2M) communications will increase the number of connected devices in 2020 by at least a factor of 10 compared with 2009 figures [48]. With an average user data rate increase by $50-100$ fold to reach a peak target of 10 Gbps [49], it is estimated that 1000-fold increase in the capacity will be required in the next decade [50].

A. *Capacity Expansion Mechanisms*

To satisfy this demand, several techniques are being studied and standardised such as massive and enhanced MIMO, beam-forming, carrier aggregation, BW expansion, CoMP and SC deployments. The capacity cube of Fig. 4 groups these techniques into three main categories: spectrum extension, spectrum efficiency and network density. Enhanced MIMO and beam-forming techniques improve the SE but they may not be sufficient to achieve the ambitious 1000-fold capacity target. In fact, SE of current systems, such as the LTE, is already close to Shannon's bounds [51], hence further improvements on the SE will have a marginal impact on the overall capacity. CoMP techniques depend on BSs cooperation to enhance the physical layer performance and to mitigate the ICI for cell edge users. Although these mechanisms may be necessary in dense deployment scenarios, they cannot achieve the 1000-fold capacity increase [52]. On the other hand, the proportional relationship between the BW and the capacity depicted by Shannon's formula [53] indicates that wider BWs give higher capacity. In this direction, carrier aggregation has been proposed in [54] for LTE-Advanced systems, where two or more carriers are aggregated together resulting into a maximum aggregated BW of 100 MHz [55]. However, this technique is limited by the allocated BW and hence wider BW allocations are required for future systems.

Two solutions to this problem are identified: spectrum sharing and new spectrum exploitation. The former shares the same portion of the spectrum between different operators under specific regulation and coordination rules, while the latter suggests exploiting new frequency bands. However, scarcity of spectrum resources in low frequency bands requires exploiting higher bands where free portions of the spectrum are available. As a result, regulatory and standardisation bodies are considering high frequency bands, such as $3.4-3.6$ GHz and above 6 GHz, as the main candidates for future cellular systems [56]. In addition, mm-wave bands (i.e., above 60 GHz) are being considered as a spectrum extension solution to satisfy the increasing capacity demand. Nonetheless, the high propagation loss of such bands limits their usability to local area and short range communications only.

Network densification allows spatial reuse of spectrum resources by reducing the cell size. The idea originates from the fact that deploying several SCs instead of one MC i.e., cell splitting, allows resource reuse across the cells. For example, in frequency reuse of one systems such as the LTE, splitting a MC into two SCs could result into doubling the capacity. Hence it can be said that spectrum extension and network densification are highly correlated. In particular, a dense deployment of SCs has been accepted to be the most promising solution to satisfy capacity demands of future cellular systems [57], [58]. Table VI summarises the capacity expansion mechanisms and highlights the spectrum reuse benefit of network densification. As a result, more SCs are being deployed within the MC coverage to offload some of the users associated with the latter. This is referred to as heterogeneous networks (HetNets), which is being considered for LTE-Advanced and beyond [59].

B. Heterogeneous Networks

In conventional HetNets, the MCs and the SCs are deployed in the same frequency band [60], thus inter-layer interference mitigation techniques such as spectrum splitting or almost blank subframes are required. However, these techniques may degrade the achievable capacity because they segment the resources between the layers either in frequency or time domains. This suggests a frequency-separated deployment, where each layer is deployed in a separate frequency band to avoid the resource splitting loss.

The CDSA is aligned with the frequency-separated deployment approach. Since the CBS provides low-rate/long-range coverage services, it can use the existing low frequency bands that offer good propagation capabilities. On the other hand, the capacity hungry plane, i.e., the DBS, can operate at high frequency bands that offer more spectrum resources and higher capacity. This approach is being investigated by several operators and research projects as a novel solution for future cellular systems. The commercial operator NTT DOCCOMO proposes a Phantom Cell Concept (PCC) for LTE-B [5], where coverage and data services are provided at low and high frequency bands respectively. The models developed in [61] show that SE of the PCC outperforms the SE offered by conventional HetNets.

In [42], a similar architecture called Cloud-HetNet has been proposed with a primary target of extending the capacity. In Cloud-HetNet, all cells (i.e., MCs and SCs) are connected to a Cloud-RAN in a star topology and they act as radio resource heads. The Cloud-RAN is the brain of this architecture, where network and medium access control layer functions and part of the baseband processing are performed in a centralised manner. Based on the Cloud-RAN, the MC can handle the CP of all users for mobility and cell discovery, while the DP is supported by the SCs. An interesting finding of [42] is that operating the DP at the 3 GHz band (where 100 MHz BW is still available) provides more capacity than the 60 GHz band (where 2.16 GHz BW is available) when the traffic load is low and vice versa. This is because [42] defines the capacity as the minimum of the achievable throughput and the traffic rate. Although the 3 GHz band offers limited BW, it can satisfy low traffic rate demands. Thus the wider BW offered by the 60 GHz band does not provide additional capacity gains and the reduced coverage minimises the achieved capacity in low load situations.

In addition, the absence of CRSs in the DP and the flexibility in switching off the DBSs reduce the DP ICI, which in turn increases the signal-to-interference-plus-noise ratio (SINR). According to Shannon's capacity formula, the latter can be translated into an increased capacity. Moreover, a higher SINR allows using high order modulation and coding schemes that provide high data rates.

C. Scalability and Reconfiguration

Heterogeneity of future networks will create high variations over spatial, time and frequency domains due to mobility, variable-rate applications and SC deployments. This requires flexible, cost efficient and reconfigurable networks that are able to adapt to such variations. In this regard, network adaptation and reconfiguration might be easily performed in the CDSA with relaxed constraints. For instance, the DP can be flexibly scaled without coverage restrictions. Thus network operators can start by deploying DBSs to satisfy the current demand only and gradually add capacity when and where it is needed. In [62], such a scalable architecture is referred to as fusion network where a host layer guarantees the connectivity while a scalable and flexible boosting layer provides on demand high data rate services.

In dense deployment scenarios, traffic tendency of each cell will be prone to high fluctuations e.g., a cell may be characterised by an asymmetric uplink (UL)/downlink (DL) traffic. In such cases, assigning static or semi-static resources for the UL and the DL could result into resource wastage. This requires flexible (re)allocation schemes to ensure efficient usage of spectrum resources. One of these schemes is dynamic time division duplex (TDD) that shares all the time slots between the UL and the DL with flexible slot reconfiguration [63]. Semi-static variations of this technique have already been implemented in current standard, for example the LTE defines seven UL/DL slot configurations [64]. However the mandatory transmission of the CRS and other periodic signals limits these techniques in the conventional architecture.

TABLE VI: Conventional capacity expansion mechanisms and their limits

Techniques/Correlation	Aim/Function	Limits
Enhanced MIMO and beam-forming	Improve the SE	Cannot achieve the 1000-fold capacity increase
CoMP	Enhance the physical layer performance and mitigate inter-cell interference	Cannot achieve the 1000-fold capacity increase
Carrier aggregation	Two or more carriers are aggregated to achieve maximum BW	Limited by the allocated BW
Spectrum extension	Spectrum sharing between operators and exploitation of new high frequency bands	High propagation loss
Network densification	Spectrum reuse by deploying multiple SCs instead of a single MC	Scalability, energy consumption, interference
HetNets	SCs are deployed within the MC coverage to offload users	Requires inter-layer interference mitigation techniques, which degrades the achievable capacity

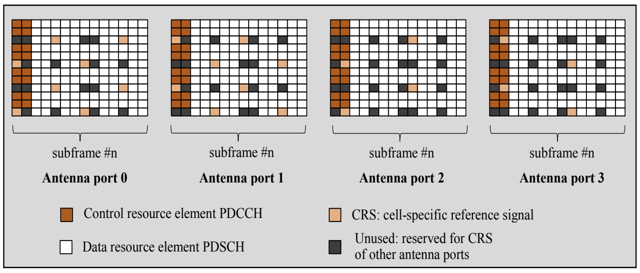

Fig. 5: LTE CRS pattern of four antenna ports [66]

On the other hand, the absence of periodic CRS/broadcast-type signalling in DBSs of the CDSA could offer a flexible implementation of dynamic TDD. Nonetheless, interference coordination between neighbouring DBSs may be required because using different UL/DL configurations in different cells implies that there could be UL-to-DL and DL-to-UL ICI in addition to the classical UL-to-UL and DL-to-DL ICI [65]. To solve this problem, a hybrid frequency division duplex (FDD) and TDD coordination scheme (hybrid FDD-TDD) has been proposed in [65] as an interference coordination technique for DBSs utilising dynamic TDD. This scheme avoids the UL-to-DL and the DL-to-UL ICI by scheduling the UL and the DL for each user in different carriers and at different subframes.

V. Centralised Control Plane

This section surveys and discusses the benefits of using a CP with central scheduling and decision entity rather than fully distributed decisions at the DP. In particular, it focuses on applications related to interference, resource and mobility management.

A. *Interference Management*

Interference control is a major concern in cellular systems especially those adopting a frequency reuse of one. To cope with this issue, several interference mitigation techniques have been standardised in the LTE, such as resource partitioning between the cells and resource muting during CRS transmission of other antenna ports, as shown in Fig. 5. Other advanced interference management strategies have been considered such as:

- Slowly-adaptive interference management [59].
- Enhanced ICI coordination (eICIC) [67].
- Autonomous component carrier selection [68].

eICIC mitigates the ICI for cell edge users by coordinating network resources in time, frequency and power domains [67]. To cope with the ICI in range expansion zones of HetNets, an extension of eICIC that complements time domain resource partitioning techniques, such as almost blank subframes, with non-linear interference cancelling receiver processing has been proposed in [69] and [70]. With dense deployment of SCs, these techniques offer limited flexibility and may not be responsive to rapid traffic variations [71]. In addition, the interference coordination will be problematic due to the increased number of interferers, and a centralised coordinator may be required to control the resource usage among different cells.

The CDSA offers flexibility in this context because the CP can play the role of the centralised coordinator. In [13], different approaches to control the ICI in control-data separation scenarios have been identified. In one scenario, the CBS fully controls the scheduling for the DBSs, which overhear the grants issued by the former. In this case, the UEs request resources from the CBS, which associates each user with the best serving DBS and schedules the users of neighbouring DBSs on different resources. This approach does not require a backhaul signalling between the CBS and the DBS, but it may introduce delay in the DL scheduling [13]. Another approach is to maintain the scheduling functionalities at the DP with scheduling constraints being defined by the CP. However, such an approach generates additional signalling between the CBS and the DBS.

B. *Resource Management*

Traditionally, cellular users camp on the network by selecting the BS that offers the strongest signal. Thus cell (re)selection is mainly UE driven with a limited control by the network i.e., the network may use offset parameters to privilege some cells [72]. Since the cell (re)selection does not require resource assignment, the UE driven approach can be justified in this case. However, the active UEs are assigned resources by the same cell initially selected by the user, which puts constraints on the resource management and optimisation process, e.g., the resources have to be assigned by this cell only without a global view of the network. This calls for network driven resource assignment strategies.

In the CDSA, the initial access procedure can be based on the received signal strength (RSS), thus the cell (re)selection could be UE driven as in the conventional architecture. However, when the UE requests resources for data transmission,

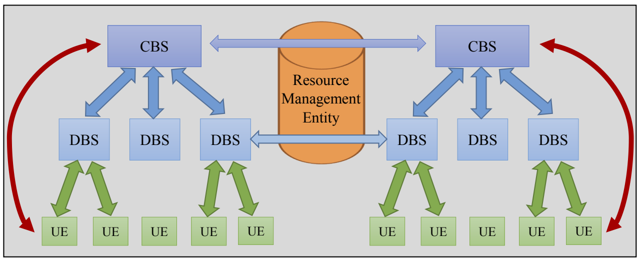

Fig. 6: Conceptual hierarchy of CDSA with RME [73]

the CBS selects the best serving DBS (or a group of candidate DBSs) with a wide view of network status and parameters such as EC, congestion, performance requirements, etc. This allows a transition from almost distributed to almost centralised radio resource management mechanisms and optimises the resource allocation process. A centralised resource management entity (RME) has been proposed in [73] for cellular networks with a CP/DP separation. The main responsibility of the RME is to select the best serving DBS based on network status and context information collected from the UEs and the network nodes. The RME is accessible to both the CBS and the DBS and several trade-offs are identified in [73] to optimise the RME decision. A conceptual hierarchy of CDSA with RME is shown in Fig. 6.

The centralised resource management schemes do not only optimise the resource selection decision, but also they could help in balancing the network load. In dense SC scenarios, the number of active UEs per cell is expected to be small. Thus each cell will be characterised by a highly fluctuating traffic profile [74]. In this case, the CBS (or a separate RME) can determine an average load threshold for each cell and then exposes only the appropriate set of DBSs to the UEs in order to balance the traffic load. Another approach is to allow the UE to conduct measurements of surrounding DBSs based on which the CBS (or the RME) can allocate the serving DBS. From another perspective, the flexible opportunities for power adjustment in the DP can be realised to temporarily reshape the coverage of a low utilised DBS to overlap with a neighbouring congested DBS, thus offloading the latter and balancing the network load. In current systems, such coverage reshaping is very limited due to CRS interference as well as the constraints imposed by the planned coverage [5].

C. Mobility Management

As the cell size decreases, mobility management becomes complex because the HOs will happen frequently even for low mobility users. In the conventional cellular architecture, the HO procedure includes transferring all channels (i.e., control and data) from one BS to another with a significant core network signalling load [75]. With a frequent HO rate, the signalling overhead and the call drop rates will increase significantly, which could degrade the quality of experience. On the contrary, the CDSA could offer simple and robust HO procedures when the RRC connection is maintained by the CBS (which is typically a MC). As a result, the intra-CBS HOs (i.e., between DBSs under the footprint of the same CBS) might be transparent to the core network. This in turn alleviates mobility signalling and reduces the HO failure (HOF) probability [76].

In [77], the HOF and the radio link failure (RLF) rates are used as key performance indicators to analyse mobility performance of the CDSA. The RLF rate is defined as the average number of RLF occurrence per UE per second, where RLF is triggered when the DL SINR is below a certain threshold ($Q_{out}$$=-8$ dB) and stays below -6 dB for at least 1 s [78]. On the other hand, the HOF rate is defined as the ratio between the number of HOF and the total number of HO attempts, where the HOF is triggered when the RLF occurs during the HO execution time [78]. The authors of [77] argue that the UE is always anchored to a MC (i.e., the CBS), thus the RLF and the HOF rates reflect the macro layer mobility. System level simulation results show that the RLF and the HOF rates of the CDSA are roughly 0.6% with a reduction in core network signalling by a factor of $3-4$ as compared with the conventional architecture. It is worth mentioning that [77] assumes that the MCs and the SCs are deployed in separate frequency bands. Thus the inter-layer interference is ignored. In co-channel deployments, however, this interference might be significant, which could degrade the SINR and increase the RLF and the HOF rates. This indicates that deploying the CP and the DP in separate frequency bands might be more appropriate from a mobility point of view.

Context information such as mobility history can play a key role in optimising the RRC and the HO process. It can be used to select the most appropriate DBS for a moving terminal, e.g., a DBS with the highest probability that the user will not leave it quickly [4], [73]. In addition, predicting the user's trajectory, i.e., a sequence of DBSs that the user will visit them, allows the CBS to make advance HO decisions. Thus each candidate DBS in the user's path can prepare and reserve resources in advance, which in turn could relax the HO requirements and minimise the interruption time [79]. Moreover, the CBS can provide the candidate DBSs with some information about the UE, such as capabilities, authentication information, etc., to minimise the air interface signalling between the UE and the DBS when the HO is executed. Nonetheless, such techniques require a reliable mobility prediction scheme; an area with a wealth of literature, see for example [79]–[84] and the references therein.

VI. Signalling Overhead

Meeting the ambitious 5G targets of 10 Gbps peak data rate and 1 ms roundtrip latency [85] needs addressing a critical issue: signalling overhead. Current signalling mechanisms are designed to operate efficiently to some extent for current density levels. However, the dominant theme for future cellular systems, i.e., network densification, may not be suitable with these mechanisms due to the expected dramatical increase in signalling overhead.

Traditionally, all cellular users are connected to the same BS irrespective of their activity state (i.e., active, idle or detached) provided that they are within the footprint of this BS. Thus the same physical layer frame is used by all UEs and hence most of the control signals are cell-specific rather than user-specific resources. For example, the CRS of the LTE is used as a pilot by active and idle UEs for channel quality measurements and for channel estimation to allow coherent demodulation of control and data channels. In addition, it is used in the initial access phase to demodulate the broadcast channel [86]. Since channel conditions of the detached and the idle UEs are usually unavailable, these signals are distributed in the time/frequency grid based on the worst-case scenario, e.g., high mobility assumptions [87]. Although this approach guarantees acceptable performance for all users including those in severe conditions, it over-provisions the physical layer frame under moderate or good channel conditions [88].

In current standards, this overhead consumes a significant part of transmission resources. In the LTE for example, the CRS has a fixed overhead of 4.76% with one antenna port. To avoid CRS interference between different antenna ports of the same BS, the LTE adopts a shifted CRS pattern with resource muting [66]. As shown in Fig. 5, when an antenna port transmits its CRS, other ports mute their transmissions. Hence the CRS overhead increases to 14.25% with four antenna ports. Similarly, the cyclic prefix has a fixed overhead of 7.14% and 25% for normal and extended cyclic prefix respectively [66]. In the six middle resource blocks, four OFDM symbols in the second time slot of the first subframe are reserved for the broadcast channel, while the primary and the secondary synchronisation signals are transmitted every 5 ms and they occupy two OFDM symbols in each transmission [66]. The overall LTE overhead depends on the used configuration i.e., BW, antenna ports, duplex mode, etc. and it can reach up to 40%.

Some proposals to reduce this overhead are being considered such as using several classes of pilots with each class being transmitted at the necessary rate, e.g., high rate UE-specific reference signal (UE-RS) for data detection and low rate channel state information reference signal (CSI-RS) for link adaptation measurements. Nonetheless, these signals also have a static pattern constrained by the worst-case conditions. In the CDSA, however, the DBS is invisible to both the idle and the detached UEs and its on-demand connection with the active UEs is established and assisted by the CBS. This relieves the DBS from the task of transmitting CRSs and removes the constraints imposed by the unknown channel conditions of the inactive UEs. As a result, the DBS-UE link lends itself to flexible, adaptive and optimised operations. For instance, the DBS pilot signal can be considered as a UE-specific resource and its transmission rate/pattern can be adaptively adjusted according to the temporal channel conditions reported by the active UEs. This feature has been investigated in [89] where a 75% reduction in pilot overhead is achieved as compared with the LTE CRS pattern. Similarly, the cyclic prefix and other signals can be adaptively and flexibly adjusted to minimise the overhead.

Depending on the adopted separation scheme, the DBS frame structure can be simplified and several signals can be removed [88]. Following the mapping of Tables II and III, the broadcast and the multicast bearer signals may not be required in the DP if these functionalities are delegated to the CP. In other words, the DBS transmits UE-specific signals only, hence its overhead scales with the amount of data being transmitted whilst removing the baseline overhead caused by the periodic CRS/broadcast-type signalling. To summarise Sections III–VI, Table VII lists limitations of the conventional architecture along with the system improvements from the CDSA.

VII. Challenges and Enabling Technologies

The CDSA aims to provide a framework where limitations of the conventional architecture can be overcome, as discussed in Sections III–VI. However, there are several research challenges and questions that need to be answered in order to concretely assess the feasibility and superiority of this architecture over the conventional one. These issues include: serving node selection, control message and data frame design, backhauling mechanisms, heterogeneous deployment with dual connectivity, channel estimation and management of discontinuous transmission techniques. In addition, the promotion of the CDSA as a candidate RAN for future cellular systems is tightly coupled to the emerging concept of SON [90]. In this section, we discuss some of the CDSA challenges and provide a survey of the preliminary work that has been done to solve these issues. Furthermore, we provide an overview of SON implementation in the CDSA.

A. *Context Information*

Traditionally, the control and the data services are provided by the same physical node and signal strength is usually used as a metric for serving node selection and HO decisions. Separating the CP from the DP makes these decisions non-trivial because different services (i.e., control and data) are provided by separate nodes. These nodes might be deployed at different locations and they could have different characteristics such as transmission power, antenna pattern, etc. Thus, the RSS (at/from the CBS) cannot be used as a metric for selecting the serving DBS. Relying on measuring the DBS signal may not be feasible either because the best serving DBS may not be discoverable by the UE (e.g., switched off for energy saving or interference reduction). As a result, the DBS-UE association requires assistance by the CBS which brings several advantages as discussed earlier. However, such network driven approaches require intelligence, context awareness and CP/DP coordination.

Position information can be considered as the simplest metric in associating the UE with a DBS. Nonetheless, this criteria does not guarantee selecting the best serving DBS because obstacles and other loss mechanisms may exist in the path between the UE and its nearest DBS. Broadly, radio channels between the UE and other DBSs may be better than the nearest DBS channel. Hence the CBS needs to obtain knowledge of channels conditions between the UE and each candidate DBS. Other parameters such as EC, mobility history, application requirements and network status can also affect the DBS selection decision [91]. For instance, assigning the UE to an already-awake DBS that is able to satisfy its requirements and excluding the inactive DBSs from the candidate set (where

TABLE VII: Comparison between the conventional architecture and the CDSA

Conventional architeture limitations	CDSA solutions and proposals	Scenarios of interest
High EC due to always-on service approach	Low EC with on-demand always-available system: • DBS with on/off operation [6], [40], [41] • DBS with high gain and selective beam-forming [11], [42]	Low utilisation, dense deployment
Wide area coverage may not be guaranteed at high frequency bands	Wide area coverage is provided at low frequency bands with dual connectivity: • Phantom Cell: CBS and DBS at low and high frequency bands, respectively [5], [12], [61]	SC at high frequency bands
Scalability and coverage trade-off	DBSs can be gradually deployed when and where they are needed without coverage constraints: • Cloud-HetNet: Radio resource heads with Cloud-RAN [42] • Fusion network: on demand deployment of DBSs [62]	SC at high frequency bands
Resource wastage with traffic tendency fluctuation due to (semi)static resource assignment constraints	Flexible DBS reconfiguration opportunities due to the absence of CRS: • DBS with Dynamic TDD [65]	Dense deployment
Limited interference coordination between the cells based on local scope	Centralised interference coordination with a wide view of network status and parameters: • CBS as a centralised coordinator [13]	Dense deployment
Resource selection is UE driven	Resource selection is network driven: • Centralised RME for DBS-UE association [73]	Dense deployment
High mobility signalling overhead and poor HO performance	UE is anchored to a MC with large coverage area: • RRC and mobility management are handled by the CBS [76], [77] • DBS mobility prediction with advance resource preparation [4], [79]	Dense deployment, high speed
large and static frame overhead due to the worst-case design approach	DBSs serve active UEs only: • DBS with UE-specific signals only [6], [88] • DBS with adaptive frame allocations [89]	Local area

possible) could reduce the EC significantly. On the other hand, mobility pattern/history optimises the resource selection process for moving terminals, as discussed in Section V-C. This highlights the importance of context awareness in the CDSA, which can be exploited to improve the EE, optimise the HO parameters and to design optimum traffic management policies.

Gathering the context information is one of the research challenges that need to be addressed. Some information can be easily and reliably gathered in current standards. Position information can be provided by a global positioning system (GPS) or other mature techniques. However, new mechanisms are needed to predict channel conditions between the UE and each candidate DBS. In this area, [92] proposes a database-aided channel quality prediction technique for cellular systems with CP/DP separation. Each CBS is equipped with a database that contains SNR measurements for each DBS under its control. As shown in Fig. 7, this database maps each measurement to the geographical location where it is reported from, with x and y being the location coordinates that depend on the required granularity e.g., longitude and latitude.

The authors of [92] use the SNR as a metric for channel quality prediction and argue that the SNR values consume less memory than the SINR. Although the latter provides a better measure than the former, the instantaneous interference depends on network status. Thus storing SINR measurements of each state may not be feasible from a memory perspective. The database training process requires the UEs to report their locations along with the DBS pilot measurements to the CBS. If there is no previous measurement for the location being reported by the UE, the reported value is added to the database. Otherwise, an exponential moving average is used to incorporate the new value. In this way, the UE can measure pilot signals of the active DBSs and use the stored values of the sleeping DBSs to determine the best channel quality. It can be noticed that this technique predicts the signal quality at the UE

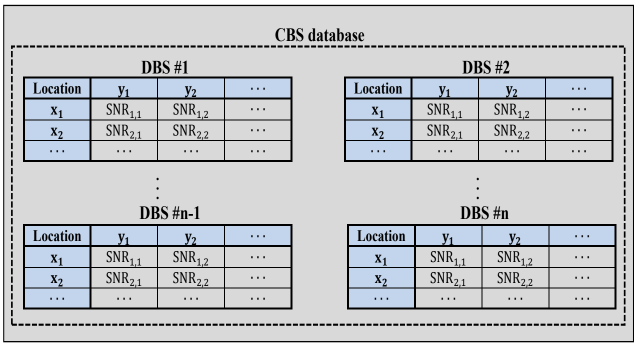

Fig. 7: SNR measurements database for channel quality prediction

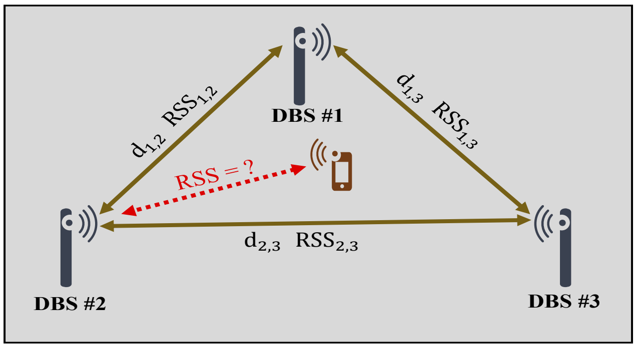

Fig. 8: Inter-DBS measurements in the RSS-to-distance map technique proposed in [93]

location given that there were previous measurements in this location. Nevertheless, it does not address the case when the propagation environment changes or when there is no previous reports from the UE position. This raises a question of how to deal with out-of-date measurements and how to interpolate between the database entries.

Two approaches have been proposed in [93] to tackle these issues. The first solution relies on a power map constructed by collecting RSS fingerprints previously reported by UEs at different locations. The expected RSS at a specific UE position can be estimated by averaging the nearest fingerprints that are weighted according to their distance from the UE. At location l, the expected signal S_l from a certain DBS based on the k nearest fingerprints can be formulated as [93]:

$$S_l = \sum_{i=1}^{k} w_i \, S_i \tag{1}$$

where S_i is the fingerprint reported from position i and w_i is a normalised weight factor inversely proportional to the distance between the UE and position i. This technique does not require channel modelling and it implicitly includes fading and non-line of sight effects. However, an accurate estimation of S_l requires a reliable fingerprint database, which can be obtained through time consuming drive tests. Wardriving (i.e., online database construction) and fingerprint prediction methods can also be used to construct the power map but they provide coarse and less accurate predictions [93].

The second approach relies on inter-DBS RSS measurements to select the serving node. Given the location of each DBS, a RSS-to-distance map is constructed, which can be used to estimate the expected signal at the UE's location, as shown in Fig. 8. In contrast to the power map technique, the RSS-to-distance map method depends on measurements between the DBSs only. As a result, it does not require pre-configuration (i.e., drive tests) or propagation constants estimation. In addition, it can adapt to the propagation environment by updating the estimated signal according to the instantaneous inter-DBS measurements [93].

B. CP/DP Backhaul

Despite the CDSA benefits discussed in Sections III–VI, a major drawback of this architecture is the backhaul connection requirements. Although the CBS and the DBS support different functionalities, they need to coordinate and communicate with each other [94]. The serving node selection decision requires the DBSs to exchange information about their current status, such as EC, congestion, etc., with the CBS. This coordination is also required to optimise scheduling, resource allocation, interference management and mobility management. As an illustration, the CBS and the DBS may negotiate whether the UE will be served by the DBS (e.g., a low speed terminal) or whether it will be served at a low rate by the CBS only to minimise the mobility overhead in high speed scenarios.

Tight collaboration and excessive signalling exchange between the CBS and the DBS provide reliable, robust and updated information. However this increases the system overhead [7] and requires an ideal backhaul connection, i.e., high throughput and low latency [13]. On the other hand, a low rate CBS/DBS signalling relaxes the backhaul requirements and reduces the overhead but it may result into unreliable or out-of-date information. This raises a research challenge of how to design efficient CBS/DBS signalling mechanisms. These are tightly coupled to the adopted separation scheme as well as the DP route from the core network to the UE. A DP path following the route: core network $\rightarrow$ DBS $\rightarrow$ UE requires a low latency signalling exchange to allow tight coordination between the CBS and the DBS. On the other hand, the DP path: core network $\rightarrow$ CBS $\rightarrow$ DBS $\rightarrow$ UE relaxes the latency requirements, however, it demands a high backhaul BW [95].

To satisfy the BW requirements of the CDSA backhaul, the MiWEBA project works on the design of efficient wireless backhaul networks operate in the mm-wave band. In the latter, a 7 GHz BW is available and a peak backhaul rate of 6.5 Gbps is achieved with 150 m separation [96]. However, the mm-wave wireless backhaul may not be a feasible solution for large CBS/DBS separation distances. In addition, it may not be suitable in dense urban environments due to the absence of line of sight. Thus, [97] proposes a hybrid optical/mm-wave backhaul for such scenarios, where repeaters are used in the backhaul link.

It is worth mentioning that the backhaul technology impacts the overall EE of the CDSA. Preliminary work in this area considers an ideal fibre optic backhaul with X2 interface between the CBS and the DBS [95]. This backhaul is used to transmit DBS activation/deactivation messages and to exchange load information in energy efficient CDSA systems with sleep modes. An extension of this study is reported in [98] to evaluate EE of the CDSA with several backhaul options. These include fibre optic, symmetric ideal fibre optic, wired cable and wireless backhaul. The results show that the wireless and the symmetric ideal fibre optic backhaul consume the least and the most, respectively, amount of active mode power. In the idle mode, however, the fibre optic backhaul is the most energy efficient among the considered options [98].

The concept of over the air (OTA) signalling [99], [100] is being investigated in METIS project as an alternative solution to avoid using the conventional direct backhaul networks. METIS approach of OTA signalling requires the CBS to take full control of scheduling functionalities while the DBSs overhear the grants issued by the CBS [101]. This alleviates the need for an ideal backhaul, however, a robust signalling design is required for interference avoidance. This calls for the design of optimum transmission gaps to allow efficient OTA signalling without a significant interruption to the physical layer transmission [94].

C. Self-Organising Networks

The prohibitive cost/effort for manual configuration and optimisation of network elements and parameters has motivated researchers and standardisation bodies to automate these procedures. In 3GPP parlance, such automatic operations are usually referred to as SON, which cover three areas [102]:

- self-configuration: concerns with pre-operational procedures such as automatic configuration and integration of newly installed BSs in a plug-and-play mode.
- self-optimisation: dynamically adjusts and optimises the operational characteristics in an automatic manner to cater for traffic patterns and propagation environment variations.
- self-healing: minimises failure impact by identifying the failing element(s) and adjusting the appropriate parameters for service recovery.

As far as the CDSA is concerned, the self-optimisation capability can be considered as the most important aspect of SON. It can play a key role in enabling most of the CDSA applications especially those related to energy saving, load balancing and mobility robustness. With rapid traffic variations, switching the DBSs on/off and controlling their beam directions manually would not be feasible and may not be timely enough. A more convenient design approach is to automate these techniques in order to ease their implementation and to maximise their effectiveness. Such SON-based mechanisms allow recognition of short term energy saving opportunities and they enable proper reconfiguration of long term EE improvement strategies [103].

An energy self-optimising scheme has been proposed in [104] to automate the BS wakeup and hibernation process by using an online optimisation algorithm. Similarly, the 3GPP investigates several energy saving deployment scenarios with an underlay coverage layer provided by macro BSs of legacy networks. In these scenarios, RAN nodes of the capacity boosting layer are switched on/off automatically with commands issued by a centralised operation, administration and management entity or by the node itself according to certain criteria and polices [105].

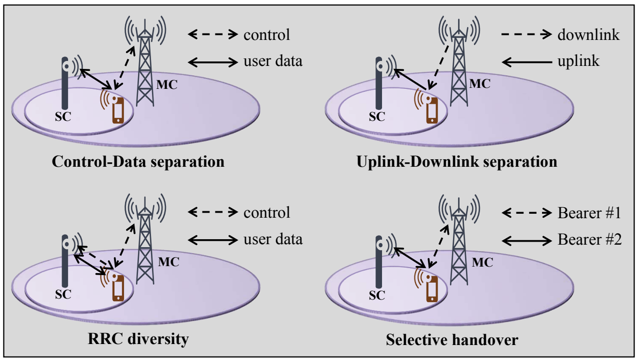

Fig. 9: Dual connectivity cases [108]

By considering load balancing as the primary objective of the self-optimisation process, [106] proposes a mobility load balancing mechanism to shift cell edge users from congested to low utilised BSs whilst minimising the associated mobility overhead. Although this technique do not consider a CP/DP separation, the basic concept can be applied to the CDSA without significant modifications. A UE-like BS that uses a M2M interface to communicate with the UEs has been proposed in [107] for load balancing purposes. This BS does not transmit any CRS and its M2M links with the UEs are controlled by a macro BS in a master-slave relationship. This configuration can be considered as a SON-based CDSA, and the results reported in [107] show that this architecture maximises the throughput whilst balancing the network load.

VIII. Preliminary Standardisation Work

This section presents an overview of preliminary standardisation proposals related to the CDSA. In particular, we focus on the aspects being studied in the 3GPP for future LTE releases.

A. Dual Connectivity

In conventional HetNets, the independent operation of SC and MC layers raises several problems as discussed earlier. Thus the standardisation bodies are considering integrating these layers by allowing the UE to communicate simultaneously with both the SC and the MC. This is something referred to as *dual connectivity* and it is being investigated by the 3GPP for LTE Release-12 and beyond under the study item "small cell enhancements". According to [108] and [109], dual connectivity may imply control-data separation, UL-DL separation, RRC diversity or selective HO, as shown in Fig. 9.

- **Control-Data separation:** This is the classical scenario relevant to the CDSA, where data transmission and RRC signalling are provided by different nodes, namely SCs and MCs, respectively, to alleviate mobility overhead and to minimise cell planning effort. As the user moves from one location to another, it transmits/receives data to/from the nearest SC. Since the UE is anchored to a MC, switching the data path from a SC to another will not trigger a HO as long as the UE stays within the coverage area of the same MC. Thus mobility robustness is considered as the main benefit of dual connectivity with control-data separation [110]. Although transferring the data channel only will incur less overhead and delay as compared with the traditional HO procedures, new signalling mechanisms are required to enable such light-weight HOs [110].

- **Uplink-Downlink separation:** The power imbalance between the MCs and the SCs in conventional HetNets implies that the best serving node in the DL may be different from the best one in the UL. Traditionally, the RSS is used as a criteria for cell (re)selection, thus the high power of the MCs indicates that they would be better candidates (from a DL perspective) than the SCs even if the UE is in the vicinity of the latter. However, this may not be the case for the UL because the limited UE power suggests UL transmission to the nearest BS that offers the lowest path loss [108]. Cell-range extension can be used to increase the uptake area of low-power nodes, which are typically better UL (but not DL) choice. Expressed differently, current standards optimise either the UL or the DL performance. This trade-off can be avoided by offloading the UL traffic to the SCs whilst keeping the DL traffic in the MCs [110], [111].
- **RRC diversity:** The dual connectivity is exploited in this case to provide the RRC signalling via multiple links in order to support a robust CP, as well as to enhance the mobility performance [109].
- **Selective handover:** This scenario aims to provide different services via different nodes, e.g., high-rate best-effort services are provided by SCs while low-rate voice services are supported by MCs. This can be achieved by using different HO thresholds for different services [108].

In the following, we focus only on dual connectivity with control-data separation, since it is relevant to the CDSA. A CP/DP split model has been proposed in [112] based on protocol stack of the LTE. In this model, most of the data radio bearers (DRBs) are established at the SCs that have a direct interface with the serving gateway (S-GW). On the other hand, the signalling radio bearers and few DRBs are established at the MCs that manage all the DRBs even though the latter are established at the SC layer. Only one RRC connection is established between the UE and the MC for connection control and mobility management, while an interface X_c between the MC and the SC is used to exchange the critical and less dynamic information. Fig. 10 shows the protocol stack of the MC and the SC in this model.

Two types of transmission/reception modes have been proposed in [113] to support the dual connectivity. These include simultaneous and time division multiplexing (TDM) modes. In the former, the UE transmits/receives data to/from both the SC and the MC at the same time. Although it utilises the resources efficiently, this mode may complicate the UE RF design and requires a careful power control when the MCs and the SCs are deployed in the same frequency band [113]. On the other hand, the TDM mode segments the resources between the SC and the MC in the time dimension. At a given time (or subframe in LTE terminology), the UE transmits/receives to/from either the SC or the MC. In contrast to the simultaneous mode, the TDM mode does not require the UE to operate with two carriers at the same time, which relaxes the requirements on the UE RF capabilities [111]. However, it results into resource splitting loss. Fig. 11 shows the simultaneous and the TDM modes.

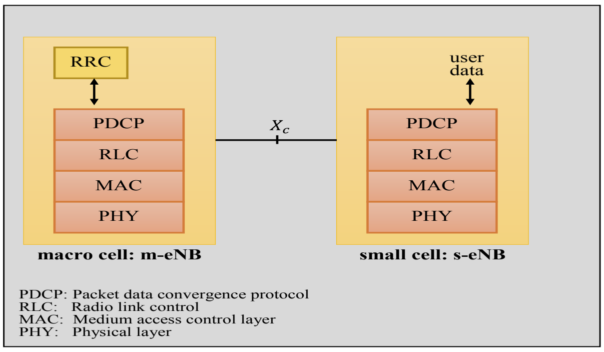

Fig. 10: Protocol stack of macro and small cells in dual connectivity with control-data separation [112]

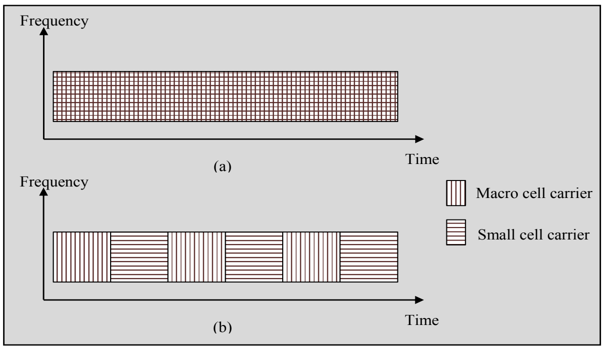

Fig. 11: Transmission/Reception modes for dual connectivity. (a) Simultaneous mode. (b) TDM mode.

B. Lean Carrier

In the LTE, the CRS is transmitted in every subframe regardless of whether the subframe contains user data or not. In addition to ICI and SE loss, this periodic transmission of CRS prevents switching off the BS transmission circuitry during unoccupied subframes. Targeting higher SE with minimum EC, [114] proposes a new carrier type, known as *lean carrier*, for LTE Release-12. In the latter, the CRS is replaced with UE-RS for channel estimation along with CSI-RS for channel quality measurements. Unlike the CRS, the UE-RS is transmitted only in resource blocks that contain users' traffic and its overhead scales with the amount of data being transmitted. As a result, the BS can be switched off during the unoccupied subframes, which could scale the EC with the traffic load. In addition, the legacy control channels that occupy the first $1-4$ OFDM symbols of each subframe across the entire BW are replaced in the lean carrier with an enhanced control channel. The latter is transmitted in few resource blocks in a similar fashion as unicast data [115], which reduces the control channel overhead.

It can be noticed that the design approach of the lean carrier is aligned with requirements and objectives of the DP in the CDSA. In [114], a scenario for using the lean carrier in a dual connectivity architecture is presented, where the UE uses the legacy LTE carrier to communicate with the MC for vital control information, while the lean carrier is utilised for data communication with the SCs.

IX. Conclusion and Future Research Directions

A comprehensive survey of RANs with CP/DP separation was presented. Potential benefits of this architecture and its superiority over the conventional one were critically discussed. In addition, the preliminary work to tackle its technical challenges was highlighted along with the ongoing discussion in standardisation forums related to this research vision. Based on this survey, we draw the following conclusions and underline potential research directions:

- The evolution of current cellular generations is driven by performance improvement with an anywhere/anytime service paradigm. Nonetheless, futuristic demands, use cases and deployment scenarios require considering additional dimensions to ensure efficient, sustainable and cost effective cellular networks.
- Green communication is a hot topic nowadays. The mature and widely available EC profiles of the most power consuming element (i.e., the BS) highlight the importance of moving towards dynamic on/off BS operations. However, coverage constraints limit opportunities and gains of such techniques. This suggests a RAN architecture with an underlay always-on connectivity layer complemented by an overlay on-demand data layer: the CDSA.
- In energy efficient CDSA networks, the best serving DBS may not be discoverable by the UE. Expressed differently, channel state information of some DBSs may not be available when they are switched off for energy saving. This in turn constrains either resource selection strategies or sleep mode polices. In the former, the UE may not be able to measure signals of the inactive DBSs, thus the serving node selection decision would exclude these DBSs from the candidate set. A simple solution is to adopt sub-optimal sleep modes where the DBS has to be switched on periodically to send pilot signals for measurement purposes. However, such an operation limits the period in which the DBS components can be turned off. This suggests indirect measurement techniques to avoid the periodic transmission of pilot signals, thus overcoming limitations of the traditional sleep mechanisms. For this purpose, contextual information plays a critical role. The preliminary models in [92] and [93] exploit position information and measurement history to predict channel quality of switched off DBSs. This opens a research direction towards developing efficient, self-organising techniques for gathering and maintaining the contextual information as well as analysing the fingerprints' accuracy/overhead trade-off [116].
- Network densification has been accepted to be the dominant theme for future cellular systems. In such scenarios, the conventional RAN architecture, where each BS makes decisions based on its local scope only, may not be suitable. A more conscious approach will require centralised decision makers that have a global view of the network. The CDSA presents itself as a promising solution to address this issue by enabling the CBSs to act as centralised coordinators (and possibly decision makers) for the DBSs under their control. In this regard, the Cloud-RAN and the SDN concepts can play a key role.
- The survey of the CDSA potential benefits presented in Sections III–V draws a key conclusion on the importance of network driven user association strategies. In contrast to the conventional architecture, network access and service provisioning are supported by different nodes in the CDSA, namely the CBS and the DBS respectively. This allows the former to select the best serving DBS with a wide view of network status and parameters such as EC, congestion and interference. Although this approach could offer an efficient resource management, optimising the association decision is not trivial. Identifying and prioritising the optimisation objectives is a challenging task due to the trade-offs involved and the dynamic nature of operational networks. For instance, an energy efficient user association strategy could lead to load imbalance across the network. Similarly, a resource assignment scheme with a primary target of maximising the data rate could degrade the network's EE. This calls for an adaptive DBS-UE association strategy with a joint optimisation in energy, load balance and capacity dimensions.
- The HO signalling overhead can be minimised with a separation scheme that allows intra-CBS HOs to be transparent to the core network. This motivates a network design with a local mobility anchor at the CBS, resulting into light-weight HO procedure between the DBSs. This approach can alleviate mobility overhead and minimise the core network signalling related to the HO process. A first attempt to investigate this claim is reported in [77] where a 75% less core network signalling, as compared with the conventional architecture, is observed. However, [77] depends on simulations only, and to the best of our knowledge there are no analytical models that evaluate the HO overhead in the CDSA, although some models exist for the conventional architecture. Thus a research effort is needed to develop models for HO signalling cost/overhead under dual connection.
- In the CDSA, the definition of coverage is different from the classical meaning that it has in conventional systems. Specifically, two types of coverage can be distinguished: area coverage and service coverage. A user with an area coverage (provided by the CBS) is a user that can camp on the network and issue a service request whenever needed. On the other hand, a user with a service coverage is a user that can get the promised quality of service such as the required data rate. Since a subset of the DBSs can be switched off during off-peak periods, some users (e.g., active and moving UEs) may not get a service coverage

although they have an area coverage. As a result, network planning requires optimised activity patterns to properly distribute the DBSs in order to guarantee service coverage for all or most of the users whilst minimising deployment and running costs [73]. From an energy perspective, new network deployment solutions are required to maximise the DBS sleeping opportunities [11].

- Backhauling is one of the major concerns surrounding the CDSA. The latter demands highly efficient backhauling networks with tighter requirements (e.g., less delay and more BW), as compared with the conventional architecture. However, advanced mechanisms such as OTA signalling can be exploited to reduce the backhaul infrastructure.
- The concepts of dual connectivity and lean carrier constitute first attempts by the 3GPP to introduce this architecture in LTE Release-12 and beyond. With these aspects being considered in the standard, the CDSA can be seen as a strong candidate in the context of 5G networks.

ACKNOWLEDGEMENT

We would like to acknowledge the support of the University of Surrey 5GIC (http://www.surrey.ac.uk/5gic) members for this work.

REFERENCES

[1] J. G. Andrews *et al.*, “What will 5G be ,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.
[2] A. Scrase, “Summit key messages,” in *Future Mobile Summit*, November 2013. [Online]. Available: http://docbox.etsi.org/Workshop/2013/201311_FUTUREMOBILESUMMIT/ZZ_CONCLUSIONS.pdf
[3] Cisco, “Visual networking index,” White Paper, February 2014.
[4] A. Capone, A. Fonseca dos Santos, I. Filippini, and B. Gloss, “Looking beyond green cellular networks,” in *Proc. of 9th Annual Conference on Wireless On-demand Network Systems and Services*, January 2012, pp. 127–130.
[5] H. Ishii, Y. Kishiyama, and H. Takahashi, “A novel architecture for LTE-B :C-plane/U-plane split and phantom cell concept,” in *Proc. of IEEE Globecom Workshops*, December 2012, pp. 624–630.
[6] X. Xu, G. He, S. Zhang, Y. Chen, and S. Xu, “On functionality separation for green mobile networks: concept study over LTE,” *IEEE Commun. Mag.*, vol. 51, no. 5, pp. 82–90, May 2013.
[7] T. Zhao, P. Yang, H. Pan, R. Deng, S. Zhou, and Z. Niu, “Software defined radio implementation of signaling splitting in hyper-cellular network,” in *Proc. of ACM 2nd workshop on Software Radio Implementation Forum*, 2013, pp. 81–84.
[8] L. Yan and X. Fang, “Decoupled wireless network architecture for high-speed railway,” in *Proc. of International Workshop on High Mobility Wireless Communications*, November 2013, pp. 96–100.
[9] J. Feng and M. Potkonjak, “Power minimization by separation of control and data radios,” in *Proc. of IEEE CAS Workshop on Wireless Communication and Networking*, vol. 9. Citeseer, 2002, pp. 112–121.
[10] S.-L. Chen, H.-Y. Lee, C.-A. Chen, H.-Y. Huang, and C.-H. Luo, “Wireless sensor network system by separating control and data path (SCDP) for bio-medical applications,” in *Proc. of European Microwave Conference*. IEEE, 2007, pp. 430–433.
[11] M. Olsson, C. Cavdar, P. Frenger, S. Tombaz, D. Sabella, and R. Jantti, “5GrEEn: Towards green 5G mobile networks,” in *Proc. of IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, October 2013, pp. 212–216.
[12] R. J. Weiler *et al.*, “Enabling 5G backhaul and access with millimeter-waves,” in *Proc. of European Conference on Networks and Communications*, June 2014.
[13] N. Brahmi *et al.*, “METIS project deliverable D4.1: Summary on preliminary trade-off investigations and first set of potential network-level solutions,” September 2013. [Online]. Available: https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D4.1_v1.pdf
[14] L. Li, Z. Mao, and J. Rexford, “Toward software-defined cellular networks,” in *Proc. of European Workshop on Software Defined Networking*, October 2012, pp. 7–12.
[15] Z. Zaidi, V. Friderikos, and M. Imran, “Future RAN architecture: SD-RAN through a general-purpose processing platform,” *IEEE Veh. Technol. Mag.*, vol. 10, no. 1, pp. 52–60, March 2015.
[16] A. Gudipati, D. Perry, L. E. Li, and S. Katti, “SoftRAN: Software defined radio access network,” in *Proc. of ACM SIGCOMM workshop on Hot topics in software defined networking*. ACM, 2013, pp. 25–30.
[17] V. Yazc, U. Kozat, and M. Oguz Sunay, “A new control plane for 5G network architecture with a case study on unified handoff, mobility, and routing management,” *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 76–85, November 2014.
[18] P. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, “Design considerations for a 5G network architecture,” *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 65–75, November 2014.
[19] H. Ali-Ahmad *et al.*, “CROWD: An SDN approach for densenets,” in *Proc. of 2nd European Workshop on Software Defined Networks*, October 2013, pp. 25–31.
[20] ——, “An SDN-based network architecture for extremely dense wireless networks,” in *Proc. of IEEE SDN for Future Networks and Services*, November 2013, pp. 1–7.
[21] S. H. Jeon, H. Oh, H. Lee, and J. Choi, “Issues and prospects for energy saving networks,” in *Proc. of International Conference on ICT Convergence*, October 2012, pp. 201–204.
[22] H. Bogucka and A. Conti, “Degrees of freedom for energy savings in practical adaptive wireless systems,” *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 38–45, June 2011.
[23] G. Fettweis and E. Zimmermann, “ICT energy consumption-trends and challenges,” in *Proc. of the 11th International Symposium on Wireless Personal Multimedia Communications*, vol. 2, no. 4, 2008.
[24] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, “The global footprint of mobile communications: The ecological and economic perspective,” *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 55–62, August 2011.
[25] G. Auer *et al.*, “How much energy is needed to run a wireless network?” *Wireless Commun.*, vol. 18, no. 5, pp. 40–49, October 2011.
[26] L. Correia *et al.*, “Challenges and enabling technologies for energy aware mobile radio networks,” *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66–72, November 2010.
[27] V. Mancuso and S. Alouf, “Reducing costs and pollution in cellular networks,” *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 63–71, August 2011.
[28] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, “Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks,” *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1525–1536, September 2011.
[29] G. Auer *et al.*, “EARTH project deliverable D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown,” 2010.
[30] H. Guan, T. Kolding, and P. Merz, “Discovery of Cloud-RAN,” in *Proc. of Cloud-RAN Workshop*, April 2010. [Online]. Available: http://212.199.115.173/~thecom/files/wordocs/article_download.pdf
[31] A. D. Domenico, E. C. Strinati, and A. Capone, “Enabling Green cellular networks: A survey and outlook,” *Computer Communications Journal*, vol. 37, pp. 5–24, January 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366413002168
[32] T. Chen, Y. Yang, H. Zhang, H. Kim, and K. Horneman, “Network energy saving technologies for green wireless access networks,” *Wireless Commun.*, vol. 18, no. 5, pp. 30–38, October 2011.
[33] B. Badic, T. O’Farrrell, P. Loskot, and J. He, “Energy efficient radio access architectures for green radio: Large versus small cell size deployment,” in *Proc. of IEEE Vehicular Technology Conference Fall*, September 2009, pp. 1–5.
[34] A. Conte, A. Feki, L. Chiaraviglio, D. Ciullo, M. Meo, and M. Marsan, “Cell wilting and blossoming for energy efficiency,” *Wireless Commun.*, vol. 18, no. 5, pp. 50–57, October 2011.
[35] Z. Hasan, H. Boostanimehr, and V. Bhargava, “Green cellular networks: A survey, some research issues and challenges,” *Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 524–540, April 2011.
[36] Z. Niu, Y. Wu, J. Gong, and Z. Yang, “Cell zooming for cost-efficient green cellular networks,” *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 74–79, November 2010.
[37] 3GPP, “Views on small cell on/off mechanisms,” 3GPP TSG RAN WG1 R1-133456, Tech. Rep., August 2013.
[38] ——, “Enhancements of small cell on/off,” 3GPP TSG RAN WG1 R1-132888, Tech. Rep., August 2013.

[39] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 56–61, June 2011.

[40] A. Capone, I. Filippini, B. Gloss, and U. Barth, "Rethinking cellular system architecture for breaking current energy efficiency limits," in *Proc. of Sustainable Internet and ICT for Sustainability (SustainIT)*, October 2012, pp. 1–5.

[41] E. Ternon, Z. Bharucha, and H. Taoka, "A feasibility study for the detached cell concept," in *Proc. of 9th International ITG Conference on Systems, Communication and Coding*, January 2013, pp. 1–5.

[42] K. Sakaguchi, S. Sampei, H. Shimodaira, R. Rezagah, G. Tran, and K. Araki, "Cloud cooperated heterogeneous cellular networks," in *Proc. of International Symposium on Intelligent Signal Processing and Communications Systems (ISPACS)*, November 2013, pp. 787–791.

[43] F. Heliot, M. Imran, and R. Tafazolli, "On the energy efficiency-spectral efficiency trade-off over the MIMO Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1345–1356, May 2012.

[44] O. Onireti, F. Heliot, and M. Imran, "On the energy efficiency-spectral efficiency trade-off in the uplink of CoMP system," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 556–561, February 2012.

[45] C. Xiong, G. Li, S. Zhang, Y. Chen, and S. Xu, "Energy- and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874–3886, November 2011.

[46] B. Romanous, S. Zaidi, A. Imran, M. Ghogho, and H. Refai, "A game theoretic approach for optimizing density of remote radio heads in user centric Cloud-based radio access network," in *Under review IEEE Globecom Workshops*.

[47] S. Zaidi, A. Imran, D. McLernon, and M. Ghogho, "Characterizing coverage and downlink throughput of cloud empowered HetNets," *IEEE Commun. Lett.*, vol. PP, no. 99, pp. 1–1, February 2015.

[48] Ericsson, "More than 50 billion connected devices," White Paper, February 2011. [Online]. Available: http://www.ericsson.com/res/docs/whitepapers/wp-50-billions.pdf

[49] Nokia Siemens Networks, "2020: Beyond 4G, Radio Evolution for the Gigabit Experience," White Paper, August 2011. [Online]. Available: http://nsn.com/file/15036/2020-beyond-4g-radio-evolution-for-the-gigabit-experience

[50] International Telecommunication Union, "ITU-R M.2243: Assessment of the global mobile broadband deployments and forecasts for International Mobile Telecommunications," Tech. Rep., 2011.

[51] P. Mogensen *et al.*, "LTE capacity compared to the Shannon bound," in *Proc. of IEEE 65th Vehicular Technology Conference-Spring*, April 2007, pp. 1234–1238.

[52] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, February 2011.

[53] C. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, January 1949.

[54] 3GPP, "TR 36.814 v. 1.2.1, Further Advancements for EUTRA: Physical Layer Aspects," Tech. Spec., June 2009.

[55] G. Yuan, X. Zhang, W. Wang, and Y. Yang, "Carrier aggregation for LTE-Advanced mobile communication systems," *IEEE Commun. Mag.*, vol. 48, no. 2, pp. 88–93, February 2010.

[56] International Telecommunication Union, "Spectrum for IMT." [Online]. Available: http://www.itu.int/ITU-D/tech/MobileCommunications/Spectrum-IMT.pdf

[57] J. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, April 2012.

[58] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Veh. Technol. Mag.*, vol. 6, no. 1, pp. 37–43, March 2011.

[59] A. Khandekar, N. Bhushan, J. Tingfang, and V. Vanghi, "LTE-Advanced: Heterogeneous networks," in *Proc. of European Wireless Conference (EW)*, April 2010, pp. 978–982.

[60] D. Astely, E. Dahlman, G. Fodor, S. Parkvall, and J. Sachs, "LTE release 12 and beyond," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 154–160, July 2013.

[61] S. Mukherjee and H. Ishii, "Energy efficiency in the phantom cell enhanced local area architecture," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC)*, April 2013, pp. 1267–1272.

[62] P. Demestichas *et al.*, "5G on the horizon: Key challenges for the radio-access network," *IEEE Veh. Technol. Mag.*, vol. 8, no. 3, 2013.

[63] B. Yu, S. Mukherjee, H. Ishii, and L. Yang, "Dynamic TDD support in the LTE-B enhanced local area architecture," in *Proc. of IEEE Globecom Workshops*, December 2012, pp. 585–591.

[64] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, "LTE: the evolution of mobile broadband," *IEEE Commun. Mag.*, vol. 47, no. 4, pp. 44–51, April 2009.

[65] B. Yu, H. Ishii, and L. Yang, "System level performance evaluation of dynamic TDD and interference coordination in enhanced local area architecture," in *IEEE 77th Vehicular Technology Conference (VTC Spring)*, June 2013, pp. 1–6.

[66] ETSI, "Evolved universal terrestrial radio access (E-UTRA); Physical channels and modulation," Tech. Spec., January 2014, 3GPP TS 36.211 version 11.5.0 Release 11. [Online]. Available: http://www.etsi.org/deliver/etsi_ts/136200_136299/136211/11.05.00_60/ts_136211v110500p.pdf

[67] D. Lopez-Perez *et al.*, "Enhanced intercell interference coordination challenges in heterogeneous networks," *Wireless Commun.*, vol. 18, no. 3, pp. 22–30, June 2011.

[68] L. Garcia, K. Pedersen, and P. Mogensen, "Autonomous component carrier selection: interference management in local area environments for LTE-Advanced," *IEEE Commun. Mag.*, vol. 47, no. 9, pp. 110–116, September 2009.

[69] 3GPP, "RP-111369: Further enhanced non CA-based ICIC for LTE," Tech. Rep., 3GPP TSG RAN meeting 53, Fukuoka, Japan, September 13–16, 2011. [Online]. Available: http://www.3gpp.org/DynaReport/FeatureOrStudyItemFile-510021.htm

[70] A. Tyrrell *et al.*, "ARTIST4G project deliverable D2.2: Advanced receiver signal processing techniques evaluation and characterization," January 2011. [Online]. Available: https://ict-artist4g.eu/projet/work-packages/wp2/deliverables/d2.2/final/d2.2-1/at_download/file

[71] T. Ihalainen *et al.*, "Flexible scalable solutions for dense small cell networks," in *Proc. of Wireless World Research Forum*, 2013.

[72] A. Lobinger, S. Stefanski, T. Jansen, and I. Balan, "Coordinating handover parameter optimization and load balancing in LTE self-optimizing networks," in *Proc. of IEEE 73rd Vehicular Technology Conference (VTC Spring)*, May 2011, pp. 1–5.

[73] I. Godor *et al.*, "EARTH project deliverable D3.3: Final report on green network technologies," July 2012. [Online]. Available: http://bscw.ict-earth.eu/pub/bscw.cgi/d70472/EARTH_WP3_D3.3.pdf

[74] Y. Kishiyama, A. Benjebbour, T. Nakamura, and H. Ishii, "Future steps of LTE-A: evolution toward integration of local area and wide area systems," *Wireless Commun.*, vol. 20, no. 1, pp. 12–18, February 2013.

[75] 3GPP, "Study on small cell enhancements for E-UTRA and E-UTRAN: Higher layer aspects," Tech. Rep., December 2013, 3GPP TR 36.842 version 12.0.0 Release 12. [Online]. Available: http://www.3gpp.org/DynaReport/36842.htm

[76] S. Liu, J. Wu, C. H. Koh, and V. Lau, "A 25 Gb/s(/km2) urban wireless network beyond IMT-advanced," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 122–129, February 2011.

[77] 3GPP, Nokia Siemens Networks, "Mobility statistics for macro and small cell dual-connectivity cases," Tech. Rep., 3GPP TSG-RAN WG2 Meeting , Chicago, USA, 15–19 April 2013. [Online]. Available: http://www.3gpp.org/DynaReport/TDocExMtg--R2-81b--30048.htm

[78] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Mobility enhancements in heterogeneous networks," Tech Rep., December 2012, 3GPP TR 36.839 V11.1.0 Release 11. [Online]. Available: http://www.3gpp.org/DynaReport/36839.htm

[79] A. Mohamed, O. Onireti, S. Hoseinitabatabae, M. Imran, A. Imran, and R. Tafazolli, "Mobility prediction for handover management in cellular networks with control/data separation," in *Proc. of IEEE International Conference on Communications (ICC)*, June 2015.

[80] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez, "Show me how you move and I will tell you who you are," in *Proc. of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*, 2010, pp. 34–41.

[81] ——, "Next place prediction using mobility markov chains," in *Proc. of the 1st ACM Workshop on Measurement, Privacy, and Mobility*, 2012.

[82] Y. Chon, E. Talipov, H. Shin, and H. Cha, "Mobility prediction-based smartphone energy optimization for everyday location monitoring," in *Proc. of the 9th ACM conference on embedded networked sensor systems*, 2011, pp. 82–95.

[83] N. Amirrudin, S. Ariffin, N. Malik, and N. Ghazali, "User's mobility history-based mobility prediction in LTE femtocells network," in *Proc. of IEEE International RF and Microwave Conference*, December 2013, pp. 105–110.

[84] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer, "Semitri: a framework for semantic annotation of heterogeneous trajectories," in *Proc. of the 14th ACM international conference on extending database technology*, 2011, pp. 259–270.

[85] G. P. Fettweis, "A 5G wireless communications vision," *Microwave Journal*, vol. 55, no. 12, pp. 24–36, 2012.

[86] F. Khan, *LTE for 4G mobile broadband: air interface technologies and performance*. Cambridge University Press, 2009.

[87] S. Sesia, I. Toufik, and M. Baker, *LTE: the UMTS long term evolution*. Wiley Online Library, 2009.

[88] A. Mohamed, O. Onireti, Y. Qi, A. Imran, M. Imran, and R. Tafazolli, "Physical layer frame in signalling-data separation architecture: Overhead and performance evaluation," in *Proc. of 20th European Wireless Conference*, May 2014, pp. 820–825.

[89] A. Mohamed, O. Onireti, M. Imran, A. Imran, and R. Tafazolli, "Correlation-based adaptive pilot pattern in control/data separation architecture," in *Proc. of IEEE International Conference on Communications (ICC)*, June 2015.

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

[91] I. Siaud *et al.*, "MIWEBA project deliverable D1.1: Definition of scenarios and use cases," December 2013. [Online]. Available: www.miweba.eu/data/MiWEBA_D1.1_v1.0.pdf

[92] E. Ternon, P. Agyapong, L. Hu, and A. Dekorsy, "Database-aided energy savings in next generation dual connectivity heterogeneous networks," in *Proc. of IEEE Wireless Communications and Networking Conference*, April 2014.

[93] A. Redondi, I. Filippini, and A. Capone, "Context management in energy-efficient radio access networks," in *Proc. of 24th Tyrrhenian International Workshop on Digital Communications– Green ICT*, September 2013, pp. 1–5.

[94] T. Nakamura *et al.*, "Trends in small cell enhancements in LTE advanced," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 98–105, February 2013.

[95] A. Prasad and A. Maeder, "Energy saving enhancement for LTE-Advanced heterogeneous networks with dual connectivity," in *Proc. of IEEE 80th Vehicular Technology Conference VTC Fall*, September 2014, pp. 1–6.

[96] K. Sakaguchi *et al.*, "Millimeter-wave evolution for 5G cellular networks," *IEICE Transactions on Communications*, vol. E98-B, no. 3, March 2015.

[97] D. Bojic *et al.*, "Advanced wireless and optical technologies for small-cell mobile backhaul with dynamic software-defined management," *IEEE Commun. Mag.*, vol. 51, no. 9, pp. 86–93, September 2013.

[98] A. Prasad and A. Maeder, "Backhaul-aware energy efficient heterogeneous networks with dual connectivity," *Telecommunication Systems*, vol. 59, no. 1, pp. 25–41, 2015. [Online]. Available: http://dx.doi.org/10.1007/s11235-014-9893-4

[99] P. Mogensen *et al.*, "LTE-Advanced: The path towards gigabit/s in wireless mobile communications," in *Proc, of International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace Electronic Systems Technology*, May 2009, pp. 147–151.

[100] C. Yang, J. Wang, M. Wang, K. Zou, K. Yang, and M. Hua, "Over-the-air signaling in cellular communication systems," *IEEE Wireless Commun.*, vol. 21, no. 4, pp. 120–129, August 2014.

[101] O. Aydin *et al.*, "METIS project deliverable D4.3: Final report on network-level solutions," March 2015. [Online]. Available: https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D4.3_v1.pdf

[102] ETSI, "Self-configuring and self-optimizing network (SON) use cases and solutions," Tech. Rep., May 2011, 3GPP TR 36.902 version 9.3.1 Release 9. [Online]. Available: http://www.etsi.org/deliver/etsi_tr/136900_136999/136902/09.03.01_60/tr_136902v090301p.pdf

[103] O. Blume, H. Eckhardt, S. Klein, E. Kuehn, and W. M. Wajda, "Energy savings in mobile networks based on adaptation to traffic statistics," *Bell Labs Technical Journal*, vol. 15, no. 2, pp. 77–94, September 2010.

[104] R. Wu, Z. Wen, C. Fan, J. Liu, and Z. Ma, "Self-optimization of antenna configuration in LTE-Advance networks for energy saving," in *Proc. of IEEE 3rd International Conference on Broadband Network and Multimedia Technology*, October 2010, pp. 529–534.

[105] ETSI, "Potential solutions for energy saving for E-UTRAN," Tech. Rep., October 2012, 3GPP TR 36.927 version 11.0.0 Release 11. [Online]. Available: http://www.etsi.org/deliver/etsi_tr/136900_136999/136927/11.00.00_60/tr_136927v110000p.pdf

[106] H. Hu, J. Zhang, X. Zheng, Y. Yang, and P. Wu, "Self-configuration and self-optimization for LTE networks," *IEEE Commun. Mag.*, vol. 48, no. 2, pp. 94–100, February 2010.

[107] H. Ishii, X. Cheng, S. Mukherjee, and B. Yu, "An LTE offload solution using small cells with D2D links," in *Proc. of IEEE International Conference on Communications Workshops (ICC)*, June 2013, pp. 1155–1160.

[108] E. Dahlman, K. Dimou, S. Parkvall, and H. Tullberg, "Future wireless access small cells and heterogeneous deployments," in *Proc. of 20th International Conference on Telecommunications*, May 2013, pp. 1–5.

[109] Ericsson, "LTE Release 12," White Paper, January 2013. [Online]. Available: http://www.ericsson.com/res/docs/whitepapers/wp-lte-release-12.pdf

[110] J. Acharya, L. Gao, and S. Gaur, *Heterogeneous networks in LTE-Advanced*. John Wiley & Sons, 2014.

[111] 3GPP, China Telecom, "Discussion on dual connectivity," Written Contribution, 3GPP TSG RAN WG1 Meeting, Chicago, US, 9-15 April 2013. [Online]. Available: http://www.3gpp.org/DynaReport/TDocExMtg--R1-72b--30042.htm

[112] 3GPP, LG electronics, "RAN1 related issues for support of dual connectivity between macro cell and small cell," Written Contribution, 3GPP TSG RAN WG1 meeting, Chicago, USA 9-15 April 2013. [Online]. Available: http://www.3gpp.org/DynaReport/TDocExMtg--R1-72b--30042.htm

[113] 3GPP, Huawei, HiSilicon, "Feasible scenarios and benefits of dual connectivity in small cell deployment," Written Contribution, 3GPP TSG-RAN WG2 Meeting 81, St. Julians, Malta, 28 January - 1 February 2013. [Online]. Available: http://www.3gpp.org/DynaReport/TDocExMtg--R2-81--30047.htm

[114] C. Hoymann, D. Larsson, H. Koorapaty, and J.-F. Cheng, "A lean carrier for LTE," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 74–80, February 2013.

[115] S. Ye, S. H. Wong, and C. Worrall, "Enhanced physical downlink control channel in LTE advanced release 11," *IEEE Commun. Mag.*, vol. 51, no. 2, pp. 82–89, February 2013.

[116] E. Ternon, P. Agyapong, L. Hu, and A. Dekorsy, "Energy savings in heterogeneous networks with clustered small cell deployments," in *Proc. of 11th International Symposium on Wireless Communications Systems (ISWCS)*, August 2014, pp. 126–130.

Abdelrahim Mohamed (S'15) received his B.Sc. degree (first class) in Electrical and Electronics Engineering from the University of Khartoum, Sudan in 2011, and his M.Sc. degree (distinction) in Mobile and Satellite Communications from the University of Surrey, UK in 2013. He secured the first place in the Electronics Engineering department of the University of Surrey, UK during his M.Sc. studies and was involved in the FP7 CoRaSat project. He is currently a Ph.D. student in the Institute for Communication Systems (ICS) at the University of Surrey, UK. He is also involved in the system architecture and co-existence work area in the 5G Innovation Centre at Surrey. His main areas of research interest include radio access networks with control/data plane separation, energy efficiency, mobility management and cognitive radio.

Oluwakayode Onireti (S'11-M'13) received his B.Eng. degree in Electrical Engineering from University of Ilorin, Nigeria, in 2005. The M.Sc. in Mobile & Satellite Communications and Ph.D. degrees in Electronics Engineering from University of Surrey, UK, in 2009 and 2012, respectively. He secured a first class grade in his B.Eng. and a distinction in his M.Sc. degree. He is currently a research fellow at the Institute for Communication Systems (ICS), University of Surrey, UK. He has been actively involved in European Commission funded projects such as ROCKET and EARTH. He is currently involved in the QSON project, a project on Self Organizing Cellular Networks. His main research interests include self-organizing cellular networks, energy efficiency, MIMO, distributed MIMO and cooperative communications.

Muhammad Ali Imran (M'03-SM'12) received the M.Sc. (Distinction) and Ph.D. degrees from Imperial College London, London, U.K., in 2002 and 2007, respectively. He is currently a Reader (Associate Professor) in the Institute for Communication Systems (ICS - formerly known as CCSR) at the University of Surrey, UK. He is leading a number of multimillion international research projects encompassing the areas of energy efficiency, fundamental performance limits, sensor networks and self-organising cellular networks. He is also leading the new physical layer work area for 5G innovation centre at Surrey. He has a global collaborative research network spanning both academia and key industrial players in the field of wireless communications. He has supervised 20 successful PhD graduates. He has contributed to 10 patents and published over 150 peer-reviewed research papers including more than 20 IEEE Transactions. He has been awarded IEEE Comsoc's Fred Ellersick award 2014, FEPS Learning and Teaching award 2014 and twice nominated for Tony Jean's Inspirational Teaching award. He was a shortlisted finalist for The Wharton-QS Stars Awards 2014 for innovative teaching and VC's learning and teaching award in University of Surrey. He is a senior member of IEEE and a Senior Fellow of Higher Education Academy (SFHEA) UK.

Ali Imran is an assistant professor in telecommunications at University of Oklahoma. He is currently leading a multinational $1.045 million research project on Self Organizing Cellular Networks, QSON (www.qson.org). His research interest include, self-organizing networks, radio resource management and big data analytics. He has authored over 40 peer reviewed articles and has presented number of tutorials at international forums such as IEEE ICC, IEEE WCNC, European Wireless and CrownCom on these topics. He is an Associate Fellow of Higher Education Academy (AFHEA), UK and Member of Advisory Board to Special Technical Community on Big Data at IEEE Computer Society.

Rahim Tafazolli (SM'09) is a professor and the Director of the Institute for Communication Systems (ICS) and 5G Innovation Centre (5GIC), the University of Surrey in the UK. He has published more than 500 research papers in refereed journals, international conferences and as invited speaker. He is the editor of two books on "Technologies for Wireless Future" published by Wiley's Vol.1 in 2004 and Vol.2 2006. He was appointed as Fellow of WWRF (Wireless World Research Forum) in April 2011, in recognition of his personal contribution to the wireless world. As well as heading one of Europe's leading research groups.