

Separation Framework: An Enabler for Cooperative and D2D Communication for Future 5G Networks

Hafiz Atta Ul Mustafa, Muhammad Ali Imran, *Senior Member, IEEE*,
 Muhammad Zeeshan Shakir, *Member, IEEE*, Ali Imran, and
 Rahim Tafazolli, *Senior Member, IEEE*

Abstract—Soaring capacity and coverage demands dictate that future cellular networks need to migrate soon toward ultra-dense networks. However, network densification comes with a host of challenges that include compromised energy efficiency, complex interference management, cumbersome mobility management, burdensome signaling overheads, and higher backhaul costs. Interestingly, most of the problems that beleague network densification stem from legacy networks' one common feature, i.e., tight coupling between the control and data planes regardless of their degree of heterogeneity and cell density. Consequently, in wake of 5G, control and data planes separation architecture (SARC) has recently been conceived as a promising paradigm that has potential to address most of the aforementioned challenges. In this survey, we review various proposals that have been presented in the literature so far to enable SARC. More specifically, we analyze how and to what degree various SARC proposals address the four main challenges in network densification, namely: energy efficiency, system level capacity maximization, interference management, and mobility management. We then focus on two salient features of future cellular networks that have not yet been adapted in legacy networks at wide scale and thus remain a hallmark of 5G, i.e., coordinated multipoint (CoMP) and device-to-device (D2D) communications. After providing necessary background on CoMP and D2D, we analyze how SARC can particularly act as a major enabler for CoMP and D2D in context of 5G. This article thus serves as both a tutorial as well as an up-to-date survey on SARC, CoMP, and D2D. Most importantly, this survey provides an extensive outlook of challenges and opportunities that lie at the crossroads of these three mutually entangled emerging technologies.

Index Terms—Separation framework, decoupled architecture, cooperative communication, energy efficiency, coordinated multipoint, D2D communication.

I. INTRODUCTION

TRADITIONAL cellular networks are designed with tight coupling of control and data planes. This architecture con-

Manuscript received December 24, 2014; revised June 8, 2015; accepted July 17, 2015. Date of publication July 21, 2015; date of current version January 27, 2016. This work was made possible by NPRP grant No. 5-1047-2-437 from the Qatar National Research Fund (a member of The Qatar Foundation). The statements made herein are solely the responsibility of the authors.

H. A. U. Mustafa, M. A. Imran, and R. Tafazolli are with the Institute for Communication Systems (ICS), University of Surrey, Surrey GU2 7XH, U.K. (e-mail: h.mustafa@surrey.ac.uk; m.imran@surrey.ac.uk; r.tafazolli@surrey.ac.uk).

M. Z. Shakir is with the Department Electrical and Computer Engineering, Texas A&M University, Doha, Qatar (e-mail: muhammad.shakir@qatar.tamu.edu).

A. Imran is with the School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK 74135 USA (e-mail: ali.imran@ou.edu).

Digital Object Identifier 10.1109/COMST.2015.2459596

forms to the main objective of ubiquitous coverage and spectrally efficient voice-oriented homogeneous services. The recent growth of data traffic overwhelmingly brought a paradigm shift from voice-traffic to data-traffic. Cisco made observations at internet service providers and predicted that the annual global Internet traffic will rise to 1.4 zettabyte by the year 2017 as compared to 528 exabyte (EB) in 2012 [1]. One of the contributors in this massive growth of Internet traffic is the proliferation of mobile devices and machine-to-machine (M2M) communication. Due to this growth, the capacity and coverage requirements exploded in recent years with worldwide mobile traffic forecast of more than 127 EB in the year 2020 [2]. An increase of thousand-fold in wireless traffic is expected in 2020 as compared to 2010 figures [3] with expected figure of 50 billion communication devices [4]. The explosive growth of mobile traffic is being handled by deploying tremendous amount of small cells resulting in heterogeneous network (HetNet) [5].

The tight coupling of planes in conventional cellular networks leaves minimum control to consider networks' energy efficiency metric. This metric had a less concern previously due to less number of subscribers, rare data services, sparse deployments, and less awareness of green cellular communication. The green attribute of the cellular communication refers to reduction of unnecessary power consumption and its subsequent impact on the environment in the form of CO₂ emissions [6]–[11]. The green cellular communication can be realized by bringing energy-awareness in the design, in the devices [12] and in the protocols of communication networks. Due to the network scaling and heterogeneity (large number of small cell deployments), this metric became prominent. In this regard, it has been estimated that the energy consumption by the information and communications technology (ICT) results in 2% of global carbon emissions [13].

Small cell deployment is an agile, cost-effective, and energy efficient solution to meet coverage and capacity requirements. However, large number of deployments (e.g., prediction of 36.8 million small cell shipments by year 2016 according to ABI research [14]), the energy efficiency gain due to small cells might be compromised. Moreover, it also poses operational expenditure (OPEX) challenges to the network operators. This heterogeneity has also imbalanced the provision of data services between macro and small cells resulting in severe interference/backhaul-limited communication. In order to overcome the threatening issues of power consumption, the awareness of energy consumption has already been realized and

a number of energy conservation techniques/approaches have been investigated in the literature.

Another core issue, rising in future ultra-dense HetNet, is the interference management. The main limiting factor in achieving the optimum capacity is intra/inter-cell interference. Although intra-cell interference, in present cellular networks, has been eliminated by using orthogonal frequency division multiple access (OFDMA) technology and radio resource management (RRM), provision of underlay co-existing networks (e.g., device-to-device (D2D), M2M), in future ultra-dense environment will again cause intra-cell interference along with existing inter-cell interference. Current interference management techniques mainly comprise mitigation, cancellation, and coordination. The first two techniques are best suited to a single cell environment, whereas for multicell scenarios, coordination techniques comprising inter-cell interference coordination (ICIC), enhanced ICIC (eICIC), coordinated beamforming (CB), and coordinated multipoint (CoMP) are more promising to provide homogeneous quality of service with small infrastructural changes over the area [15]. The ICIC techniques were introduced to mitigate inter-cell interference for cell-edge users. The main idea is to use either different set of resource blocks (RBs) throughout the cell or partition RBs for cell-centre and cell-edge users. In another scheme of ICIC techniques, this RB partitioning can be coupled with different power levels (e.g., power boost for cell-edge users and low power for cell-centre users) to mitigate inter-cell interference. The ICIC techniques have been enhanced to eICIC for HetNet in 3rd generation partnership project (3GPP) Rel-10. These techniques, unlike ICIC, consider both control and traffic channels either in time, frequency or power domains to mitigate inter-cell interference. The main idea of eICIC is based on almost blank subframe (ABS). These blank subframes are reserved for different purposes for macro tier and small cell tiers. The macro tier mostly uses these subframes for control channels with low power, whereas small cell tiers use them for traffic channels to serve cell-edge users.

The CB and CoMP fall into the category of interference exploitation as compared to interference avoidance schemes (e.g., ICIC, eICIC). In such techniques, joint scheduling, transmission, and processing are carried out to exploit inter-cell interference and enhance cell-edge performance. In CB, the user equipment (UE) is served by a single base station (BS), however, interference is coordinated between cooperating BSs. To enhance the data rate of individual UE, it can cooperatively be served by a number of BSs in CoMP, however, this approach requires sharing data between cooperating BSs which results in huge backhaul capacity requirements. As compared to interference avoidance, the exploiting techniques (e.g., multicell cooperation (CB and CoMP)) have been identified as a key solution in long term evolution (LTE) and LTE-Advanced (LTE-A) to improve the cell-edge performance, average data rate, and spectral efficiency by mitigating and exploiting inter-cell interference [16]–[18].

The green aspects of future 5G cellular networks require energy efficient communication which can be realized effectively by completely switching-off under-utilized BSs. However, the switch-off mechanism has severe limitations in current cellular

architecture due to coverage holes. In order to avoid coverage holes, one of the candidate solution is the new cellular architecture where control and data planes are separated, i.e., decoupled or separation architecture, to provide ubiquitous coverage and more localized high-rate data services. Another potential advantage in this architecture is the flexible mobility management due to reduced handover signaling. In present architecture, the mobile user is handed over to nearby BS even if there is no active data session. Since, control plane is coupled with data plane, it is mandatory to handover in-active mobile terminals to ensure coverage. This results in handover signaling which is required for coverage but not for data services. On the other hand, the mobile user with no active data session in decoupled architecture can move freely without initiating handover due to ubiquitous coverage. Huge potential savings can be realized in this case, due to reduced handover signaling resulting in energy efficient communication.

In order to realize thousand-fold capacity enhancements in future cellular networks, much higher bandwidth is required. This higher bandwidth is available in millimeter wave (mm-Wave) spectrum. The higher frequency has poor propagation characteristics, however, the corresponding spot-beam coverage is more feasible for low-range high-rate data services. Therefore, coverage at lower frequencies (with good propagation characteristics) and high-rate data services (with limited coverage) requires decoupled architecture. Another aspect that severely limits the system capacity is the ultra-dense cellular environment in future networks (due to more granular tiers in the form of D2D, and M2M overlay/underlay communication). The underlay system offers higher system capacity but causes intra-cell interference and therefore, interference management becomes more complex in this case. For such an environment, cooperation and coordination is the promising solution for interference management in decoupled architecture.

Keeping in view the above vision, we structure the article in three sections. The first section introduces separation framework and provides survey of existing literature on separation architecture. Since energy efficiency is the key enabler for separation framework, we provide extensive literature review of existing approaches that realize energy efficient communication in current cellular architecture. This is followed by highlighting future requirements of cellular networks from the perspective of system capacity, interference management, and mobility management. We highlight several shortcomings due to coupled planes and provide motivation for separation architecture. The shortcomings in current architecture and potential gains due to decoupling are tabulated at the end of first section. The second section provides a brief tutorial on cooperative communication including underlay D2D cooperation. This section serves as a background to discuss cooperation in separation architecture. The third section presents different scenarios where cooperative communication can be realized in separation framework by highlighting potential advantages and associated complexities. The article is organized as follows. We provide the list of acronyms in Table I. Section II provides system performance reviews of traditional and separation architecture. In Section III, the general context of cooperative communication, clustering, and D2D communication for traditional cellular system

TABLE I
LIST OF ACRONYMS

Acronym	Definition
3GPP	3 rd Generation Partnership Project
ABRB	Almost Blank Resource Block
ABS	Almost Blank Subframe
BBU	Base Band Unit
BS	Base Station
C-RAN	Cloud Radio Access Network
CARC	Conventional Architecture
CB	Coordinated Beamforming
cBS	Control BS
CCU	CoMP Central Unit
CoMP	Coordinated Multipoint
CRC	Cyclic Redundancy Check
CRS	Common Reference Signal
CSI	Channel State Information
CSI-RS	Channel State Information Reference Signal
CU	Central Unit
dBS	Data BS
eICIC	enhanced Inter-cell Interference Coordination
eLA	enhanced Local Area
eNB	evolved NodeB
FDD	Frequency Division Duplex
HARQ	Hybrid Automatic Repeat Request
HetNet	Heterogeneous Network
ICIC	Inter-cell Interference Coordination
IMT-Advanced	International Mobile Telecommunications-Advanced
ISM	Industrial, Scientific, and Medical
JD	Joint Detection
JT	Joint Transmission
LTE	Long Term Evolution
LTE-A	Long Term Evolution-Advanced
MIMO	Multiple Input Multiple Output
OFDMA	Orthogonal Frequency Division Multiple Access
PID	Physical Cell Identification
PSS	Primary Synchronization Signal
RAN	Radio Access Network
RB	Resource Block
RRH	Remote Radio Head
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
SARC	Separation Architecture
SC-RAN	SARC in C-RAN
SON	Self-organizing Network
SSS	Secondary Synchronization Signal
TDD	Time Division Duplex
UE	User Equipment
UMTS	Universal Mobile Telecommunications System

has been presented. Section IV describes different perspectives to extend cooperative communication to the separation framework. Section V concludes the survey and highlights future research directions in this area.

II. SEPARATION FRAMEWORK: PERFORMANCE MEASURES AND POTENTIAL GAINS

The current cellular networks comprise tightly coupled control and data planes in the same radio access network (RAN). This architecture meets the main objective of ubiquitous coverage and spectral efficiency for voice services in homogeneous deployments. The massive growth of data traffic overwhelmingly dominated the voice traffic resulting into a paradigm shift from homogeneity to heterogeneity and voice services to data services. The traditional architecture (designed for homogeneous voice services) meets the current requirements of ubiquitous coverage and high spectral efficiency, however, it provides these services by overlooking signaling overheads, backhaul cost, and energy efficiency of the system. In order to enhance the coverage and capacity of current cellular systems, it is common practice to deploy small cells for peak-load scenarios at the cost of reduced energy efficiency, increased overhead signaling (e.g., in terms of frequent handovers) and increased backhaul requirements. In order to mitigate the rising concerns of power consumption, number of solutions, based on dynamic BS switching mechanism, are suggested to exploit the temporal and spatial variations in traffic load. However, the tight coupling of user and control planes restricts the flexibility and leaves less degree of freedom to optimize the system performance (discussed in subsequent subsections). To this end, the idea of control and data planes separation was proposed by the project beyond green cellular generation (BCG2) of GreenTouch consortium in Jan., 2011 [19]. Similar approaches have been suggested in study group of 3GPP on “New Carrier Type.” The Mobile and wireless communications Enablers for Twenty-twenty Information Society (METIS) [20] aims to lay the foundation of 5G where control and data plane separation is being considered as a candidate system architecture. The green 5G mobile networks (5grEEn) is focusing on green aspects of future 5G networks by considering separation of control and data planes. The joint European Union-Japan project Millimeter-Wave Evolution for Backhaul and Access (MiWEBA) is investigating the use of separated control and data planes for mm-Wave based small cells [21].

In order to highlight potential gains due to decoupling of control and data planes, we present conventional architecture (CARC) and futuristic separation architecture (SARC) in Fig. 1. As shown in Fig. 1(a), CARC is a conventional HetNet (comprises macrocell and large number of small cells) where coverage and data services are simultaneously provided at same frequency either by macro or small cell on coupled control and data planes. The advantage of this approach is ubiquitous coverage, however, the serving cell cannot sleep and it has to provide coverage even at low load conditions resulting in under-utilization of resources. The mobile users, irrespective of active or in-active sessions, are always covered by dedicated channels (ubiquitous coverage). However, it results in under-utilization of data plane (since it is coupled with control plane). In Fig. 1(b), SARC is a hierarchical HetNet comprising conventional HetNet and an additional tier of D2D/M2M communication, where control and data planes are decoupled. In such an architecture, the ubiquitous coverage and low-rate data

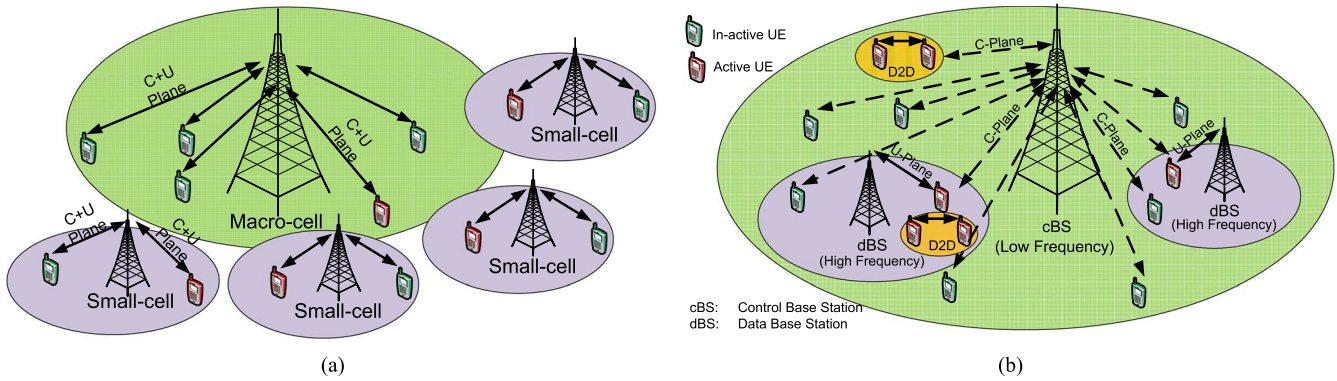


Fig. 1. Conventional and control/data plane separation architecture. (a) Conventional architecture (CARC). (b) Separation architecture (SARC).

services¹ are provided by control BS (cBS) at lower frequency bands with good channel characteristics. The data services are provided on demand at higher frequency bands by short range high-rate data BSs (dBSs). The advantages of this architecture are ubiquitous coverage (by decoupled control plane for active or in-active users), small cell sleeping possibility without coverage holes, temporal and spatial traffic adaptation, and high-rate data services for active users without compromising the energy efficiency of the system. The reader is referred to [22] for feasibility study of detached cells from the perspective of reliability and energy savings.

The control plane is responsible for system configuration and management. It provides system information, synchronization, and reference signals etc. The system information is broadcast and it mainly comprises the information required to join the network. The synchronization information includes frame timings as well as symbol level timings. The reference signals are used to know channel state which is indispensable for scheduling and resource allocation. In contrast to this, data plane is responsible to provide the requested contents along with some acknowledging mechanism (e.g., hybrid automatic repeat request (HARQ)). In order to give insights into information exchanges in both the planes, we provide a case study of LTE/LTE-A networks in Table II.

The SARC for HetNet offers many potential gains such as energy efficiency, capacity enhancement, reduced overhead signaling, flexible interference and mobility management. Control signaling is provided by cBS, however, certain types of control signaling cannot be fully decoupled. For example, frame/symbol level synchronization and channel state information (CSI) is required in both planes.

The separation of planes for future cellular networks has been realized very recently. To this end, the control and data plane separation has been suggested in [23], [24], where the provision of coverage has been provided by a long range low rate control evolved Node B (eNB). The data services, on the other hand, are provided by dedicated data eNBs. In [23], it is proposed that signaling will provide wider coverage to all UEs regard-

less of active or in-active data session under data eNB. Such network-wide adaptation provides flexibility to power down certain BSs when no data transmission is needed. In simple strategy of powering down the dBSs, neither control signaling (e.g., synchronization, reference signals, system information etc) nor associated backhaul to the access network is required; no data services are requested by UEs, only coverage is required which is ubiquitously provided by the cBS. The powering down strategy can, therefore, save approximately 80% of RAN power per BS switch-off [23], [25], [26] besides power savings due to backhaul communication links. Therefore, separation of planes promises tremendous increase in energy efficiency, reduced overhead signaling, and relaxed backhaul requirements. In [23], the energy efficiency gain has been emphasized by considering system level approach where under-utilized BSs are realized in sleep mode. In this study, no expected gains in energy efficiency are highlighted. Certain technical challenges including context awareness, resource management, and radio technologies for the signaling network are highlighted without proposing any design guidelines for the separation architecture.

The design of the signaling network in SARC is more challenging as compared to conventional approach. In CARC, the BSs usually do not sleep due to the possibility of coverage holes. Therefore, all BSs are active and no wake-up signaling is required. The handover procedure is usually UE driven based on reference signal received power (RSRP) values. In contrast, data services in SARC, in case of sleeping dBS, can be ensured by (i) optimal dBS selection from sleeping dBSs, and (ii) initiation of wake-up mechanisms. The optimal dBS selection can be quite challenging since cBS has no instantaneous knowledge of channel conditions. This results in more complex signaling procedures as compared to CARC. The new design is required to be robust and energy efficient. Use of low frequencies provides better propagation and obstacle penetration. Moreover, mobility management is flexible in HetNet using SARC architecture. This is because, control plane handover is rarely required since the coverage area of cBS is large as compared to the coverage area of BSs in conventional system. The data plane handover is only required in case of active data requests and in case of in-active users, none of the handovers (control plane or data plane) are required. This has been discussed in more details in Section II-D2.

¹The control BS has ubiquitous coverage over a large area as compared to small cell coverage area. Hence, it is more feasible to provide data services to high mobility users by cBS to avoid signaling overhead and frequent handovers in small cells.

TABLE II
CONTROL DATA PLANE INFORMATION EXCHANGE IN LTE/LTE-A

Signals	Information Exchange	Direction	Plane
Physical random access channel (PRACH)	Initial synchronization with eNode B (eNB).	Uplink	Control
Reference Signals (RS)	Demodulation RS (DRS) - Channel estimation for coherent demodulation, Sounding RS (SRS) - Channel quality estimation over a span of bandwidth.		
Physical uplink control channel (PUCCH)	(HARQ ACK/NACK)*, channel quality indicator (CQI), precoding matrix indicator (PMI), rank indicator (RI), scheduling requests.		
Physical uplink shared channel (PUSCH)	User uplink data.	Downlink	Data
Synchronization	Primary and secondary synchronization (PSS, SSS) for cell identity and frame timing.		
Reference Signals (RS)	Channel state information (CSI-RS), demodulation (DM-RS), cell-specific (CRS), positioning (PRS).		
Control Indicators	Physical control format indicator channel (PCFICH) to indicate size of PDCCH, physical HARQ indicator channel (PHICH) to ACK/NACK user data on PUSCH.		
Multicast/Broadcast	Physical broadcast channel (PBCH) carrying master information block (MIB), multicast/broadcast single frequency network (MBSFN), multicast channel (PMCH).		
Physical downlink shared channel (PDSCH)	User multiplexed data.	Data	

* HARQ is sent either as a feedback message on control channel or piggybacking feedback on user's data plane.

In [24], a two-layer network functionality separation scheme, targeting low control signaling overhead and flexible network reconfiguration for future green networks has been proposed. A frame structure level detail has been proposed in which network functionality including synchronization, system information broadcast, paging, and multicast (synchronization, pilot, frame control, and system/paging/multicast information bearer signals) is incorporated in control network layer (CNL). Whereas, the network functionality of synchronization and unicast (synchronization, pilot, frame control, and unicast information bearer signals) is incorporated in data network layer (DNL). In this study, the main focus is given on advantages of low control signaling overhead. The network area power consumption has been plotted for two architectures showing significant potential gain for separation architecture leading towards future energy efficient green mobile networks. Unlike [23], the authors in [24] proposed abstract level network design for control and data planes separation. The categorization of different wireless signals and their mapping relationship with physical channels are presented. However, the challenges highlighted in [23] are not discussed in [24]. The study also lacks in addressing interference management issues, backhaul requirement, realization of underlay networks (e.g., D2D), mobility management and corresponding handover procedures in separation architecture.

The important focus areas for energy efficient 5G mobile network are highlighted in [27]. These areas include system architecture with decoupled control and data planes, ultra-dense HetNet deployment, radio transmission using multiple input multiple output (MIMO) configuration and energy efficient backhaul. The transmission planes are categorized into data, control, and management planes. It is emphasized that if these planes are decoupled from each other then independent scaling is possible at most energy efficient locations. Furthermore, the logical separation of control and data planes can provide most efficient discontinuous transmission/reception (DTX/DRX) functionality to save energy in idle modes. Similar to [23], the authors in [27] highlighted the requirements and technical challenges to realize future green 5G mobile network. However, the system architecture and radio transmissions

design guidelines are not outlined in details as in [24]. The solutions to these important areas are considered as deliverables of 5GrEEn.

In [28], hyper-cellular network is introduced as decoupled control and traffic network to realize energy efficient operation of BS. In such a network, data cells are flexible to adapt traffic variations and network dynamics while control cells can flexibly and globally be optimized. The hyper-cellular network is considered as a novel architecture for future mobile communication systems. The approach realizes control and data planes separation using open source radio peripherals and legacy global system for mobile (GSM) network. In this testing, signaling BS provided coverage whereas data BS ensured phone call connectivity. A very promising formulation has been setup by using open base transceiver station (OpenBTS), universal software radio peripheral (USRP) front end, wide bandwidth transceiver (WBX) daughter board, and dell PCs. This formulation provides an insight into real-time practical setup for prototype testing. However, system improvements are not shown in this paper. Moreover, none of the performance metrics (energy efficiency, backhaul relaxation, and throughput) have been analyzed and validated for this simple and basic approach.

The control and data planes separation concept has been presented from the perspective of energy optimized connectivity management in seventh framework programme (FP7) CROWD [29]. To this end, software defined networking (SDN) based medium access control (MAC) and mobility management has been proposed to complement huge deployments of cellular nodes. Two key challenges, interference and mobility management, are considered for next generation dense wireless mobile networks. The functional architecture has been proposed and several key control applications are identified. More focus is given on mobility management and an SDN-based distributed mobility management (DMM) approach has been suggested. The control applications for interference management range from existing multi-tier scheduling scheme (e.g., eICIC) to LTE access selection schemes. The radio transmission aspects and backhaul limitations have not been outlined in any of the control applications identified in this study.

TABLE III
SUMMARY OF APPROACHES FOR CONTROL AND DATA PLANES SEPARATION

Project/Paper/Ref.	Aim	Working/Highlights	Impacts/Conclusion
“Looking beyond green cellular networks” [23]	To switch-off BSs flexibly in case of no data transmission	⇒ Coverage → by long-range low-rate control eNB ⇒ Data Service → by short-range high-rate data eNBs. ⇒ Ubiquitous coverage by signaling plane.	<ul style="list-style-type: none"> • Selection and activation of BS is not a difficult task compared to optimizing the decision process.
“On Functionality Separation for Green Mobile Networks: Concept Study over LTE” [24]	To reduce control signaling overhead & realize flexible network reconfiguration	⇒ Separation scheme based on two-layer network functionality: CNL/DNL ⇒ CNL → multicast information bearer signals ⇒ DNL → unicast information bearer signals	<ul style="list-style-type: none"> • Rare Handover in CNL • Call re-establishment is not required • HO signaling is reduced significantly
“5GrEEn: Towards Green 5G mobile networks” [27]	To provide general outlook on system architecture for energy efficient 5G network	⇒ Ultra-dense HetNet deployment ⇒ Radio transmission using MIMO configuration ⇒ Energy efficient backhaul ⇒ Transmission planes: data, control, and management	<ul style="list-style-type: none"> • Separation of control and data plane provides most effective DTX/DRX functionality to save energy in idle modes
“Software defined radio implementation of signaling splitting in hyper-cellular network” [28]	Energy efficient operation of BS	⇒ Hyper cellular network: Decoupled signaling and data services ⇒ Handset is provided coverage by signaling BS ⇒ Phone calls are connected with the help of data BS	<ul style="list-style-type: none"> • Provides an insight into real-time practical setup for prototype testing
“FP7 project CROWD” [29]	Energy optimized connectivity management	⇒ SDN based MAC control and mobility management	<ul style="list-style-type: none"> • Complements huge deployments of cellular nodes
“Dual connectivity in LTE HetNets with split control- and user-plane” [30]	Dual connectivity and use of CSI-RSs for CSI measurements	⇒ Different MA small cells are considered as different antennae of MIMO/CoMP array	<ul style="list-style-type: none"> • CSI-RS is also used to estimate the downlink path loss for uplink power control
“A novel architecture for LTE-B: C-plane/U-plane split and Phantom Cell concept” [31], [32]	To provide high data rate to UE through spatial reuse of spectrum	⇒ Phantom Cell architecture: high frequency band solution with decoupled control/data plane ⇒ Macro cell controls the small cells for connection establishment ⇒ Small cells use high frequency bands to provide high-rate data coverage	<ul style="list-style-type: none"> • Outperforms conventional small cell architecture in both spectral and energy efficiency metrics

In [30], the authors measure CSI by using the concept of dual connectivity (using macrocell assisted small cells) and proposing the use of CSI reference signals (CSI-RSs) instead of common reference signal (CRS). Since, CSI-RSs are traditionally used by UEs to differentiate between different antennas of a MIMO system, therefore in the proposed network layout, different macrocell assisted small cells are considered as different antennae of MIMO/CoMP array. This strategy results in energy efficient operation (by reducing number of CRS) and provides network-triggered handover (unlike UE-triggered handover in CARC) to realize flexible and enhanced mobility management. Due to the absence of CRS for macrocell assisted small cells, the authors proposed to use CSI-RSs to estimate the downlink path loss for uplink power control. Similar to the previous approaches, the authors in [30] focuses only on reducing control signaling to realize energy efficient operation without emphasizing context awareness, radio frame structure, backhaul issues, and interference management.

The 3GPP is presently standardizing enhanced local area (eLA) small cell HetNet (LTE Rel-12) to provide high data rate to UEs through spatial reuse of the spectrum. In [31], a particular eLA architecture called Phantom Cell is proposed by NTT DOCOMO. This architecture is based on control and data planes separation; suggested as a novel architecture for LTE-B. The approach in [31] suggests deployment of massive small cells by leveraging high frequency reuse under the coverage of macrocell to achieve high capacity, seamless mobility, and scalability. The two tier configuration is realized as a master-slave configuration where macrocell controls the small cells

dynamically for connection establishment and small cells use high frequency bands to provide high-rate data coverage. This high frequency band solution with decoupled control and data planes, where small cells do not transmit cell-specific reference signals, is introduced as Phantom Cell architecture. In order to evaluate the energy efficiency performance of the Phantom Cell architecture, the stochastic geometry is used to compare the results with the conventional frequency division duplex (FDD) based LTE picocell deployment in [32]. The numerical results indicate that the Phantom Cell architecture outperforms conventional small cell architecture in both spectral and energy efficiency metrics. The authors in [31] provide preliminary results for capacity enhancements in separation architecture without considering energy efficiency aspects, whereas [32] provides more rigorous analysis for both spectral and energy efficiency of separation architecture. Some interesting conclusions are made about higher spectral efficiency and higher energy efficiency, however, both these studies focused on spectral and energy efficiency metric and did not include other aspects such as context awareness, signaling network, and functional description of the separation architecture. The reader is referred to [33], [34] for Phantom cell operation at super high and extremely high frequency and related technical issues such as larger path loss in small cell, human body shadowing, massive MIMO architecture, and precoding algorithms to achieve super high data rates. The comparative summary of different approaches for control and data planes separation is presented in Table III.

In the following subsections, we provide motivation for control and data planes separation architecture. In this context, we

consider several key performance measures and analyze them in existing architecture. We provide survey of existing approaches, highlight the shortcomings and discuss these measures from the perspective of SARC architecture.

A. Energy Efficiency

The energy efficiency of RAN mainly depends on power consumption of BS. According to energy aware radio and network technologies (EARTH) project [25], the BS power consumption model comprises power consumed by radio frequency chain (especially power amplifier), signal processing units, and supply units (mains supply, DC-DC, and active cooling) as follows:

$$P_{BS} \propto (P_{RFC}, P_{SPU}, P_{SU}),$$

In order to ensure energy efficient communication, one simple strategy can be adopted where under-utilized BS, in case of low traffic conditions, should go to sleep mode (hence reducing power consumption P_{RFC} and P_{SU}). This situation, however, causes coverage holes due to tight coupling of control and data planes unlike futuristic architecture where coverage and data services will be decoupled to provide ubiquitous coverage and on-demand data services.

The power consumption had not been a problem in past due to homogeneous networks and sparse deployments. Therefore, energy efficiency metric had not been considered while designing such cellular networks. Due to technology scaling and proliferation of large number of smart devices, the capacity demands increased tremendously with more energy consumption worldwide. This huge increase in capacity was predicted by wireless world research forum (WWRF) more than a decade ago. The key technological vision from WWRF expected around 7 trillion wireless devices serving 7 billion people by 2017. Moreover, it was predicted that approximately 80–95% subscribers will be mobile broadband users [35], [36]. The huge increase in number of subscribers motivated the network operators to deploy small cells in order to quickly meet the customer needs. According to ABI research, by 2016, small cells will cover up to 25% of all mobile traffic and small cells shipments (both indoor and outdoor) will likely to reach 36.8 million units worth \$20.4 billion. It further predicts that, outdoor small cell units alone will reach over 3.5 million units by 2018 [14], [37]–[39]. The coverage and capacity requirements of subscribers can be met by deploying increased number of small cells, however, the associated power consumption will increase significantly in future.

In order to reduce the power consumption of under-utilized BSs and ensure energy efficient communication in existing HetNet, different techniques are reported in literature such as dynamic BS switch-off, cell range expansion etc. These techniques provide substantial gain in power saving, however, they come with the inherent problem of coverage holes (in case of BS switch-off), increased interference (due to increased transmit power in cell range expansion techniques), and huge backhaul requirements. To address these problems and ensure energy efficient communication, a paradigm change in control

and data planes coupling has been suggested in literature and research community. This approach not only provides ubiquitous coverage and reduced transmit power but also reduces control signaling associated with each BS.

The current cellular systems are designed for worst case ubiquitous coverage scenarios. In such a design, the BS needs to be active even for few subscribers. This goal can be justified in remote sparsely populated areas covered by few BSs where the spatio-temporal variations of traffic patterns follow a near constant trend. However, in urban areas, the BS deployment is dense and traffic variations are more abrupt. In such dense deployments, the coverage goal is achieved at the cost of increased power consumption and reduced energy efficiency of the system. The most power expensive element of RAN is BS, consuming around 80% of overall power [23], [26]. In full-load conditions, the power consumption of BS is justified, however, in low load conditions, BS is still consuming most of the power to provide coverage. Moreover, in design of cellular systems, the short-term and long-term traffic variations (e.g., temporal effects on traffic loads due to day/night times and spatial effect due to weekends/weekdays) are not considered due to which the existing cellular networks cannot be fully optimized from this perspective.

1) *Evaluation Framework*: In order to quantify the power consumption of wireless networks, the EARTH project provides a holistic energy efficiency evaluation framework (E3F). This framework provides power consumption breakdown of each entity of RAN. A BS power model has been proposed that maps the radio frequency (RF) output power (radiated from the antenna) to the total supply power consumption of BS. The power consumption for macro, micro, pico, and femto cells are compared. The traffic models (short-term and long-term) are investigated to emphasize the energy saving potentials. The deployment areas of Europe are segregated into dense urban, urban, suburban, rural and sparse. The traffic variations for a single day are depicted to give an insight into the energy efficiency evaluation of the wireless cellular network. Number of key findings are presented as follows:

- On average, the vast majority of the resources are idle in wireless networks.
- The supply power scales linearly to the number of transmit/receive chains.
- The RF output power and power consumption of BS are nearly linear.
- For macro BSs, the consumption of power amplifier (PA) scales with BS load.
- For micro BSs, the PA scaling is present to a lesser extent, whereas for pico/femto BSs, this scaling is negligible.

It has been mentioned that DC power consumption of a typical 3-sector site at zero load is still 50% of the peak power [40]. The conventional model without power supply and active cooling/air conditioning can be 400 W lower than the total power consumption of a site [41]. In [42], a parameterized linear power model is proposed to encompass the two general power saving techniques that are based on either design change or operating procedures. The former is based on changing the layout of the network (e.g., by introducing HetNet) whereas the

latter is more attractive for existing architecture. This approach saves energy by reducing transmission power, adapting transmission bandwidth, deactivating unused antennas, and incorporating BS sleep modes.

The model presented in [42] is the simpler parameterized model of [25]. The authors did not discuss the implications of coverage holes due to sleep mode operation.

2) *BS Switch-Off*: Power consumption of cellular systems has been addressed from two perspectives. The first one motivates the use of low-powered components in cellular networks and hence focuses on reducing the energy consumption at local scope. The second perspective takes the holistic approach of network design, planning, and management phases to conserve the energy of the overall cellular network. In both cases, the most power expensive element in access network is BS. A lot of research has been carried out to propose switch-off mechanisms for BSs. In [43], BS switch-off has been proposed by quantifying the reduction in activity probability for cooperative scenario. It has been shown that for a fixed distance between BSs, the expected number of enabled BSs reduces up to 11% depending on the user density. By changing the distance between cooperative BSs to an optimal value, an additional 39% reduction in activity probability can be achieved which results corresponding reduction in power consumption per unit area. The proposed analysis assumed perfect hexagonal grid which is non-realistic in practical BS deployments. Moreover, finding the optimal distance and changing the BS deployment is practically infeasible and very hard to realize. The authors in [44] suggest probabilistic data BS sleeping mechanism in separation architecture. The formulated problem jointly optimizes the sleeping probability and spectrum resource allocation to minimize the overall power consumption, however, this study does not consider mobility of the users and their impact on cell sleeping probability.

In [45]–[47], traffic profile based BS switching has been proposed to save energy. The cell switch-off has been suggested for cellular access networks [45] and universal mobile telecommunications system (UMTS) access networks [46] with the assumption that the radio coverage will be provided by neighboring cells by increasing transmit power. The smaller number of BSs for long-term switch-off vs. larger number of BSs for short-term switch-off have been investigated. However, these studies considered ideal networks (hexagonal and Manhattan models) and introduced the energy saving by dynamic switching algorithms. Though both of the approaches [45], [46] target to reduce power consumption, this strategy can cause severe interference to the neighboring active BSs due to increased transmit power. This can be ideal for the scenario where all neighboring BSs need switch-off which is not practical. The approach in [47] considered first and second order statistics of traffic profile to propose dynamic switching strategy. The users are handed over to the neighboring cells before the reference cell can be switched off. The statistics based switching strategy can save energy, however, it is suitable for near-constant traffic pattern (e.g., night times). In case of slowly varying traffic profile, an instantaneous switching strategy is more promising which can flexibly be realized in SARC. A simple approach may consider traffic profile and provide data service either by

cBS (in case of low-data rates) or by the near-by dBS (in case of high-data rates). In case of sleeping dBS, the cBS (having global coverage) may initiate wake-up mechanism which can be reactive or pro-active by predicting user mobility patterns. Since, no transmit power of dBS is increased, therefore an energy efficient communication, without increasing interference, can be achieved in SARC as compared to approaches proposed for CARC.

The macro BSs provide bigger coverage with high transmit power as compared to small cells. In order to conserve energy, the capacity enhancements are carried out by deploying large number of low-powered small cells. This brings heterogeneity in the network. For such networks, an area power consumption metric has been investigated in [48], [49], to quantify energy savings. The small cell deployment offers substantial power savings, however, this strategy scales poorly with number of small cells envisioned for future ultra-dense cellular environment. The scaling of small cells can be compensated by dynamic BS switch-off mechanisms which can be realized in separation architecture without producing coverage holes. The approaches in [48], [49] considered mixed deployment scenarios by considering macro and micro cells at fixed positions. This strategy is suitable for new deployments but it is not applicable to existing deployments of small cells. Assuming perfect hexagonal grid is a theoretical interest. These studies also lack in presenting realistic operating algorithm where area power consumption scales with any change in deployment e.g., due to network scaling or BS failure.

The BS switching-on/off based energy saving (SWES) algorithm has been proposed in [50] to exploit the temporal and spatial variation in the network traffic profile. The algorithm works in a distributed manner with reduced computational complexity. A notion of network-impact has been introduced that ensures minimal effects on neighboring BSs by turning off BSs gradually (one by one). In order to reduce overheads over the air and backhaul, three other heuristic versions of SWES are proposed that take network-impact as decision metric. The authors claim around 50–80% potential savings for real traffic profile of metropolitan urban area. Several extensions of this research are proposed as follows:

- To consider more realistic BS power consumption model.
- To consider HetNet, consisting of different types of BSs, such as macro, micro, femto BSs and even WiFi APs.
- To develop a dynamic BS switching algorithm that considers downlink and uplink traffics jointly.

Besides these extensions, the authors did not consider quality-of-service (QoS) requirements of the handed-over users. For example, in homogeneous deployments with large coverage area, cell-centre users have certain QoS requirements. In case, the serving BS of these users has to be switched-off, the neighboring BSs cannot guarantee same QoS without increasing the transmit power which results in inter-cell interference. This situation can be avoided in HetNet where neighboring small cells can cover handed-over users with moderate increase in transmit power, however, SWES techniques are proposed for only homogeneous deployments.

The theoretical framework for BS energy saving is presented in [51]. It encompasses dynamic BS operation, and related

problem of user association together. The problem is formulated as total cost minimization that allows for a flexible trade-off between flow-level performance (e.g., file transfer delay) and energy consumption. For user association problem, an optimal energy-efficient user association policy has been proposed, whereas for BS operation problem (i.e., BS switching-on/off), a simple greedy-on/off algorithm, based on mathematical background of sub-modularity maximization problem, is proposed. A number of heuristic algorithms, based on the distances between BSs or the utilization of BSs that do not impose any additional signaling overhead, are also proposed. The numerical results show 70–80% reduction in total energy consumption while depending on the arrival rate of traffic and its spatial distribution as well as the density of BS deployment. Unlike [50], the theoretical framework in [51] considers HetNet, however, to ensure mathematical tractability, no fast fading is considered and inter-cell interference is assumed as Gaussian-like noise which restricts practical realization of the proposed technique. Since an under-utilized BS consumes nearly the same power as a fully loaded BS [25], the logical solution to this problem is to switch off idle BSs while providing the same coverage and quality of the service. However, switching-off BSs will create coverage holes as the signaling and data services are provided by the same BS. A number of different techniques are proposed in literature to solve this problem. A paradigm shift in control and data planes coupling has been suggested in [23] where the coverage is provided by long-range BSs and high-rate data services are provided by small cell BSs. Hence, these short-range BSs can be activated/deactivated according to user demands without creating coverage holes.

3) *Renewable Energy Resources*: The energy efficiency of cellular systems has also been addressed using renewable energy sources. The cellular networks are scaled according to developed environment (e.g., urban, sub-urban, rural) and network traffic, however, the rural areas usually dominate on a country-wide coverage [25]. In developing countries, many remote locales do not have access to national electricity grid. To provide coverage in these areas, usually diesel is used as an energy source to operate BSs. The situation gets worse in low load conditions where the BS remains powered up to provide coverage for few active mobile terminals. The BS switch-off strategies cannot be adopted due to possibility of coverage holes in sparse deployments of BSs in remote areas of the country. In such cases, using renewable energy sources can be more advantageous. In [52], a reference model for renewable energy BS (REBS) has been suggested along with the concept of renewable energy-aware BS. The REBS comprises BS, energy control unit (ECU), and energy sources (renewable and non-renewable). The ECU is the important element that utilizes the energy storage unit in case of excess demand/supply and hence compensates the potential un-reliability of renewable energy sources. However, the presented reference model is very simple and the overall approach does not cover the complexities involved in designing ECU.

The renewable energy sources (solar, wind, fuel cell) are suggested in [53] for eco-friendly green 5G cellular networks. In the year 2004, Japanese cell phone operator NTT DOCOMO operated an experimental 3G BS (DoCoMo Eco Tower). This

self-powered tower used solar and wind power simultaneously [54]. In the year 2010, world wide fund for nature (WWF) annual report [55] was published showing substantial reduction in CO₂ emission in China because of using alternative energy sources.² In [56]–[58], an energy efficient communication and the dynamics of the smart grid are considered in designing green wireless cellular networks. The author in [56] proposed a novel game-theoretical decision making strategy to analyze the impact of smart grid on cellular network. The retailer and consumer are formulated as two players of a Stackelberg game. The proposed decision making scheme considers real time pricing in demand side management mechanism and gives insights into system parameters that affect the retailer's procurement and price decisions. The idea has been extended in [57] by considering CoMP to ensure QoS when certain BSs are switched off. Both of these strategies are further extended in [58] where service blocking probability is included in the system model. The analysis of the two player game has been enhanced by proving existence as well as uniqueness of the Stackelberg equilibrium. Though the approaches of energy efficient smart grid communication ensures reduction in OPEX and CO₂ emissions, the inherent problem of coverage holes due to BS switch-off, more control signaling at air interface due to coupled planes, and much higher backhaul requirements in case of CoMP operation renders such approaches impractical. Moreover, using CoMP to provide coverage for all users of switched-off BS can cause severe blockage and poor QoS. This is because CoMP has originally been designed to ensure cell-edge coverage not for the coverage of all users due to severe backhaul capacity limitations.

The existing approaches for CARC ensures power savings, however, all these approaches have certain shortcomings discussed previously. For example, BS switch-off mechanism in CARC causes coverage holes and in order to provide coverage by the cell range expansion techniques, the transmit power of covering BS increases resulting in inter-cell interference. The existing energy efficient approaches for CARC and the corresponding shortcomings are summarized in Table IV. The problem of coverage holes and subsequent problem of increased transmit power does not exist in SARC due to inherent ubiquitous coverage of cBS. Similarly, the problem of continuous operation of sparsely deployed BSs in remote locales of the country can best be tackled by providing data services by cBS during off-peak hours. Therefore, SARC can scale with two extreme load conditions (i.e., remote locales and ultra-dense environments).

B. System Capacity

In the past, voice services dominated data services due to which the cellular systems were mainly designed for the voice traffic. Such systems offered very low system capacity

²The use of solar and wind energy saved China 48.5 million metric tons of CO₂ emissions in the year 2008 and 58.2 million metric tons in the year 2009. Based on the result for China Mobile, and with conservative estimates, 70 million tons of carbon emission reductions had been estimated in the year 2008 which is equivalent to the total CO₂ emissions from countries like Sweden, Finland and Norway.

TABLE IV
SUMMARY OF APPROACHES FOR ENERGY EFFICIENT COMMUNICATION

General Approach	Proposed Technique	Shortcomings	Ref.
Dynamic BS Switch-off	Reduction in activity probability and cooperation/coordination	Not suitable for already deployed BSs High backhaul capacity requirements	[43]
	Probabilistic sleeping mechanism	Mobility impact not considered	[44]
	Cell range expansion	Inter-cell interference due to high transmit power	[45], [46]
	First/second order statistics of traffic profile	Not suitable for varying traffic patterns	[47]
	Small cell deployment	Scales poorly with number of small cell deployments	[48], [49]
	Temporal and spatial variations of traffic profile	Suitable for homogeneous deployment	[50]
	Flow level dynamics	Not suitable for fast fading channels Assumption of Gaussian-noise like inter-cell interference	[51]
Renewable Energy Resources	Alternate energy as main source	Not addressing problem of under-utilized network resources	[52]–[54]
Smart Grid	Game theoretical approach		[56]–[58]

TABLE V
TRAFFIC FORECAST

Category/Year	2010	2015	2020
Global Mobile Subscribers (Million)	5328	7490	9684
Total Mobile Traffic (EB)	4	45	127

complaint to the capacity requirements of voice services at that time. In the year 2009, the mobile data overtook the voice traffic in terms of total traffic generated on the network. With the emergence of mobile data services, the capacity requirements increased and the total worldwide mobile traffic is now expected to reach very high numbers. A brief view on number of worldwide mobile subscribers excluding WiFi traffic off-loading and including M2M communication [2] is shown in Table V.

The capacity requirements in terms of average area throughput for future mobile networks beyond international mobile telecommunications-advanced (IMT-Advanced) are studied to be 25 Gb/s/Km² [59] with peak data rate of 4.5 Gb/s/cell in downlink and 2.5 Gb/s/cell in uplink. The spectrum and bandwidth requirements for future IMT-2000 and IMT-Advanced are presented in [60]. Such high requirements and explosive growth of mobile data require huge system capacity. In literature, mainly three approaches are considered to meet the capacity requirements. These include spectrum efficiency, spectrum aggregation [61], and network densification [31]. The same has been identified by DOCOMO as “The Cube” for future 5G systems [62]. The spectrum efficiency targets the capacity enhancements by considering CB, multi-user MIMO, and CoMP. The spectrum aggregation includes carrier aggregation either contiguous or non-contiguous to meet the capacity requirements of different applications. However, the spectrum efficiency/aggregation have a local scope as compared to network densification that has been globally accepted as the cost-effective and agile solution to meet the capacity demands of future cellular systems. The huge number of small cell deployment results in heterogeneity in the network. This heterogeneity is expected to increase in the future by the increased number of D2D and M2M communications. In such ultra-dense HetNet, virtually a personal cell might be required in future to meet the

capacity and coverage requirements. The idea of personal cell has been introduced as pCell technology by Artemis Networks [63] where each wireless device will be provided the full bandwidth even in high load conditions and hence each mobile device will have virtually a dedicated personal BS. However, the pCell technology has yet not been commercialized.

In SARC, the capacity enhancements can be realized flexibly. For example, the spectral efficiency can be higher due to reduced control signaling interference. In CARC, every BS is responsible to provide control signaling as well as data services in its coverage area. Therefore, there will be as many control signaling interferers as there are BSs in specific area. In contrast to CARC, smaller number of cBSs will provide global coverage and hence control signaling interferers will be reduced in SARC. Moreover, due to sleeping dBSSs, there will be reduced inter-cell interference in data plane. The beam forming and CoMP can be realized centrally at cBS. The adaptive dBSS clustering for CoMP operation can be flexible by considering cell-sleeping into account (which is not possible in CARC). By having global coverage of the cBS, traffic off-loading may be realized by establishing D2D communication for common content exchange.

The network densification in SARC includes deployment of dBSSs in the coverage of cBS. In order to enhance capacity, dBSSs can be deployed at higher frequency bands with much more bandwidth. In this context, huge bandwidth at mm-Wave spectrum is an attractive choice for high-rate data transmissions [64]. In [65], the authors provide detailed design trade-offs and performance requirements to support wireless communication at 60 GHz frequency. The challenges associated with data transmission at this frequency include poor propagation, blocking/shadowing, atmospheric and rain effects [66], [67]. In order to model mm-Wave channel and analyze access performance, ray optics techniques have been used [68]. The ray tracing simulations at 72 GHz show that the propagation at such a high frequency can be approximated with limited diffraction and scattering phenomenon. The agreement between channel model and the measurement at mm-Wave band can also be observed in [69]. In [70], an air interface design, based on null cyclic prefix single carrier, has been proposed. The ray tracing results and the propagation measurements at mm-Wave show that it

is the best candidate for communication at this frequency. The measurement results at mm-Wave (28 and 38 GHz) spectrum with steerable directional antennas are presented in [71]. The novel hybrid beamforming scheme and mm-Wave prototyping for indoor and outdoor environment [72] asserts the feasibility of wireless communication at this frequency band. All these studies ensure that mm-Wave spectrum has potential gain to ensure high data rate transmission for dBSSs in future 5G cellular networks.

In future cellular communication, the mobile devices require several changes from the view point of hardware, software/firmware design, and protocol stack. Though existing smartphone and mobile terminals are multiple random access technology (multi-RAT) capable, however, large antenna array in small form factor is indispensable for mm-Wave transmission. In order to operate in multiple scenarios (e.g., high/low mobility, under legacy network coverage, as relay node, or operating as D2D underlay node, etc), dynamic radio frame and corresponding protocol stack is required for 5G mobile devices. The reader may refer [73], [74] for further details.

The smartphones and mobile devices for 5G networks are introduced as NanoEquipment (NE) in [75]. The author has discussed 5G RAN and 5G mobile device (i.e., NE) from the perspective of nanocore technology. Using this technology, large antenna array in small form factor can be realized to meet the requirements of the data plane for mm-Wave communication. For high mobility users, low-rate data services may be provided at lower frequencies. Since, control and data planes are expected to operate at different frequencies (i.e., lower frequencies for control/low-rate data and mm-Wave for data plane), dedicated RF chains for control and data planes are mandatory. These are few hardware changes which we can expect for 5G mobile devices. The software and protocol changes are expected to be transparent to mobile devices due to futuristic SDN approach. To realize future cellular communication, dual connectivity for control/data planes and multi-RAT technology for seamless communication (at any available legacy or new air interface) will require self-organized sophisticated radio frame and protocol stack.

In SARC, network densification of dBSSs at mm-Wave spectrum can be achieved to meet the capacity requirements. Since, the capacity requirements of future ultra-dense environment are much higher, therefore assigning physical cell identification (PID)³ to each active/sleeping dBSS can be quite challenging. The PID is the physical ID of the cell which is required by UEs to uniquely identify the serving dBSS and acquire time/slot synchronization. In current LTE systems, the cell ID can be calculated during initial cell search using primary and secondary synchronization signals (PSS/SSS). If the under-utilized dBSS is set to sleep mode by the cBS, UE can never know the presence of near-by sleeping dBSS. The dBSS localization, waking-up and assignment of cell ID, introduced as PID management, are the responsibilities of cBS. Since, cell sleeping is a rare phenomenon in CARC, therefore static PID assignment to always running BSs is a feasible strategy. In SARC, the simple

conventional solution of static PID assignment will result into inefficient PID utilization. Many cell IDs will be unused in case of large number of sleeping dBSSs. The optimum PID management in SARC can follow on-demand PID assignment in a self-organized manner. This strategy can scale well in case more dBSSs are deployed to meet capacity demands. However, this solution comes at the expense of centralized PID management and tight synchronization. For active UEs, the cBS will not only localize the near-by dBSS but also assign the PID (in case of sleeping dBSS); hence assisting the required time synchronization between dBSS and active UEs. Once, the sleeping dBSS is active, it can use the assigned ID and corresponding PSS/SSS to provide time/slot synchronization to UEs. In spite of complex processing, the centralized PID management in SARC can bring self-organization which is indispensable for sleeping dBSSs in future cellular networks.

Another perspective to meet capacity requirement is to select optimal dBSS for data services. Since, cBS has global context information (e.g., positions of dBSSs and UEs), it can use simple path-loss, statistical CSI, and load conditions to associate UE to the optimal dBSS. Using this simple strategy, the UE can be handed over to the dBSS with highest capacity provision. This can be possible because cBS has global knowledge of the coverage area, however, the optimal dBSS selection can be challenging due to possibility of cell-sleeping. In such a scenario, cBS has to initiate wake-up mechanisms, assign PID, arrange initial synchronization, and handover UE to the dBSS. Once a successful handover is accomplished, the reduced flow control (minimum required control signaling) with the dBSS as compared to the full flow control in CARC, offers a higher degree of freedom to achieve higher data rate. The inherent benefit of SARC architecture is the reduced control signaling in radio frame of dBSS. For example, in current LTE and LTE-A systems, the radio frame is 10 ms where control signaling is required to be sent periodically along with requested data. Thus, control signaling takes substantial portion of radio frame to provide connection establishment, handover mechanism, and other control procedures. Such restriction does not hold for dBSS as majority of the control signaling will be provided by cBS via dual connectivity mechanism. Since control signaling is reduced to minimum in dBSS of SARC, the frame size can carry maximum data traffic to meet higher capacity demands.

C. Interference Management

In future cellular systems, interference management will be a real challenge due to heterogeneity (small cells, remote radio head, D2D, M2M, multi-RAT services etc.), dense spectrum reuse (overlay/underlay D2D, M2M), and network densification. Although this hierarchical heterogeneity promises tremendous capacity and coverage enhancements, the resulting interference will be manifolds higher as compared to present deployments.

In present OFDMA based cellular systems, intra-cell interference is mitigated using orthogonal sub-carriers, however, inter-cell interference exists due to frequency reuse (reuse-1). This inter-cell interference has negligible effect on cell-centre users and severe effects on cell-edge users. In literature, this

³In LTE/LTE-A, physical cell ID (PCI) is used by UEs to differentiate between neighboring cells and perform signal strength measurements.

inter-cell interference is addressed using different mitigation techniques like randomization, cancellation, and coordination [76]. The randomization techniques average out the interference across the whole spectrum using scrambling, interleaving etc. Hence, the interference is not mitigated rather distributed equally and fairly over the system bandwidth [77]. The cancellation techniques apply advanced signal processing at the receiver (e.g., interference rejection combining (IRC)) to reject the interference in a single-cell environment whereas coordination techniques push the interference to the cluster level comprising multicell environment. Hence, the notion of interference for cooperative networks has been changed to inter-cluster instead of inter-cell interference. In ideal coordination techniques, intra-cluster interference is completely removed, whereas inter-cluster interference limits the system performance. The ICIC techniques employ either selective frequency reuse, selective power reuse or selective invert power frequency reuse. The frequency reuse (fractional frequency reuse (FFR), partial frequency reuse (PFR), and soft frequency reuse (SFR)) improves the cell-edge performance, however, the major drawback of such techniques is the spectrum under-utilization that directly degrades the overall system performance. The selective power reuse technique is based on higher power for cell-edge users as compared to cell-centre users by keeping orthogonal frequencies to avoid inter-cell collision. This approach overcomes the spectrum under-utilization, however, no significant capacity gain is achievable since interference avoidance in this case is entirely dependent on good channel conditions [78]. The invert power frequency reuse technique is the hybrid of frequency and power reuse. This technique is suggested to achieve performance trade-off between under-utilized spectrum with higher cell-edge throughput and fully-utilized spectrum with lower cell-edge throughput [79], [80].

The ICIC techniques have further been evolved as eICIC for LTE-A. These techniques are categorized into time, frequency, and power domains. In time domain, the interference is handled by sending either ABS or employing symbol shift for the two interfering cells (aggressor and victim) [81]. The cell selection bias is introduced to ensure received signal strength based user association in favor of picocell. In frequency domain, the physical signals and control channels are completely orthogonal among aggressor and victim cells, thereby mitigating interference at the cost of reduced bandwidth [82]. In order to optimize the resources and employ interference control, almost blank resource block (ABRB) is suggested in [83]. The ABRB is defined over both time and frequency domains unlike simple time-domain ABS approach. Hence, it provides more granularity in resource allocation. The ABRB is a generalization of ABS approach and it provides further improvements by providing co-tier (macro-macro) interference control along with cross-tier (macro-pico) interference control. The power domain techniques employ power control mechanism in indoor low-power nodes, however, reducing the maximum transmit power of low-power nodes may degrade the overall performance especially in case of femtocells [77].

In SARC, the interference management has some potential flexibility. Since, the cBS and dBSSs are operated on different bands, hence the cBS UEs can roam even in the coverage area

of dBSSs without causing interference. In the coverage area of cBS, the cell-centre cBS UEs will see no interference from neighboring cBSs due to longer path-loss. However, the cell-edge cBS UEs will be effected by the inter-cell interference. In case of dBSS UEs, inter-cell interference will be higher due to ultra-dense deployments of dBSSs. In such dense environment, interference-aware transmission may be realized at two levels. The first may consider interference mitigation between cBSs by realizing long-range cBS clusters. The second may consider CB for clusters of densely deployed dBSSs.

D. Mobility Management

The optimal mobility management ensures the capacity and coverage of mobile cellular networks. In literature, different approaches are reported for mobility management. In [84], macrocell cooperation and Manhattan grid layout has been proposed for mobility management. The simulation results show that without this cooperation, dense small cells would require at least 4 times more re-connection load. This study investigates potential advantages in mobility management due to macrocell cooperation, however, cell sleeping phenomenon has not been considered in this study. The authors in [85] analyzed the mean handover rate and the mean sojourn time in macrocell assisted small cell architecture. The BSs are deployed as poisson point process (PPP) with serving zones as poisson voronoi tessellations (PVT). The random waypoint is considered for the user mobility. The analytic expression show that the handover rate and sojourn time are simply a function of user velocity, transmission probability, and BS density, however, similar to [84], cell sleeping phenomenon has not been considered. The study in [31] considered 4 Phantom cells per macrocell to evaluate the handover performance. It has been observed that the handover failure gets worse as the density of phantom cells increases. The handovers are only considered for Phantom-Cell-to-Phantom-Cell without investigating the impact of macrocell handovers and cell sleeping on overall mobility management.

The potential advantages of SDN technology for mobility management has been discussed in [86]. The functional description of three approaches for handover management are presented. These include 1) Centralized SDN, 2) Semi-centralized SDN, and 3) Hierarchical SDN. The main problem of preserving session continuity and scalability of handovers is discussed. Similarly, the functional description and architecture of DMM in SDN/OpenFlow has been presented in [87]. Both these studies are mainly a functional level discussion without providing design guidelines for mobility management. For further details on SDN networking, the reader is referred to a recent survey [88].

The mobility management for high-speed railway wireless communication networks has been considered in [89], [90]. The evolution of GSM for railway (GSM-R) to LTE for railway (LTE-R) has yet not been standardized [91], however, LTE has been considered to study the impact on performance of European Train Control System (ETCS) railway signaling [89]. This study did not consider control and data planes separation unlike [90] where theoretical analysis and simulation results are presented to emphasize higher security of train control system and larger capacity for passenger services by using separation

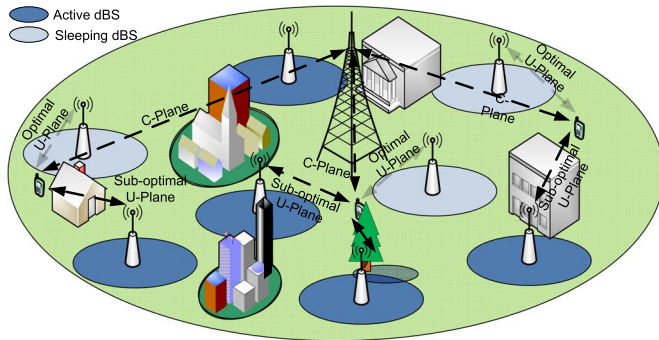


Fig. 2. Active and sleeping dBSs in coverage area of cBS.

framework. This study, however, did not mention the impact of fast handovers on system performance.

In future cellular systems, mobility management should be as seamless as possible for multicell and multi-RAT technologies in order to provide ubiquitous coverage and meet the capacity demands of UEs. In CARC, coverage is provided distributively by different BSs in their respective coverage areas. In such architecture, each BS has limited local knowledge of the network. Since, the coverage and control is not centralized, hence the serving BS might not be knowing the possible optimal sleeping BS in the vicinity to initiate handover. It might be the case that the serving BS initiate handover to the first tier of neighboring sub-optimal multicell or multi-RAT BS whereas the optimal sleeping BS is present in the higher tiers of neighboring BSs. This is due to the non-availability of signal strength of sleeping BSs as well as lack of global knowledge of the multiple tiers of BSs. In SARC, the cBS is centralized with global knowledge of all dBSs (active or sleeping) in multiple tiers in the whole coverage area. The cBS will have the context information and it can predict the signal strength of sleeping dBS e.g., using simple distance dependent path-loss and statistical channel conditions, to select, awake and initiate handover which is not possible in CARC. However, this flexibility comes at the expense of more intense signal processing in cBS to find the optimal set of BSs for handover procedures. One of the limitation of SARC is that if cBS fails due to any reason, the whole coverage area might black out, whereas in CARC, failure of one BS just affects a small portion of the coverage area. However, this limitation can be mitigated by a self-organized backup cBS. The global coverage with sleeping data cells in coverage area and optimal/sub-optimal dBS selection scenarios are depicted in Fig. 2.

In this figure, the channel conditions vary for different deployment scenarios. It might be possible that the closest dBS provides worst data service due to deep fades and shadowing effects as compared to the farthest sleeping dBS. However, the central control of cBS provides flexibility in optimal/sub-optimal dBS selection by incorporating wake up mechanism for sleeping dBSs.

1) *Cell (Re)Selection*: The cell (re)selection in SARC is different from CARC. Two possible scenarios of UE activity are described. The first comprises in-active mode where UE does not require data services. Only coverage is required which is provided by the cBS. The UE can be stationary or moving in the coverage area of either dBSs or cBS without requiring

any handover procedures. In CARC, the cell (re)selection and subsequent handover is carried out for even in-active UEs to provide ubiquitous coverage. This is due to the coupling of control and data planes. The second case consists of active mode where UE requires data services. This case is quite complex as compared to CARC where data session needs to be established in the same BS that is providing the control signaling. In CARC, if the BS is sleeping, then the coverage has to be provided by neighboring BSs whereas in SARC, cell-sleeping is more flexible as global coverage is provided by the cBS (even in coverage areas of dBSs). Therefore, the notion of cell sleeping in SARC is different than CARC. Since the SARC is more feasible for cell sleeping mechanisms, the cell (re)selection procedures become complex as compared to CARC. The mobility management in SARC has to consider the sleeping cells into account while optimizing the dBS selection for the requested data service. Although cBS has global knowledge of dBSs and UEs in the coverage area, it does not know the channel conditions between sleeping dBS and associated near-by UEs. In best channel conditions, the near-by sleeping cell is the best candidate for data services so the cBS can wake up the sleeping cell by assuming simple path-loss model and performs the cell (re)selection as well as handover procedures. However, in worst channel conditions, there can be a case that the near-by dBS might provide worse received signal strength as compared to a far sleeping dBS. In CARC, the UE reports RSRP measurements of neighboring BSs to the serving BS. In SARC, the cBS is long range and hence UE can report the RSRP measurements of dBSs directly to the cBS that can manage cell (re)selection globally (further potential gains of this strategy are highlighted in Section IV-A). However, the measurements by the UE will exclude sleeping dBSs and therefore some other mechanism should be devised for predicting the channel conditions of sleeping dBSs. The cell (re)selection may be based on conventional procedures (RSRP based), though, such procedures are more challenging in SARC due to the sleeping cells in coverage area. The complexity is traded-off with more centralized control on cell-sleeping to conserve energy for green cellular communication.

2) *Handover Procedures*: The handover requirements in SARC and CARC are different. In CARC, complete handover is initiated for cell-centre or cell-edge users. However, in SARC, partial handovers might be required depending on cell-centre or cell-edge users. Therefore, SARC handovers can be classified into partial and full handovers. In partial handover, only data plane handover (DPHO) is required for cell-centre users and control plane is intact. For cell-edge users, complete handover consisting of control plane handover (CPHO) and DPHO might be required. Therefore, SARC offers significant reduction in CPHO overheads for cell-centre users. However, the complete handover in SARC is complex as compared to CARC where the handover is performed softly as both control and data sessions are handed over to a single neighboring BS. In SARC, the control handover is made to the neighboring cBS and data sessions are handed over to the active or sleeping dBSs in the neighboring coverage area. This procedure might produce delays in case of sleeping dBSs. Therefore, an agile and robust soft handover is needed in SARC for cell-edge

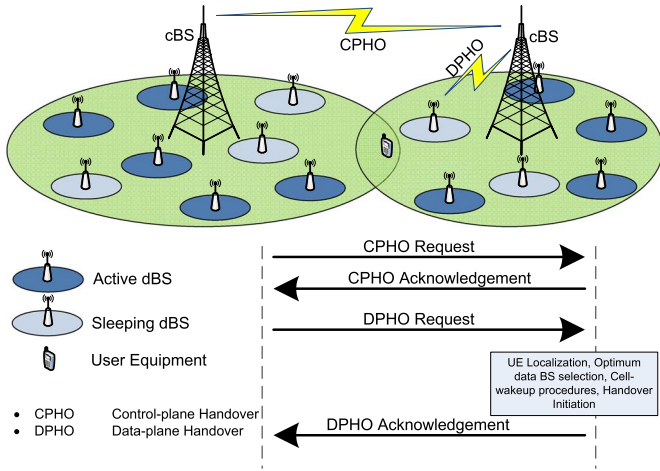


Fig. 3. Handover procedure.

users. In spite of this complexity, the potential gains in SARC due to CPHO overhead reduction can be significant due to the increased number of cell-centre users. The simplest handover procedure for CPHO and DPHO is depicted in Fig. 3. In this figure, we can have different scenarios for handovers which are discussed below.

- a) *Control Plane Handover (CPHO)*: In CPHO, the cell-centre users do not require handover as discussed above. However, for cell-edge users, CPHO is required whether the UE is in active or in-active data sessions. In case of in-active data sessions, the CPHO is not complex and it can be initiated and performed quickly without incurring delays and hence providing ubiquitous coverage. However, in active data sessions, handover procedures might be challenging due to decoupling of control and data BSs.
- b) *Data Plane Handover (DPHO)*: In current LTE systems, the handover is initiated based on events (A1–A5) where the main theme behind these events is to set a certain threshold between serving and neighboring BSs for handover initiation. When the signal strength of neighboring cell is higher than the serving cell, a handover procedure is initiated. A brief description of events [92] is described as follows:-
 - **A1**: Serving cell becomes better than threshold.
 - **A2**: Serving cell becomes worse than threshold.
 - **A3**: Neighbor becomes offset dB better than serving cell.
 - **A4**: Neighbor becomes better than threshold.
 - **A5**: Serving becomes worse than threshold 1 and neighbor becomes better than threshold 2.

In SARC, the cell-centre users will require DPHO based on UE activity. This case can be quite complex as compared to CARC. In case the dBSs are active, then the DPHO procedures are not very complex and no waking-up mechanisms are required. However, the sleeping dBSs may pose challenge to the cBS. Since in this case, the cBS may require localization of UE and prediction of sleeping dBSs channel conditions for optimum dBS selection among the neighboring dBSs. Many factors need to be considered before actually waking-up the dBS. It might not be advisable to wake-up a dBS just for

TABLE VI
MOBILITY MANAGEMENT IN CARC AND SARC

Quick View of Mobility Management		
	CARC	SARC
BS Knowledge	Knowledge of the multiple tiers of BSs is not available.	Global knowledge of all dBSs (active or sleeping) due to centralized cBS
Sleeping BSs	RSRP is not available.	cBS can predict statistical RSRP.
Handover	Serving BS cannot initiate handover to sleeping BS.	Initiate wake-up mechanism and subsequent handover
BS failure	Failure of one BS affects small portion of the coverage area.	Whole coverage area can be affected.

Cell (re)selection	
CARC	SARC
If the BS is sleeping, then the coverage has to be provided by neighboring BSs.	Flexible cell sleeping mechanisms require complex cell (re)selection procedures.
UE reports RSRP measurements of neighboring BSs to the serving BS.	UE reports the RSRP measurements of dBSs directly to the cBS that can manage cell (re)selection globally.

Handover Procedures	
Complete handover is initiated for cell-center or cell-edge users	Partial handover for cell-center users and full handovers for cell-edge users
Complete handover is performed softly as control/data sessions are handed over to a single neighboring BS.	Complete handover in SARC is complex due to decoupled control and data planes.

short-time roaming users. In order to provide optimum DPHO, mobility trends might be predicted based on context and history to differentiate between short/long term UE camping and avoid ping-pong effects before waking up the dBS. The quick view of mobility management, cell (re)selection, and handover procedure is given in Table VI.

The discussion of above mentioned performance measures target two aspects: 1) shortcomings of the existing approaches, and 2) potential gains due to SARC architecture. In order to provide a quick view of this discussion, we provide shortcomings of CARC and potential gains due to SARC in Table VII.

III. COOPERATIVE COMMUNICATION (ESSENTIAL BACKGROUND)

The cooperative communication is a broad term encompassing mainly two categories of wireless networks e.g., cellular, and ad-hoc. The objective of cooperation in both cases is same i.e., nodes should act as cooperative agents for other nodes in order to improve, for example, coverage probability, interference management, and capacity of the overall system [93]. However, the cooperation strategies are different due to the presence and absence of infrastructure in former and later cases, respectively. For example, in ad-hoc networks, wireless nodes spontaneously and dynamically self-organize into an arbitrary and temporary infrastructure [94] without relying on central

TABLE VII
SHORTCOMINGS/POTENTIAL GAINS DUE TO CARC/SARC

Perf. Measure	General Approach	Shortcomings due to CARC	Potential Gains due to SARC
Energy Efficiency	<ul style="list-style-type: none"> ⇒ Dynamic BS switch-off ⇒ Renewable energy resources ⇒ Smart grid 	<ul style="list-style-type: none"> ⇒ Generation of coverage holes due to BS switch-off. ⇒ Higher interference due to increased transmit power in cell range expansion. ⇒ Higher overhead/HO signaling results in more power consumption. ⇒ Renewable and smart grid resources do not address the problem of under-utilized network resources rather provide alternate source of energy. 	<ul style="list-style-type: none"> ⇒ No coverage holes due to ubiquitous coverage. ⇒ No cell range expansion is required due to ubiquitous coverage. ⇒ Reduced power consumption due to reduced overhead/HO signaling. ⇒ Renewable and smart grid resources can be added in main power source.
System Capacity	<ul style="list-style-type: none"> ⇒ Spectrum efficiency ⇒ Spectrum aggregation ⇒ Network densification 	<ul style="list-style-type: none"> ⇒ High frequency (mm-Wave) cannot be used for data services since it will result in reduced coverage (due to spot beams). ⇒ Since mm-Wave communication cannot be realized, the huge contiguous spectrum cannot be utilized for capacity enhancements. ⇒ Control/overhead signaling restricts payload size resulting in reduced system capacity. ⇒ Limited pro-active caching for cooperation set due to distributed and local context. 	<ul style="list-style-type: none"> ⇒ High frequency (mm-Wave) communication to enhance system capacity without losing coverage. ⇒ The huge contiguous spectrum can be utilized for very high rate data services. ⇒ Control is decoupled, therefore, radio frame may contain maximum payload size. ⇒ Global context allows wake-up mechanism as well as pro-active caching for cooperation.
Interference Management	<ul style="list-style-type: none"> ⇒ Interference avoidance ⇒ ICIC/eICIC ⇒ CB and CoMP 	<ul style="list-style-type: none"> ⇒ Low degree of freedom for active/inactive users ⇒ While moving, in-active users require handovers that causes interference. ⇒ Active users require full handover (CPHO+DPHO) ⇒ Ping-pong effect due to high mobility near cell-edge. 	<ul style="list-style-type: none"> ⇒ No CPHO for in-active users. ⇒ Only DPHO for active users. ⇒ Reduced interference due to less CPHO and DPHO. ⇒ Ping-pong effect can be controlled by providing data services via control plane.
Mobility Management	<ul style="list-style-type: none"> ⇒ Macrocell cooperation ⇒ Macrocell assisted small cells ⇒ SDN approach 	<ul style="list-style-type: none"> ⇒ Cell sleeping is not flexible. ⇒ CSI acquisition is challenging in sleeping cell scenario. ⇒ Cooperation requires overhead signaling on backhaul. 	<ul style="list-style-type: none"> ⇒ Mobility management is ensured by always-running cBS with flexible dBSSs sleeping possibility. ⇒ Centralized control of SARC may exploit statistical CSI for sleeping cells. ⇒ Ubiquitous control in SARC may exploit pro-active network caching to overcome overhead signaling on backhaul.

controller (e.g., BS) for the signaling flow and connection management, whereas, in cellular cooperative networks, the wireless nodes are controlled and dependent on the serving BSs. However, cooperative communication with little involvement of BS can be seen in case of D2D communication. This type of communication can be considered as infrastructure based ad-hoc links where peers act as either mobile relays (e.g., content dissemination) or the source nodes (e.g., file transfer, exchange of common contents etc.) with little involvement (control signaling) of access and core network. In this context, D2D communication can also be categorized as cooperative communication to assist the network for content dissemination or ad-hoc type direct communication.

In this section, we provide essential background to understand infrastructure based multicell BS coordination (e.g., CoMP), self-organized BS clustering and network-controlled D2D communication.

A. CoMP Classification

CoMP can be classified from a number of different perspectives. For example, if the transmission direction is taken into consideration, then CoMP is classified as either joint

detection (JD) (uplink) or joint transmission (JT) (downlink). From the cooperation system architecture, CoMP can either be centralized, decentralized or distributed. The level of CoMP coordination is quantized into no, limited, and full cooperation. Based on this quantization, CoMP scales into either intra-cell beamforming, multicell CB, or fully coordinated CoMP. The classification of CoMP is illustrated in Fig. 4.

1) *Joint Detection*: In multicell JD, each BS receives signals from its respective UEs and exchanges either quantized or un-quantized signal between cooperating BSs. In a typical scenario, a BS suffering from high co-channel interference sends cooperation request to the participating BSs. This request includes the cooperation mode and physical resource block (PRB) associated with the effected UE. The cooperating BSs exchange the quantized signal of the requested UE depending on the cooperation mode as follows:

- **IQ Samples Transfer**: A frequency-domain in-phase quadrature (IQ) samples representing complex constellation points of the requested UE are extracted from Fast Fourier Transform (FFT) module and transferred to the serving BS. The serving BS processes the IQ samples as if they were received by its own antennas.

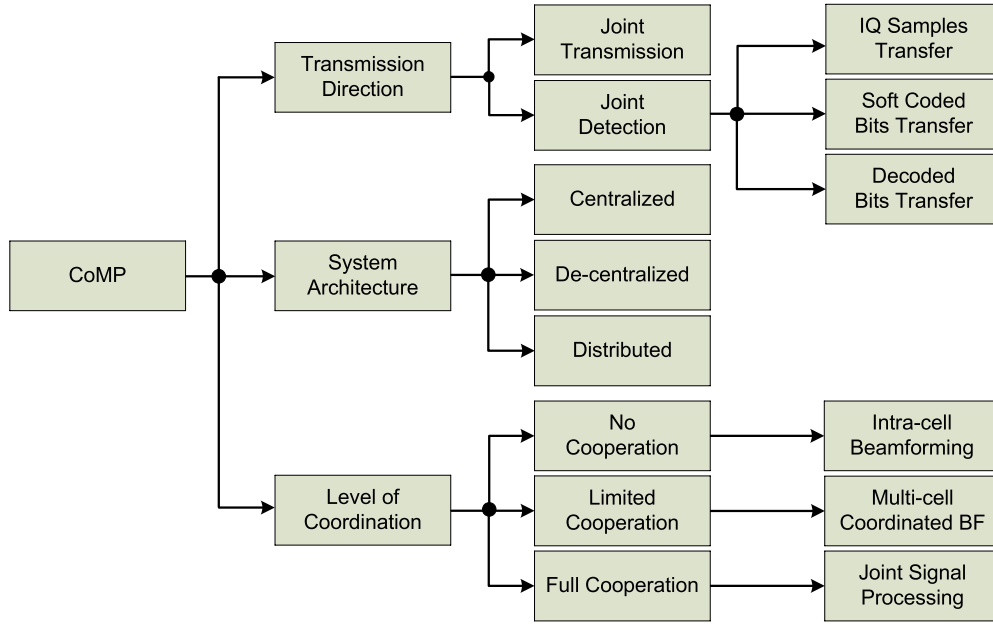


Fig. 4. CoMP classification.

- **Soft Coded Bits Transfer:** In this cooperation mode, the cooperation request of serving BS must contain not only the PRBs of the transmitted signal, but also its modulation and reference signals. After equalizing and demodulating the received signal, the cooperating BSs transfer the quantized soft values of the coded bits back to the serving BS.
- **Decoded Bits Transfer:** The serving BS also mentions the decoder of the associated UE and shares it with cooperating BSs which demodulate and decode the signal, perform cyclic redundancy check (CRC) and transfer the decoded data back to the serving BS on successful CRC check. After receiving the response message, the serving BS performs selection combining. For details on JD algorithms, the reader is referred to [95]–[98].

2) *Joint Transmission:* In multicell JT, CSI and user data of each UE in cooperation set is exchanged between cooperating BSs. Each BS designs beamformers and jointly transmit the data to the target UE. In this scheme, coherent transmission plays a key role to achieve maximum performance gain of JT. The reader is referred to [99]–[101] for optimal JT strategies.

3) *Centralized:* In centralized JD, the cooperating BSs decode the received signal of the corresponding UEs according to the cooperation mode (mentioned in Section III-A1) and share it with the BS that acts as a centralized node to jointly decode all UEs. In case of centralized JT, each participating BS of cooperation cluster send the CSI to the centralized controller which finds global optimal precoding vectors. These precoding vectors are then shared between participating BSs to exploit inter-cell interference.

4) *De-Centralized:* In decentralized JD, every cooperating BS individually and independently decodes the uplink transmission of respective UEs by exploiting CSI that has been shared between all BSs in the cooperation cluster. In case of JT, every cooperating BS has different extent of CSI knowledge and no

BS in the cooperation cluster has full knowledge of global CSI at transmitter. The global optimal precoding is not possible in this case and hence this decentralized JT provides sub-optimal solutions.

5) *Distributed:* This scheme is similar to the centralized approach with only difference that there is no dedicated central unit (CU) and any participating BS can act as a centralized node in a distributed manner.

6) *Intra-Cell Beamforming:* In case of non-cooperative CoMP, BSs do not exchange information, rather perform individual intra-cell beamforming based on limited feedback from their respected UEs. Based on feedback, each serving BS performs interference-aware scheduling and the corresponding UEs have the capability of IRC receiver.

7) *Multicell Coordinated Beamforming:* In this case, the cooperating BSs exchange CSI between each other in order to reduce the inter-cell interference. This level of coordination requires small backhaul capacity and is known as coordinated beamforming in 3GPP LTE-A literature.

8) *Full Cooperation:* In case of full cooperation, CSI and user data of each CoMP-enabled UE is exchanged between cooperating BSs. This scheme requires very large backhaul capacity and strict synchronization requirements to perform joint signal processing. The full extent of this cooperation may be exploited by adopting coordinated and coherent transmission to the target UE.

B. CoMP Clustering

Due to signaling overheads on air interface and backhaul, the number of BSs in cooperation cluster is limited in practice. For such cooperating BSs, the clustering can be static or dynamic. The static clustering is designed on the basis of geographical positioning of BSs and is kept constant over time and channel conditions. However, dynamic or adaptive clustering adapts the

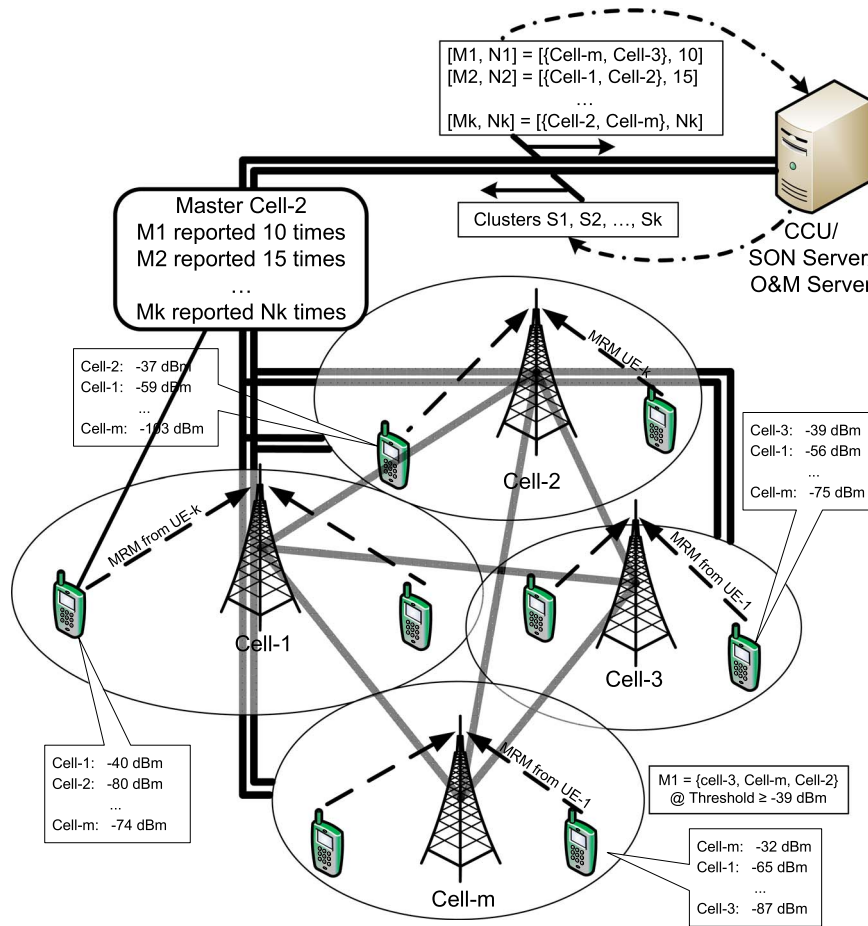


Fig. 5. Self-organizing network based adaptive clustering.

channel conditions and is comparatively more complex. The adaptive clustering for current cellular system is suggested by exploiting existing RF measurements reported by UEs to the serving BS [102] as shown in Fig. 5.

In such an approach, huge number of average RSRP measurements are extracted from the measurement report messages by serving BSs of respective UEs. These huge measurements are categorized in the form of reporting sets and sent to the CoMP CU (CCU) which selects the cooperation cluster based on some performance indicator. These indicators may, for example, include system load, delay, system complexity, combined signal strength, user priority classifications, or other network related metrics [102]. The advantage of this approach is that it can utilize the existing framework of 3GPP (functions such as automatic neighbor relation (ANR), neighbor relation tables (NRTs)) to provide self-organizing network (SON) based clustering solution.

C. Device-to-Device Cooperation

In present cellular systems, we have HetNets that comprise macrocells, small cells (micro, pico, femto), access points, and smart mobile devices. In future cellular systems, ultra-dense HetNets are expected where capacity and coverage can be met by cooperation between different nodes. In this context, even more smaller granularity of cooperation is expected e.g.,

CB/CoMP at device level (D2D CoMP) and D2D cooperation for content dissemination or common information exchange.

D2D communication has an old origin in the form of ad-hoc and personal area networking technologies in unlicensed spectrum bands e.g., industrial, scientific, and medical (ISM) bands. In this case, short range communication is possible without infrastructure unlike cellular communication where network control is mandatory. Although such ad-hoc communication requires very less control signaling, it inherits certain drawbacks such as limited content sharing, no point-to-multipoint links, synchronization issues, authentication, and security concerns. D2D communication has also been proposed in licensed spectrum especially in cellular bands in either ad-hoc or network-assisted mode. The ad-hoc mode of D2D communication in licensed spectrum offers limited applications similar to the unlicensed counterpart, however, network-assisted D2D communication in cellular band has many applications and services including proximity-based commercial services, social networking, video sharing, mobile relaying, gaming, traffic offloading, capacity enhancement (frequency reuse), extended cellular coverage, and improved energy efficient communication.

D2D communication has been studied by research community quite long. In early 2006, mobile communication system Aura-Net, based on wireless technology FlashLinQ, was proposed. This communication system exploited D2D communication for proximity-aware inter-networking to enhance and

augment the capacity and coverage of wireless wide area network (WWAN) [103]. The proposed system features distributed spatial spectrum reuse protocol that is scalable to different levels of proximal granularity. It is mentioned that Aura-Net provides a template for future proximal aware “Internet of Things.”

The smart communication devices have the capability to be virtually connected to any device, any time, anywhere. This global connectivity offers remoteness as well as proximity at the same time. Coupled with proximity services, the ultra-dense heterogeneity of future cellular networks can be exploited to achieve potential advantages of low-range high-rate D2D data communication to enhance capacity and coverage. D2D communication is considered as a sub-feature of 3GPP LTE-Direct Rel-12 [104]. It comprises two main features:

- 1) Device to Device Peer Discovery
- 2) Device to Device Data Communications

In order to complement huge SC deployments and overcome OPEX and energy efficiency concerns, traffic off-loading from cellular to multi-RAT networks, other unlicensed wireless infrastructures (e.g., WiFi) and multi-hop ad hoc links between devices drew much attention recently. The MOTO project [105] funded by the European Commission under FP7 proposes traffic offloading where D2D communication is one of the ingredients. The establishment of D2D links can be considered as ad-hoc network in infrastructure where the network resources are reused by mobile peers directly with little involvement (control signaling) of access and core network. In this hybrid architecture (infrastructure based ad-hoc links), huge capacity, ubiquitous coverage, energy efficiency, and backhaul gains are promised by exploiting maximum D2D links and reusing the resources optimally.

D2D communication is being considered as an integral part of next generation cellular networks where proximity services and social networks are dominating over conventional services. The network-assisted D2D communication offers another tier of communication within a cell by reusing the spectrum resources. The reduced distance between nodes improves spectral efficiency, throughput per area, energy efficiency, and latency. The link reliability can be improved by migrating from multi-hop to single hop communication (mesh-like topology). The coverage can be enhanced by multi-hop cooperation between devices which can be the only communication in case of no coverage-zone, coverage holes, and emergency situation. The load balancing and load management can be optimized by network and device pro-active caching of common information and offloading the devices to establish direct links [106]. Hence, D2D links in future cellular networks are key enablers for traffic off-loading, reducing access delays, optimal resource utilization, capacity and coverage enhancements, and energy efficient communication.

The huge potential performance gains due to direct communication are coupled with certain challenges that include quality-of-experience (QoE), quality-of-protection (QoP), user consent, battery issues, and cellular aspects. These factors are very important and can directly effect the performance gains of D2D communication. The QoE includes user perception,

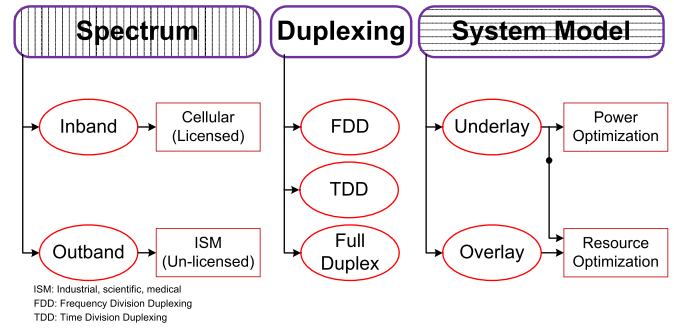


Fig. 6. D2D taxonomy.

expectations, and experience that needs to be maintained in cellular and direct mode of communication. The QoE is a measure of user’s desired or expected experience about cellular services. Though user might not be interested in specific mode of communication (cellular or D2D), he can be considered as perceiving seamless switching between two modes and enjoying services at agreed QoS. The QoP refers to the confidentiality and privacy which is even more severe when the locations and contents may be compromised by intruding D2D partner. However, this can be tackled by incorporating simple authentication, authorization and accounting (AAA) procedures to block such attacks. Even with this solution, using the device without the consent and permission of the mobile owner is a big problem along with battery consumption issues. Using the device for D2D relaying, for example, without incentivizing the mobile owner can not be realized practically. The cellular aspects include interference management due to underlay D2D network, optimal number of D2D nodes, exploitation of common interests (social relationship strength to harness D2D communication), CSI between nodes, and synchronized switching between cellular and D2D nodes.

D2D communication can be classified in a taxonomic representation as shown in Fig. 6.

The spectrum used for D2D communication can be either in-band or outband. The inband spectrum is considered as licensed cellular band whereas outband spectrum means unlicensed e.g., ISM band. D2D links can operate either in FDD, time division duplex (TDD) [107]–[109], or full duplex mode [110]–[113]. In FDD mode, two separate frequencies for transmit and receive are required at both nodes of D2D link. This results in under-utilization of spectrum by underlay D2D network. In order to overcome under-utilization of frequencies, TDD mode can be used where single frequency is required for transmit and receive. This comes with more complex transceiver design. The full duplex model allows single frequency without slot sharing (as in TDD mode), however, this can be possible if self-interference due to simultaneous transmission/reception can be canceled. The interested reader is referred to [113] for further details. The TDD and full duplex modes have potential advantages of cost-effective transceiver design in small form factor. D2D communication is possible without network, however, it has limited applications as compared to network-assisted direct communication. The capacity and coverage can be enhanced either by overlay or underlay system model. In case of overlay communication, the dedicated spectrum is allocated for D2D

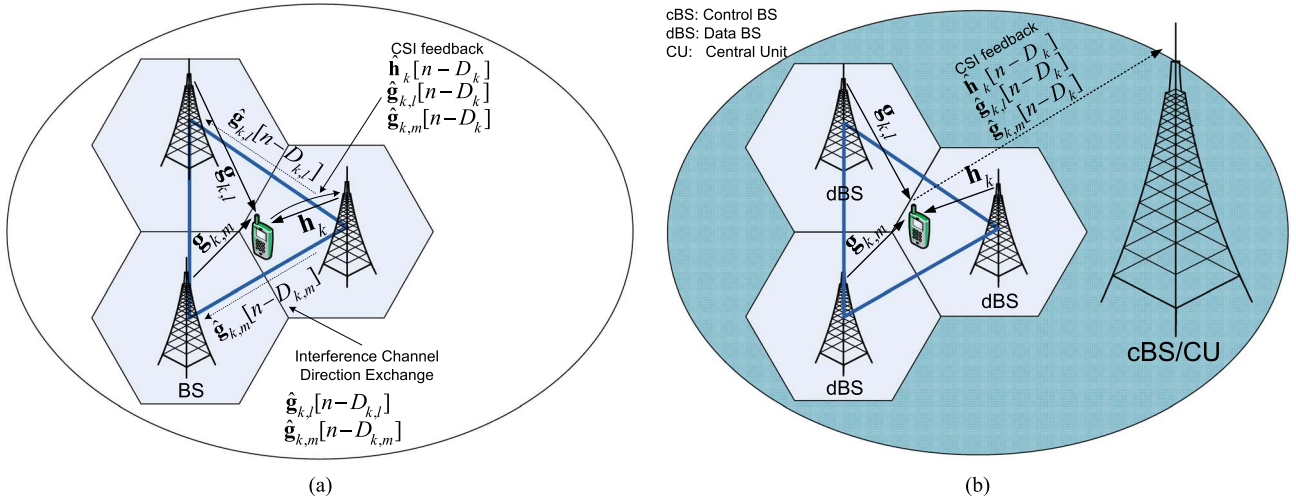


Fig. 7. System model for exchange of desired and interfering channels. (a) Conventional architecture (CARC). (b) Split architecture (SARC).

network. This can be done by partitioning the available spectrum for cellular and D2D users. In this system model interference management can be relaxed due to allocating dedicated spectrum. However, this model results in low frequency reuse and waste of cellular resources [114], [115]. A more complex underlay model can be realized where maximum capacity and coverage can be achieved by sharing the same spectrum between cellular and D2D users (full frequency reuse) by incorporating more sophisticated interference management techniques. The interference management comprises either power or RRM depending on uplink or downlink spectrum reuse.

The reader is referred to [116] and [117] for further details on CoMP and more recent study on CoMP for 5G networks, respectively. For D2D communication, [118] and [119] provide comprehensive survey and tutorial on the subject.

IV. COOPERATION IN SARC

The cooperation in next generation ultra-dense HetNet is indispensable especially when huge D2D links are exploited. In this section, we provide preliminary discussions for possible extension of cooperation framework in SARC. In this context, we first present coordinated beamforming followed by D2D clustering and D2D CoMP in SARC. We further discuss realization of SARC in cloud-RAN architecture, fronthaul/backhaul limitations and possible solution in the form of proactive caching.

A. Coordinated Beamforming

In coordinated beamforming, the desired and interfering CSI (ICI) is required at each participating BS of the cooperation set. In conventional multicell CARC, UE measures channel state of serving and neighboring BSs and reports the quantized channel information to the serving BS. The serving BS sorts out ICI and exchanges corresponding interference information to the participating BSs. The participating BSs receive delayed interference information via backhaul and choose appropriate beamformers. In this mechanism, there are two drawbacks. First, the exchange of CSIs between cooperating BSs incurs backhaul delay in

addition to the feedback delay from UEs (refer [120], [121] for further details). Secondly, in case the CSI is perturbed (due to quantization effects, noise etc) during exchange via backhaul, the interference at the neighboring cells cannot be perfectly removed resulting in sub-optimal performance [122]. In order to highlight these problems, a simple system model of three cells is considered where exchange of desired and interfering channels for SARC and CARC are, respectively, compared in Fig. 7. In this figure, the downlink (uplink) desired and interfering channels at UE (BS/dBS) are, respectively \mathbf{h}_k and $\mathbf{g}_{k,x}$ for $x \in \{l, m\}$. The UE normalizes and quantizes these channels to $\hat{\mathbf{h}}_k[n]$ and $\hat{\mathbf{g}}_{k,x}[n]$, respectively. These channels are fed back by the UE to the serving BS/dBS. The purpose of limited (quantized) feedback is to send the channel direction to the serving BSs [121] where multi-antenna beamforming (single-cell) or CB (multicell) vectors are chosen in such a way that they lie in the null space of interference channel directions [123] to achieve inter-cell interference nulling.

The feedback delay associated with CSI is D_k . Upon receiving the CSI, each BS segregates and forwards ICI to the respective cooperating BSs via backhaul which causes an additional delay $D_{k,x}$ resulting into a total delay of $D_{bh} = D_k + D_{k,x}$ where $D_{k,x} \geq D_k$. The relation between the current and delayed CSI and ICI is given by Gauss-Markov auto-regressive model [124] that assumes slowly time varying channels as follows [123]:

$$\begin{aligned} \hat{\mathbf{h}}_k[n] &= \eta_k \hat{\mathbf{h}}_k[n - D_k] + \sqrt{1 - \eta_k^2} \mathbf{e}_{h_k}[n], \\ \hat{\mathbf{g}}_{k,x}[n] &= \eta_{k,x} \hat{\mathbf{g}}_{k,x}[n - D_{k,x}] + \sqrt{1 - \eta_{k,x}^2} \mathbf{e}_{g_{k,x}}[n], \end{aligned} \quad (1)$$

where $\mathbf{e}_{h_k}[n]$ and $\mathbf{e}_{g_{k,x}}[n]$ are, respectively, desired and interferer channel error vectors distributed as $\mathcal{CN}(0, 1)$. The auto-correlation function of desired and interfering channel are η_k and $\eta_{k,x}$, respectively, defined by the Clarke's auto-correlation model [124], [125] as:

$$\begin{aligned} \eta_k &= b_0 J_0(2\pi D_k f_d T_s), \\ \eta_{k,x} &= b_0 J_0(2\pi D_{k,x} f_d T_s), \end{aligned} \quad (2)$$

where b_0 is the variance of the underlying Gaussian process, $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind, f_d is the maximum Doppler frequency, and T_s is the symbol duration.

Based on above formulation, in Fig. 7(a), it can be seen that each BS exchanges the quantized interference channel between cooperation set. This ICI experiences asymmetric backhaul delay $D_{k,x}$. In such a distributed architecture, the coherent beamforming can not be achieved and, hence, the benefits of CB can not be fully exploited. However, in SARC, the CSI/ICI is fed back directly to the cBS, therefore $D_{k,x}$ associated with the dBS x is reduced to D_x . For this case, the auto-correlation function of the interfering channel and corresponding ICI becomes

$$\eta_{k,x} = b_0 J_0(2\pi D_k f_d T_s) = \eta_k,$$

$$\hat{\mathbf{g}}_{k,x}[n] = \eta_k \hat{\mathbf{g}}_{k,x}[n - D_k] + \sqrt{1 - \eta_k^2} \mathbf{e}_{g_{k,x}}[n], \quad (3)$$

By reducing $\eta_{k,x} = \eta_k$ in (2) and $D_{k,x} = D_k$ in (1), we can see that, in Fig. 7(b), the backhaul delay has been eliminated due to direct feedback from the UEs to cBS and hence all beamformers for the participating dBSs can be designed coherently. Although the coherent beamforming can be carried out in SARC, the real problem is to share the beamformers to the corresponding dBSs via asymmetric backhaul links. This problem can be tackled by incorporating centralized timing advance mechanism in cBS CU to allow the participating dBSs adjust the transmission to achieve coherent CB.

The distributed beamforming suits to CARC architecture where CSI/ICI is available in a distributed manner and beamformers are designed at every participating BS. In this case, the perturbation of ICI and corresponding backhaul delay directly effects the performance of CB. However, in SARC, due to inherent centralized ubiquitous coverage, the CSI/ICI from UEs can directly be fed back to cBS. The cBS can act as a CU to design coherent beamformers based on large number of measurement reports. The advantage of this approach is that the backhaul signaling for exchange of interference information and corresponding asymmetric delay can be removed. This approach can further adapt the channel conditions more rapidly since the beamforming does not depend on backhaul delays. In order to address perturbation issue due to exchange of ICI via backhaul links, the availability of global CSI at cBS can be leveraged to jointly design beamforming matrix. Although the exchange of jointly designed beamforming matrix may also be perturbed while exchanging beamformers to the cooperating dBSs via backhaul, the perturbation will effect the overall matrix and results in fair system performance unlike distributed beamformer design in case of CARC.

B. D2D Clustering

In order to enhance spectral efficiency of cellular systems, intra-cell interference has been tackled in LTE and LTE-A by using OFDMA technology and RRM. Therefore, intra-cell interference is not a problem in such systems, however, inter-cell interference exists for which cooperative communi-

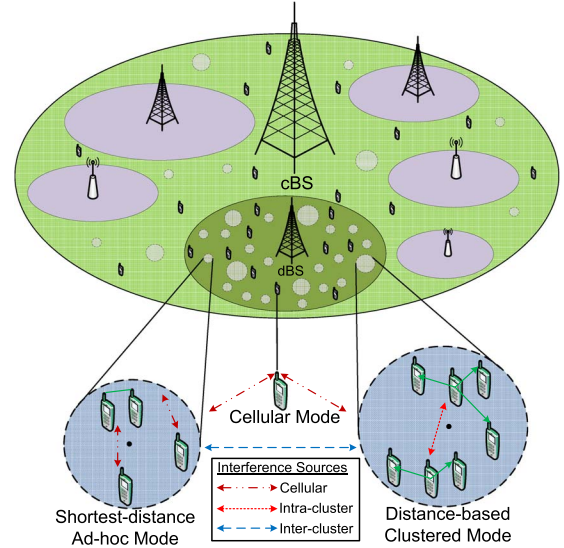


Fig. 8. D2D clusters and interference sources in SARC.

cation (CB, and CoMP) has been suggested [17] to coordinate interference between clusters of BSs and improve cell-edge performance.

In future ultra-dense HetNets, underlay D2D network is being considered as an integral part for rapidly evolving proximal inter-networking. This smallest communication tier reuses the resources of primary users within a cell and hence again generates intra-cell interference which was previously mitigated by OFDMA technology. If we extend the granularity of cooperation at device level and utilize centralized context and CSI (due to separation framework) at cBS CU, we can flexibly control intra/inter-cell interference and hence meet huge capacity gains and spectral efficiency demands of future cellular systems without compromising energy efficiency, and overhead signaling cost (e.g., at air interface or backhaul links). We can further improve these metrics by exploiting self-organized D2D clusters and network controlled D2D communication.

In the following, we consider hierarchical HetNet (i.e., D2D tiers in cBS as well as dBS tiers [126]) in SARC and realize D2D communication using channel condition and/or social relationship between nodes as shown in Fig. 8.

1) *Channel Conditions Based D2D Clusters*: D2D communication can be realized either in ad-hoc mode or in the form of clusters. In ad-hoc mode, we consider point-to-point links between devices. Since two nodes are allowed to communicate based on shortest distance (reduced path-loss) criterion therefore, in this case, small cooperation radius is required. This mode is feasible for exchange of already cached common information between two devices. However, this mode undermines the potential capacity gain due to the rejection of other nodes that might come in the cooperation radius and request the same common information. In clustered mode, we consider point-to-multipoint links between devices, therefore, requiring comparatively higher cooperation radius. This mode is feasible for content dissemination. Based on the channel conditions or simple reduced path-loss criterion, one node can be selected by

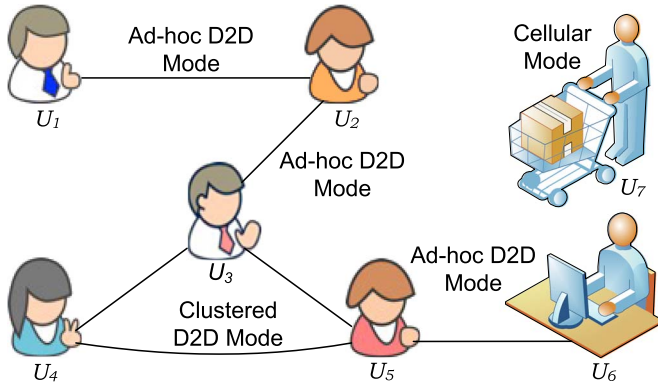


Fig. 9. Social network offers cellular, ad-hoc, and clustered D2D modes.

the network to disseminate contents to the requesting nodes. This mode offers higher capacity gain as compared to ad-hoc mode of D2D communication. For cluster regions in ad-hoc and clustered mode, we foresee interference due to:

- Primary cellular user.
- Intra-cluster D2D nodes.
- Inter-cluster D2D nodes.

In order to minimize mutual interference between cellular and D2D users, the power optimization at conventional serving BS should consider uplink power control of not only cellular users but also transmit power of near-by D2D nodes. This can be possible if serving BS request near-by D2D nodes to share CSI between the nodes. The CSI may also be used for network-assisted centralized or distributed beamforming to mitigate intra-cluster interference. Similarly, if we incorporate inter-cluster level cooperation, further capacity gains may be envisaged.

2) *Social Network Based D2D Clusters*: The channel conditions based clustering of D2D nodes is realistic, however it provides overestimated spectral gains due to the assumption that every node has common information to exchange with every other node. In order to assume realistic assumption about common information exchange or content dissemination, social-aware D2D communication should be considered. The social influence of different mobile users may be quantized into different levels of social impact by exploiting the history and logs of each user. For example, some mobile users have limited social influence in terms of assisting the network for content dissemination or offloading and they fit into the category of cellular mode or ad-hoc mode D2D communication. On contrary, many mobile users fall into the category of clustered type D2D communication where they can actively assist the network for exchange of common information and content dissemination. Such social influence may be exploited to model realistic and optimum D2D links/clusters. The exemplary social network for different levels of social interaction is shown in Fig. 9.

In this figure, D2D link establishment can be done by considering different decision criterion. In this context, different users may be assigned different levels of social influence. The social influence can be calculated using measures of centrality.

TABLE VIII
SOCIAL INFLUENCE USING CLOSENESS CENTRALITY

Social Influence							
Closeness Centrality							
Node	U_1	U_2	U_3	U_4	U_5	U_6	U_7
Closeness	0.52	0.78	1	0.78	0.78	0.56	0
Mode Selection							
User	Social influence		Mode				
U_7	No		Cellular				
U_1 and U_6	Low		Ad-hoc				
$U_2 :: U_5$	High		Clustered				

For example, we can use simple measure of closeness centrality to assign weights to different users in Fig. 9. The closeness centrality is defined as the shortest distance between a reference node and all other nodes reachable from it [127], [128]. The simple mathematical relation of closeness centrality can be given as:

$$C_c(U_i) = \frac{\left[\sum_{j=1}^N d(U_i, U_j) \right]^{-1}}{N-1}, \quad (4)$$

where N is the total number of nodes and $d(\cdot)$ is the shortest distance between reference node and all other nodes. Using (4), the closeness centrality⁴ weights can be measured as shown in Table VIII.

According to closeness centrality calculated in Table VIII, U_7 has no social influence, therefore it is suitable for cellular mode. The users U_1 and U_6 have low level of social influence and hence they are feasible for ad-hoc mode D2D communication. The users U_2 , U_4 , and U_5 have slightly higher influence as compared to users U_1 and U_6 that allows them to be considered for clustered mode D2D communication. In case of ties (e.g., U_2 , U_4 , U_5), reduced path-loss or better channel conditions based criterion may be used to establish link. User U_3 has highest influence which make it suitable for content dissemination in clustered mode of D2D communication.

3) *Prediction Based Adaptive D2D Clustering*: As mentioned in Section III-B, the clusters can be static or dynamic where the latter offers more gains as compared to former. The dynamic clustering and cooperation framework is suitable for nomadic users [17]. Since D2D communication is being evolved for proximity services and inter-networking, dynamic clustering and cooperation framework is very feasible for such type of communication. The dynamic clustering can be extended into self-organized adaptive clustering if the user mobility is predicted. For example, by predicting dwell times of potential D2D users at serving dBS, the required signaling for D2D clustering may be performed in a self-organized manner. Another advantage of this approach is that the prediction of dwell times may allow to tackle ping pong effects and reduce handover cost for switching between cellular and D2D modes. The adaptive clusters can further be optimized by considering

⁴The closeness centrality has been normalized by the maximum weight in Table VIII.

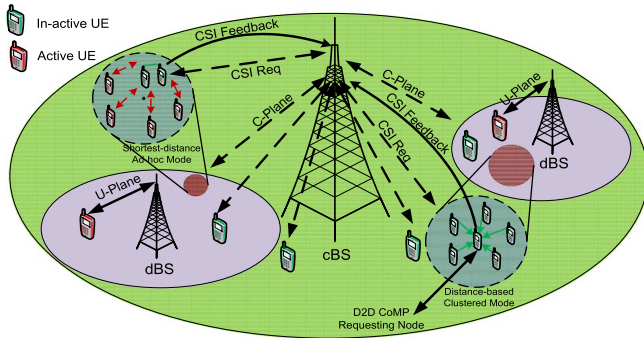


Fig. 10. D2D CoMP to manage interference in underlay network.

mobility patterns along with reduced path-loss, common contents and channel condition criterion.

C. D2D CoMP

In previous sub-section, we have presented two modes of D2D communication i.e., ad-hoc and clustered (Fig. 8). In both cases, cooperation framework for multicell BS i.e., CB and CoMP can be realized in SARC for D2D communication. This type of cooperation coupled with common information exchange (ad-hoc mode) or content dissemination (clustered mode) is introduced as D2D CoMP. Since cBS has global context of every node in the coverage area, it can discover nodes for either ad-hoc or clustered mode communication e.g., by localizing nodes and applying shortest distance/reduced path-loss criterion.

In order to get CSI between cooperating and requesting nodes, cBS can send a reference signal and request a CSI feedback. Based on RSRP values, one of the node in cooperation cluster may send CSI directly to the cBS. The cBS can use this CSI to design beamformers and share with nodes in cooperation set for proactively cached common information exchange or content dissemination. D2D CoMP in SARC is shown in Fig. 10.

In this figure, D2D cooperation regions are shown for ad-hoc and clustered mode D2D CoMP operation. In case of ad-hoc mode, cBS needs to localize and discover an influential partner node⁵ with shortest distance (reduced path-loss) criterion. Once an influential node (containing common information) is identified within proximity of requesting node, cBS can command influential node to send reference signal and subsequently request CSI feedback from the requesting D2D node. For example, in a simple scenario, zero-forcing (ZF) or minimum mean-square-error (MMSE) [129] can be used to design precoder to realize CB for ad-hoc mode D2D communication.

In case of clustered mode D2D communication, cBS needs to localize a set of influential nodes (known as cooperating nodes in traditional CoMP) that can make cooperation cluster for content dissemination. At this stage, cBS needs to know CSI between requesting and influential nodes. Similar to the ad-hoc mode, cBS can command influential nodes to send ref-

erence signal and subsequently request CSI feedback from the requesting node. However, CSI acquisition is more complex as compared to ad-hoc mode due to higher number of distributed influential nodes. Here, we present one strategy to acquire CSI at cBS. In this strategy, cBS will schedule different time slots in a time division multiple access (TDMA) fashion and allocate these slots to the influential nodes. Meanwhile, cBS will command requesting node to acquire time division multiplexed (TDM) reference signals, measure CSI and feedback to the cBS. Once CSI is acquired by the cBS, ZF or MMSE, as mentioned for ad-hoc mode, can be used to design precoders at cBS and shared with influential nodes. The D2D CoMP has potential gains to mitigate interference, however, it comes with the additional cost of higher signaling for CSI acquisition.

D. SARC in Cloud-RAN

The realization of control and data planes separation has been discussed briefly in [30], [130], [131] through Carrier Aggregation (CA) and multiple remote radio head (RRH). Similarly, in [132], the integration of software-defined RAN (SD-RAN) and BCG2 architecture (i.e., decoupled control and data planes) has been suggested to achieve greater benefits and faster realization of both technologies. Motivated by such studies, we present arguments to support SARC in existing cloud RAN (C-RAN) architecture. The C-RAN solution comes into two types [133]. The first one is fully centralized where RRH provides radio function and the baseband functions (layer 1, layer 2, etc) are provided by the base band unit (BBU). The second is partially centralized where layer 1 functionality of baseband function is integrated into the RRH. Both C-RAN solutions comprise RRH, the radio function and antennas (located at remote sites as close to the UEs as possible), mobile fronthaul, the fiber link between RRH and BBUs (which can be distributed or centralized at the central office (CO)). In order to realize SARC in C-RAN (SC-RAN), some RRHs can be deployed at cBS for ubiquitous coverage and the remaining RRHs for data services. The proposed SC-RAN is shown in Fig. 11.

In this figure, SC-RAN is equivalent to traditional C-RAN with decoupled control and data planes. The BBU stack in CO brings flexibility in C-RAN for joint management of resources and the co-existence of control and data BBUs in SC-RAN can extend this flexibility to share signaling, channel conditions (e.g., CSI), and user data. This results into higher potential to perform joint signal processing e.g., CB and CoMP [133]. The adaptive clustering is more manageable in centralized BBUs in SC-RAN due to global control of the coverage area (cBS BBU). The notion of cell-sleeping can be realized and load balancing, mobility management, and interference management can be accomplished more flexibly with reduced OPEX and higher energy efficiency resulting into future green cellular networks.

The flexibility of realizing SC-RAN comes with the expensive requirement of fronthaul/backhaul links. Since, huge information needs to be exchanged between cooperating dBSs in case of CoMP, high capacity fronthaul/backhaul links are required. In order to address the problem of high capacity backhaul requirements, the distributed caching of contents in femtocells has been proposed in [134], [135]. These approaches

⁵An influential node can be identified by utilizing the history/context of different nodes and assigning some weight based on the activity of the node, e.g., time duration of active sessions, file upload/download frequency etc.

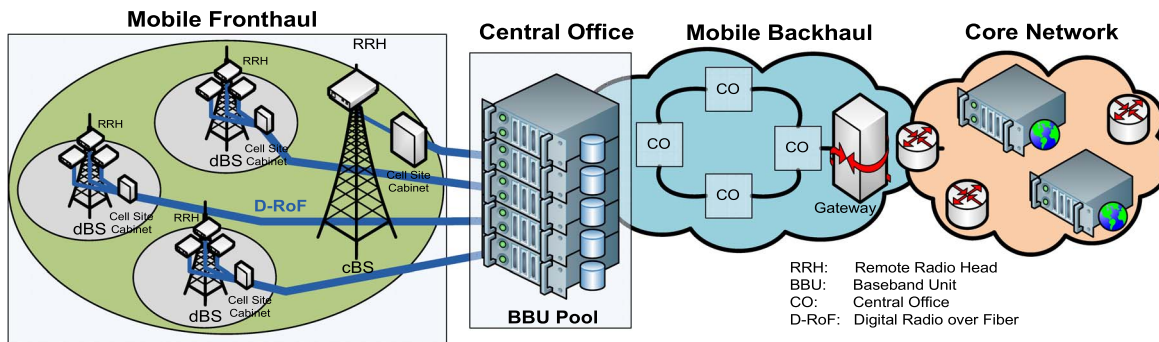


Fig. 11. Split C-RAN architecture.

use high storage capacity at femto BS to cache most popular contents and harnessing D2D communication for content delivery. Recently, the backhaul problem in CoMP has been addressed using cache-enabled relays and BSs [136]–[138]. All these approaches are based on cache-enabled opportunistic cooperative MIMO (CoMP) framework where a portion of contents are cached at cooperating set of relays or BSs to relax backhaul capacity requirements. Such approaches may be used in SC-RAN, where partially centralized C-RAN (with layer 1 functionality integrated into RRH) can be incorporated so that cache-enabled dBSs can provide high-rate data services in CoMP fashion without requiring huge capacity requirements.

V. CONCLUSION

In this article, we outline several performance measures to highlight potential gains and give motivation for evolution of traditional coupled architecture towards control and data planes separation. The different perspectives of energy efficiency, system capacity, interference management and mobility handling are discussed. Since, control and data planes separation approach is in its early stage, little literature exists that addresses some of the performance measures (e.g., [31], [32] evaluates energy and spectral efficiency). Wherever possible, we provided survey of the approaches proposed for separation architecture; otherwise, we provided our view point for potential advantages and associated complexities in SARC. By considering different scenarios from the perspective of outlined performance measures, it is revealed that there is a huge potential for capacity and energy efficiency enhancements by separating control and data planes. Moreover, the SARC provides flexibility in mobility management at the cost of more complex signaling network. The second part of the article provides background for cooperation framework for interference management in multicell environment. It is emphasized that there are several potential advantages of sending CSI to the cBS and exploiting pro-active caching to realize backhaul relaxed CB and CoMP for interference management in future ultra-dense cellular environment.

Another perspective of cooperation has been presented where cooperation means assisting network for common information exchange or content dissemination between near-by devices in the form of ad-hoc or clustered mode direct communication.

D2D CoMP has been introduced where conventional cooperation framework has been suggested to handle intra-cell interference. Due to ubiquitous coverage in SARC, centralized cBS offers more flexibility in CSI acquisition and corresponding beamforming for CB and CoMP operation. The centralized cBS also offers higher degree of freedom to predict nodes for content sharing and it can even be combined with network proactive caching and adaptive clustering for self-organized D2D communication.

Motivated by the control and data planes separation framework, in the following, we outline the lessons learned and several potential research directions in this area:

- Energy efficiency is the most important aspect of future cellular systems. Among many approaches mentioned in Section II-A (e.g., BS switch-off, smart grid, renewable energy sources), dynamic BS switch-off mechanism can play an important role in realizing green cellular communication. The inherent drawback of coverage holes (due to BS switch-off techniques) and more interference (due to increased transmit power in cell range expansion) does not exist in SARC due to ubiquitous coverage. Some of the research studies (e.g., [31], [32], [44]) investigated the potential gains in energy efficiency due to control and data planes separation. In order to investigate full energy efficiency gains, the realistic power consumption models are required. For such models, existing approaches for traditional architecture can be investigated followed by more advanced and sophisticated energy management techniques for SARC.
- The higher spectrum and more bandwidth are envisioned to ensure capacity requirements of future cellular networks. In this context, mm-Wave spectrum and carrier aggregation are potential candidates for next generation cellular networks. A lot of research is being conducted to investigate feasibility of mm-Wave spectrum. Designing new channel models for dual connectivity (i.e., mm-Wave for data plane and lower frequency for control plane) has a lot of research potential that can lead towards communication in SARC.
- For current HetNet, intra-cell interference does not exist and inter-cell interference management has been standardized. In future ultra-dense networks, intra-cell

interference will again be a problem due to underlay systems e.g., D2D communication. In order to overcome this interference, existing techniques of CB and CoMP can be extended at device level (D2D) and the backhaul limitations can be complemented by exploiting proactive caching techniques (e.g., [136]–[138]).

- In future ultra-dense environment, cells at mm-Wave spectrum will have spot beam coverage. This results in huge capacity enhancements which can further be leveraged by harnessing D2D cooperation for content sharing or content dissemination.
- Mobility management is flexible due to higher degree of freedom in SARC. However, this comes at the price of complex signaling network in SARC. The control plane design will be more complex due to more tiers (underlay networks). In this context, lot of research endeavors are required to realize seamless handovers and higher coverage probability while ensuring QoS requirements of each user.

ACKNOWLEDGMENT

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

REFERENCES

- [1] S. Cass, "IP Traffic in 2017: 1.4 Zettabytes," Feb. 2014. [Online]. Available: <http://spectrum.ieee.org/telecom/internet/ip-traffic-in-2017-14-zettabytes>. Accessed: Jun. 2, 2015.
- [2] "UMTS Forum Report: Mobile Traffic Forecasts: 2010–2020," Universal Mobile Telecommunications System (UMTS), Rep. 44, May 2011. [Online]. Available: http://www.umts-forum.org/component/option,com_docman/task,doc_download/gid,2348/Itemid,213/. Accessed: Jun. 2, 2015.
- [3] A. Osseiran *et al.*, "The foundation of the mobile and wireless communications system for 2020 and beyond: Challenges, enablers and technology solutions," in *Proc. IEEE VTC*, Jun. 2013, pp. 1–5.
- [4] "More than 50 billion connected devices," Ericsson, Plano, TX, USA, White Paper, Feb. 2011. [Online]. Available: http://www.akos-rs.si/files/Telekomunikacije/Digitalna_agenda/Internetni_protokol_Ipv6/More-than-50-billion-connected-devices.pdf. Accessed: Jun. 2, 2015.
- [5] J. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. Reed, "Femto-cells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.
- [6] A. Bianzino, C. Chaudet, D. Rossi, and J. Rougier, "A Survey of Green Networking Research," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 1, pp. 3–20, 1st Quart. 2012.
- [7] J. Wu, S. Rangan, and H. Zhang, Eds., *Green Communications: Theoretical Fundamentals, Algorithms and Applications*, 1st ed. Boca Raton, FL, USA: CRC Press, Sep. 2012.
- [8] H. Scott Matthews *et al.*, "Planning energy-efficient and eco-sustainable telecommunications networks," *Bell Labs Tech. J.*, vol. 15, no. 1, pp. 215–236, Jun. 2010.
- [9] 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, Aug. 2011.
- [10] C. Gao *et al.*, "Relax, but do not sleep: A new perspective on green wireless networking," in *Proc. IEEE INFOCOM*, Apr. 2014, pp. 907–915.
- [11] M. Shakir *et al.*, "Green heterogeneous small-cell networks: Toward reducing the CO₂ emissions of mobile communications industry using uplink power adaptation," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 52–61, Jun. 2013.
- [12] A. Dejonghe *et al.*, "Green reconfigurable radio systems," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 90–101, May 2007.
- [13] G. Rittenhouse, S. Goyal, D. Neilson, and S. Samuel, "Sustainable telecommunications," in *Proc. Tech. Symp. ITU WT*, Oct. 2011, pp. 19–23.
- [14] ABResearch, "Market Gets Primed to Rollout Half a Million Outdoor Small Cells in 2013," Jan. 2013. [Online]. Available: <https://www.abiresearch.com/press/market-gets-primed-to-rollout-half-a-million-outdo/>. Accessed: Jun. 6, 2015.
- [15] P. Marsch, S. Khatkhat, and G. Fettweis, "A framework for determining realistic capacity bounds for distributed antenna systems," in *Proc. IEEE ITW*, 2006, pp. 571–575.
- [16] "Coordinated multi-point operation for LTE physical layer aspects," 3rd Generation Partnership Project, Sophia Antipolis Cedex, France, TS 36.819 V11.1.0, Dec. 2011.
- [17] P. Marsch and G. P. Fettweis, Eds., *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
- [18] S. Parkvall *et al.*, "LTE-Advanced—Evolving LTE towards IMT-Advanced," in *Proc. IEEE VTC*, 2008, pp. 1–5.
- [19] "GreenTouch-Project." [Online]. Available: http://www.greentouch.org/index.php?page=mobile_networks_working_group_projects. Accessed: Jun. 6, 2015.
- [20] "METIS 2020." [Online]. Available: <https://www.metis2020.com>. Accessed: Jun. 6, 2015.
- [21] "MiWEBA—Project." [Online]. Available: http://www.miweba.eu/?page_id=80. Accessed: Jun. 6, 2015.
- [22] E. Ternon, Z. Bharucha, and H. Taoka, "A feasibility study for the detached cell concept," in *Proc. Int. ITG Conf. SCC*, Jan. 2013, pp. 1–5.
- [23] A. Capone, A. Fonseca dos Santos, I. Filippini, and B. Gloss, "Looking beyond green cellular networks," in *Proc. Annu. Conf. WONS*, 2012, pp. 127–130.
- [24] 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.
- [25] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.
- [26] J. Louhi, "Energy efficiency of modern cellular base stations," in *Proc. INTELEC*, 2007, pp. 475–476.
- [27] M. Olsson *et al.*, "5GrEEn: Towards Green 5G mobile networks," in *IEEE Int. Conf. WiMob*, Oct. 2013, pp. 212–216.
- [28] T. Zhao *et al.*, "Software defined radio implementation of signaling splitting in hyper-cellular network," in *Proc. 2nd Workshop Softw. Radio Implement. Forum*, 2013, pp. 81–84.
- [29] H. Ali-Ahmad *et al.*, "An SDN-based network architecture for extremely dense wireless networks," in *Proc. IEEE SDN4FNS*, Nov. 2013, pp. 1–7.
- [30] A. Zakrzewska, D. Lopez-Perez, S. Kucera, and H. Claussen, "Dual connectivity in LTE HetNets with split control- and user-plane," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 391–396.
- [31] H. Ishii, Y. Kishiyama, and H. Takahashi, "A novel architecture for LTE-B :C-plane/U-plane split and Phantom Cell concept," in *Proc. IEEE GLOBECOM*, Dec. 2012, pp. 624–630.
- [32] S. Mukherjee and H. Ishii, "Energy efficiency in the Phantom Cell enhanced local area architecture," in *Proc. IEEE WCNC*, Apr. 2013, pp. 1267–1272.
- [33] Y. Okumura, "5G mobile radio access system using SHF/EHF bands," in *Proc. APMC*, 2014, pp. 908–910.
- [34] S. Suyama, J. Shen, A. Benjebbour, Y. Kishiyama, and Y. Okumura, "Super high bit rate radio access technologies for small cells using higher frequency bands," in *Proc. IEEE MTT-S IMS*, Jun. 2014, pp. 1–4.
- [35] K. E. S. WWRf and L. Sorensen, "User scenarios 2020," Jul. 2009. [Online]. Available: <http://www.wireless-world-research.org>. Accessed: Jun. 2, 2015.
- [36] R. Tafazolli, *Technologies for the Wireless Future: Wireless World Research Forum (WWRf)*. New York, NY, USA: Wiley, 2006.
- [37] Y. Q. Bian and D. Rao, "Small cells big opportunities," Huawei, Shenzhen, China, Feb. 2014. [Online]. Available: http://smallcellforum.org/smallcellforum/Files/File/SCF-Small_Cells_White_Paper.pdf. Accessed: Jun. 2, 2015.
- [38] F. Donovan, "Infonetics: In-building, outdoor small cells to handle quarter of mobile traffic by 2016," FierceMobileIT, Washington, DC, USA, Jan. 2013. [Online]. Available: <http://www.fiercemobileit.com>. Accessed: Jun. 2, 2015.
- [39] "Small cells—What's the big idea?" Small Cell Forum, Feb. 2012. [Online]. Available: http://smallcellforum.org/smallcellforum/Files/File/SCF-Small_Cells_White_Paper.pdf. Accessed: Jun. 2, 2015.
- [40] D. Ferling *et al.*, "Energy efficiency approaches for radio nodes," in *Proc. FutureNetworkSummit*, Jun. 2010, pp. 1–9.
- [41] L. Correia *et al.*, "Challenges and enabling technologies for energy aware mobile radio networks," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 66–72, Nov. 2010.

- [42] H. Holtkamp, G. Auer, V. Giannini, and H. Haas, "A parameterized base station power model," *IEEE Commun. Lett.*, vol. 17, no. 11, pp. 2033–2035, Nov. 2013.
- [43] M. Herlich and H. Karl, "Energy-efficient assignment of user equipment to cooperative base stations," in *Proc. ISWCS*, Aug. 2013, pp. 1–5.
- [44] S. Zhang, J. Wu, J. Gong, S. Zhou, and Z. Niu, "Energy-optimal probabilistic base station sleeping under a separation network architecture," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 4239–4244.
- [45] M. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular access networks," in *Proc. IEEE ICC Workshop*, Jun. 2009, pp. 1–5.
- [46] L. Chiaraviglio, D. Ciullo, M. Meo, and M. Marsan, "Energy-efficient management of UMTS access networks," in *Proc. ITC*, Sep. 2009, pp. 1–8.
- [47] E. Oh and B. Krishnamachari, "Energy savings through dynamic base station switching in cellular wireless access networks," in *Proc. IEEE GLOBECOM*, Dec. 2010, pp. 1–5.
- [48] A. Fehske, F. Richter, and G. Fettweis, "Energy efficiency improvements through micro sites in cellular mobile radio networks," in *Proc. IEEE GLOBECOM*, Nov. 2009, pp. 1–5.
- [49] P. Rost and G. Fettweis, "Green communications in cellular networks with fixed relay nodes," in *Cooperative Cellular Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, pp. 300–323, 2010.
- [50] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-off strategies for green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2126–2136, May 2013.
- [51] 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, Sep. 2011.
- [52] H. Al Haj Hassan, L. Nuaymi, and A. Pelov, "Renewable energy in cellular networks: A survey," in *Proc. IEEE Online Conf. GreenCom*, Oct. 2013, pp. 1–7.
- [53] L.-C. Wang and S. Rangapillai, "A survey on green 5G cellular networks," in *Proc. Int. Conf. SPCOM*, Jul. 2012, pp. 1–5.
- [54] "FOMA base station using solar and wind power," NTT DOCOMO, Tokyo, Japan, Jul. 2004. [Online]. Available: <https://www.nttdocomo.co.jp/english/corporate/csr/ecology/>. Accessed: Jun. 2, 2015.
- [55] Z. P. Yang Tianjian, H. Yiwen, and D. Pamlin, "Low carbon telecommunications solutions in China," Jul. 2010. [Online]. Available: http://www.pamlin.net/other_documents/ChinaMobilereport-summary-en.pdf. Accessed: Jun. 2, 2015.
- [56] S. Bu, F. Yu, and P. Liu, "A game-theoretical decision-making scheme for electricity retailers in the smart grid with demand-side management," in *Proc. IEEE Int. Conf. SmartGridComm*, Oct. 2011, pp. 387–391.
- [57] S. Bu, F. Yu, Y. Cai, and P. Liu, "Energy efficient cellular networks with CoMP communications and smart grid," in *Proc. IEEE ICC*, Jun. 2012, pp. 5921–5925.
- [58] S. Bu, F. Yu, Y. Cai, and X. Liu, "When the smart grid meets energy-efficient communications: Green wireless cellular networks powered by the smart grid," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 3014–3024, Aug. 2012.
- [59] 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, Feb. 2011.
- [60] "Estimated spectrum bandwidth requirements for the future development of IMT-2000 and IMT-advanced," Int. Telecommun. Union, Geneva, Switzerland. [Online]. Available: <http://www.itu.int/pub/R-REP-M.2078>. Accessed: Jun. 2, 2015.
- [61] R. Zhang, Z. Zheng, M. Wang, X. Shen, and L.-L. Xie, "Equivalent capacity in carrier aggregation-based LTE-A systems: A probabilistic analysis," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6444–6460, Nov. 2014.
- [62] "5G radio access: Requirements, concept and technologies," pp. 1–13. [Online]. Available: https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/whitepaper_5g/DOCOMO_5G_White_Paper.pdf. Accessed: Jun. 2, 2015.
- [63] "Artemis." [Online]. Available: <http://www.artemis.com/pcell>. Accessed: Jun. 2, 2015.
- [64] S. Rangan, T. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proc. IEEE*, vol. 102, no. 3, pp. 366–385, Mar. 2014.
- [65] R. Daniels and R. Heath, "60 GHz wireless communications: Emerging requirements and design recommendations," *IEEE Veh. Technol. Mag.*, vol. 2, no. 3, pp. 41–50, Sep. 2007.
- [66] M. Marcus and B. Pattan, "Millimeter wave propagation: Spectrum management implications," *IEEE Microw. Mag.*, vol. 6, no. 2, pp. 54–62, Jun. 2005.
- [67] J. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [68] M. Jacob *et al.*, "Diffraction in mm and sub-mm wave indoor propagation channels," *IEEE Trans. Microw. Theory Tech.*, vol. 60, no. 3, pp. 833–844, Mar. 2012.
- [69] M. Kyro, V. Kolmonen, and P. Vainikainen, "Experimental propagation channel characterization of mm-wave radio links in urban scenarios," *IEEE Antennas Wireless Propag. Lett.*, vol. 11, pp. 865–868, Jul. 2012.
- [70] S. Larew, T. Thomas, M. Cudak, and A. Ghosh, "Air interface design and ray tracing study for 5G millimeter wave communications," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 117–122.
- [71] T. Rappaport *et al.*, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, May 2013.
- [72] W. Roh *et al.*, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 106–113, Feb. 2014.
- [73] T. Janevski, "5G mobile phone concept," in *Proc. IEEE CCNC*, Jan. 2009, pp. 1–2.
- [74] A. Tudzarov and T. Janevski, "Protocols and algorithms for the next generation 5G mobile systems," *Netw. Protocols Algorithms*, vol. 3, no. 1, pp. 94–114, Jun. 2011.
- [75] "5G the Nanocore," pp. 1–24. [Online]. Available: http://www.telecoms.com/wp-content/blogs.dir/1/files/2011/05/5G_The_NanoCore.pdf. Accessed: Jun. 2, 2015.
- [76] M. Rahman and H. Yanikomeroglu, "Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1414–1425, Apr. 2010.
- [77] C. Kosta, B. Hunt, A. Qudus, and R. Tafazolli, "On interference avoidance through inter-cell interference coordination (ICIC) based on OFDMA mobile systems," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 973–995, 3rd 2013.
- [78] G. Boudreau *et al.*, "Interference coordination and cancellation for 4G networks," *IEEE Commun. Mag.*, vol. 47, no. 4, pp. 74–81, Apr. 2009.
- [79] Z. Xie and B. Walke, "Frequency reuse techniques for attaining both coverage and high spectral efficiency in OFDMA cellular systems," in *Proc. IEEE WCNC*, Apr. 2010, pp. 1–6.
- [80] L. Dong, Z. Song, L. Wenxin, and W. Wenbo, "A frequency reuse partitioning scheme with successive interference cancellation for OFDMA uplink transmission," in *Proc. IEEE Int. Symp. PIMRC*, Sep. 2009, pp. 1362–1366.
- [81] S. Deb, P. Monogioudis, J. Miernik, and J. Seymour, "Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets," *IEEE/ACM Trans. Netw.*, vol. 22, no. 1, pp. 137–150, Feb. 2014.
- [82] D. Lopez-Perez *et al.*, "Enhanced inter-cell interference coordination challenges in heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 22–30, Jun. 2011.
- [83] A. Liu, V. Lau, L. Ruan, J. Chen, and D. Xiao, "Hierarchical radio resource optimization for heterogeneous networks with enhanced inter-cell interference coordination (eICIC)," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1684–1693, Apr. 2014.
- [84] M. Karabacak, D. Wang, H. Ishii, and H. Arslan, "Mobility performance of macrocell-assisted small cells in Manhattan model," in *Proc. IEEE VTC*, May 2014, pp. 1–5.
- [85] C.-H. Lee and Z.-S. Syu, "Handover analysis of macro-assisted small cell networks," in *Proc. IEEE Int. Conf. iThings, IEEE GreenCom IEEE Cyber, CPSCom*, Sep. 2014, pp. 604–609.
- [86] S. Kuklinski, Y. Li, and K. T. Dinh, "Handover management in SDN-based mobile networks," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 194–200.
- [87] L. Valtulina, M. Karimzadeh, G. Karagiannis, G. Heijnen, and A. Pras, "Performance evaluation of a SDN/openflow-based distributed mobility management (DMM) approach in virtualized LTE systems," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 18–23.
- [88] D. Kreutz *et al.*, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14–76, Jan. 2015.
- [89] A. Sniady and J. Soler, "LTE for railways: Impact on performance of ETCS railway signaling," *IEEE Veh. Technol. Mag.*, vol. 9, no. 2, pp. 69–77, Jun. 2014.

- [90] L. Yan and X. Fang, "Decoupled wireless network architecture for high-speed railway," in *Proc. IEEE Int. Workshop HMWC*, Nov. 2013, pp. 96–100.
- [91] H. Y. Choi, Y. Song, and Y.-K. Kim, "Standards of future railway wireless communication in Korea," *Recent Adv. Comput. Eng., Commun. Inf. Technol.*, pp. 360–367, 2014. [Online]. Available: <http://www.wseas.us/e-library/conferences/2014/Tenerife/INFORM/INFORM-50.pdf>. Accessed: Jun. 2, 2015.
- [92] "Evolved universal terrestrial radio access, radio resource control," 3rd Generation Partnership Project, Sophia Antipolis Cedex, France, TS 36.331 V11.5.0, Sep. 2013.
- [93] C. W. Ahn and J.-H. Lee, "Wireless cooperative communication: A survey," in *Proc. ICUIMC*, 2011, p. 78.
- [94] M. G. Rubinstein, I. M. Moraes, M. E. M. Campista, L. H. M. K. Costa, and O. C. M. B. Duarte, "A survey on wireless Ad Hoc networks," in *Mobile and Wireless Communication Networks*, vol. 211, G. Pujolle, Ed. New York, NY, USA: Springer-Verlag, 2006, pp. 1–33. [Online]. Available: http://link.springer.com/chapter/10.1007/978-0-387-34736-3_1
- [95] B. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.
- [96] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Syst., Electron.*, Sep. 1998, pp. 295–300.
- [97] X. Wang and H. Poor, "Iterative (turbo) soft interference cancellation and decoding for coded CDMA," *IEEE Trans. Commun.*, vol. 47, no. 7, pp. 1046–1061, Jul. 1999.
- [98] H. Dai, A. Molisch, and H. Poor, "Downlink capacity of interference-limited MIMO systems with joint detection," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 442–453, Mar. 2004.
- [99] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.
- [100] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 768–783, May 2004.
- [101] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.
- [102] A. Garavaglia, R. Weber, M. Schulist, and S. Brueck, "Adaptive cell clustering in a multi-cluster environment," U.S. Patent US 8 639 256 B2, Jan. 2014.
- [103] M. Corson *et al.*, "Toward proximity-aware internetworking," *IEEE Wireless Commun.*, vol. 17, no. 6, pp. 26–33, Dec. 2010.
- [104] S. Balraj, "LTE direct overview." [Online]. Available: <http://s3.amazonaws.com/sdieee/205-LTE+Direct+IEEE+VTC+San+Diego.pdf>. Accessed: Jun. 2, 2015.
- [105] "Radio access and spectrum FP7—Future networks cluster." [Online]. Available: <http://www.ict-ras.eu/index.php/ras-projects/moto>. Accessed: Jun. 2, 2015.
- [106] E. Bastug, M. Bennis, and M. Debbah, "Social and spatial proactive caching for mobile data offloading," in *Proc. IEEE ICC Workshop*, Jun. 2014, pp. 581–586.
- [107] Y. Li, T. Zhou, J. Xu, Z. Li, and H. Wang, "Adaptive TDD UL/DL slot utilization for cellular controlled D2D communications," in *Proc. GMC*, Oct. 2011, pp. 1–6.
- [108] H. Sun, M. Sheng, M. Wildemeersch, and T. Quek, "Modeling of D2D enhanced two-tier dynamic TDD heterogeneous cellular networks," in *Proc. IEEE/CIC ICC*, Oct. 2014, pp. 609–614.
- [109] H. Sun, M. Wildemeersch, M. Sheng, and T. Quek, "D2D enhanced heterogeneous cellular networks with dynamic TDD," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4204–4218, Aug. 2015.
- [110] S. Kim and W. Stark, "Full duplex device to device communication in cellular networks," in *Proc. ICNC*, Feb. 2014, pp. 721–725.
- [111] S. Ali, N. Rajatheva, and M. Latva-Aho, "Full duplex device-to-device communication in cellular networks," in *Proc. EuCNC*, Jun. 2014, pp. 1–5.
- [112] K. Hemachandra, N. Rajatheva, and M. Latva-Aho, "Sum-rate analysis for full-duplex underlay device-to-device networks," in *Proc. IEEE WCNC*, Apr. 2014, pp. 514–519.
- [113] S. Ali, A. Ghazanfari, N. Rajatheva, and M. Latva-aho, "Effect of residual of self-interference in performance of full-duplex D2D communication," in *Proc. Int. Conf. 5GU Connectivity*, Nov. 2014, pp. 46–51.
- [114] C. B. Das, "A study on device to device communication in wireless mobile network," *Int. J. Modern Commun. Technol. Res.*, vol. 3, no. 3, pp. 1–18, Mar. 2015.
- [115] S. Mumtaz and J. Rodriguez, Eds., *Smart Device to Smart Device Communication*. Berlin, Germany: Springer-Verlag, 2014. [Online]. Available: <http://link.springer.com/10.1007/978-3-319-04963-2>
- [116] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.
- [117] S. Schwarz and M. Rupp, "Exploring coordinated multipoint beamforming strategies for 5G cellular," *IEEE Access*, vol. 2, pp. 930–946, Aug. 2014.
- [118] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 1801–1819, 4th Quart., 2014.
- [119] J. Liu, N. Kato, J. Ma, and N. Kadowaki, "Device-to-device communication in LTE-advanced networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 1923–1940, Fourth Quarter 2015.
- [120] B. E. Godana and D. Gesbert, "Coordinated beamforming in multicell networks with Channel State Information exchange delays," in *Proc. IEEE Int. Symp. PIMRC*, Sep. 2013, pp. 713–718.
- [121] K. Huang, B. Mondal, R. Heath, and J. Andrews, "CTH07-1: Effect of feedback delay on multi-antenna limited feedback for temporally-correlated channels," in *Proc. IEEE GLOBECOM*, Nov. 2006, pp. 1–5.
- [122] R. Bhagavatula and R. Heath, "Impact of delayed limited feedback on the sum-rate of intercell interference nulling," in *Proc. IEEE ICC*, Jun. 2011, pp. 1–5.
- [123] R. Bhagavatula and R. Heath, "Adaptive bit partitioning for multicell intercell interference nulling with delayed limited feedback," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3824–3836, Aug. 2011.
- [124] C. Tan and N. Beaulieu, "On first-order Markov modeling for the Rayleigh fading channel," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2032–2040, Dec. 2000.
- [125] R. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Techn. J.*, vol. 47, no. 6, pp. 957–1000, Jul. 1968.
- [126] H. Mustafa *et al.*, "Spectral efficiency improvements in HetNets by exploiting device-to-device communications," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 857–862.
- [127] E. Yan and Y. Ding, "Applying centrality measures to impact analysis: A coauthorship network analysis," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 10, pp. 2107–2118, 2009.
- [128] C. Ni, C. Sugimoto, and J. Jiang, "Degree, closeness, and betweenness: Application of group centrality measurements to explore macro-disciplinary evolution diachronically," in *Proc. Int. Conf. ISSI*, 2011, pp. 1–13.
- [129] S. Mumtaz, K. Saidul Huq, and J. Rodriguez, "Coordinated paradigm for D2D communications," in *Proc. IEEE Conf. INFOCOM WKSHP*, Apr. 2014, pp. 718–723.
- [130] Ericsson, "Physical layer aspects of dual connectivity," 3rd Generation Partnership Project, Sophia Antipolis Cedex, France, 3GPP Std. Contrib. R1-130566, St. Julian's, Malta, Feb. 2013.
- [131] K. Sakaguchi *et al.*, "Cloud cooperated heterogeneous cellular networks," in *Proc. Int. Symp. ISPCS*, Nov. 2013, pp. 787–791.
- [132] 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, Mar. 2015.
- [133] "C-RAN the road towards green RAN," pp. 1–44. [Online]. Available: http://labs.chinamobile.com/cran/wp-content/uploads/CRAN_white_paper_v2_5_EN.pdf. Accessed: Jun. 2, 2015.
- [134] N. Golrezaei, K. Shanmugam, A. Dimakis, A. Molisch, and G. Caire, "FemtoCaching: Wireless video content delivery through distributed caching helpers," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1107–1115.
- [135] N. Golrezaei, A. Molisch, A. Dimakis, and G. Caire, "FemtoCaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Commun. Mag.*, vol. 51, no. 4, pp. 142–149, Apr. 2013.
- [136] A. Liu and V. Lau, "Mixed-timescale precoding and cache control in cached MIMO interference network," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6320–6332, Dec. 2013.
- [137] A. Liu and V. Lau, "Cache-enabled opportunistic cooperative MIMO for video streaming in wireless systems," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 390–402, Jan. 2014.
- [138] A. Liu and V. Lau, "Exploiting base station caching in MIMO cellular networks: Opportunistic cooperation for video streaming," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 57–69, Jan. 2015.



Hafiz Atta Ul Mustafa received the B.E degree in avionics engineering from the National University of Sciences and Technologies and the M.Sc. degree in telecommunication engineering from National University of Computer & Emerging Sciences, Pakistan, in 2003 and 2010, respectively. He is currently pursuing the Ph.D. degree with the Institute for Communication Systems (ICS—formerly known as CCSR) at University of Surrey, U.K. He started as Design Engineer in NexTek Service, where he was involved in embedded system design, telecommunication system engineering, product evaluation, and project management. He has also provided consultancy services in satellite communication systems to the Ministry of Defense, Pakistan, from 2005 to 2013. His main research interests include digital signal processing, telecommunication link analysis, stochastic geometry, and optimization of communication systems.



Ali Imran is an Assistant Professor in telecommunications at the University of Oklahoma, Tulsa, OK, USA. 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 Conference on Communications, IEEE Wireless Communications and Networking Conference European Wireless, and CrownCom on these topics. He is an Associate Fellow of the Higher Education Academy (AFHEA), U.K., and a member of the advisory board to the Special Technical Community on Big Data of the IEEE Computer Society.



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) with the Institute for Communication Systems (ICS—formerly known as CCSR) at the University of Surrey, U.K. He is leading a number of multimillion international research projects encompassing the areas of energy efficiency, fundamental performance limits, sensor networks, and self-organizing 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 Ph.D. 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 been 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 Fellow of Higher Education Academy (SFHEA), U.K.



Rahim Tafazolli (SM'09) is a Professor and the Director of the Institute for Communication Systems (ICS) and 5G Innovation Centre (5GIC), University of Surrey, U.K. He has published more than 500 research papers in refereed journals, international conferences, and as invited speaker. He is the Editor of two books, *Technologies for Wireless Future* (vols. 1 and 2; Wiley, 2006). He was appointed as Fellow of the 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.



Muhammad Zeeshan Shakir (M'03) received the Ph.D. in electronic and electrical engineering from the University of Strathclyde, Glasgow, UK, in 2010. He has been an Assistant Research Scientist at Texas A&M University at Qatar in Doha, Qatar, since July 2012. From January 2006 to September 2009, he was the joint recipient of an industrial research fund and a prestigious overseas research scholarship by the University of Strathclyde. His research interests include design and deployment of diverse wireless communication systems, including hyper-dense heterogeneous small-cell networks and Green communication systems. He has published more than 75 technical journal and conference papers and has contributed to six books, all in reputable venues. He is co-author of two research monographs. Most of his research has been sponsored by Qatar National Research Fund and national industrial partners. He has been/is giving tutorials on emerging Green wireless communication systems at IEEE flagship conferences including IEEE BlackSeaCom 2015, Constanta, IEEE Globecom 2015, San Diego, IEEE ICC 2014, Sydney, and IEEE GLOBECOM 2014, Austin. He has been/is serving as an associate technical editor/lead guest editor for IEEE COMMUNICATION MAGAZINE, IEEE WIRELESS COMMUNICATIONS MAGAZINE, and IET Journal on Signal Processing. He has been/is serving as technical co-chair and member of technical program committee of several special sessions/workshops and symposia in flagship conferences, such as ICC, Globecom, and GlobalSIP. Currently, he has been serving as an elected secretary to the IEEE DySPAN 1900.7 working group and the standard representative of ComSoc technical committee on Signal processing and communication electronics. He is an active member of IEEE Standard Association.

He has published more than 75 technical journal and conference papers and has contributed to six books, all in reputable venues. He is co-author of two research monographs. Most of his research has been sponsored by Qatar National Research Fund and national industrial partners. He has been/is giving tutorials on emerging Green wireless communication systems at IEEE flagship conferences including IEEE BlackSeaCom 2015, Constanta, IEEE Globecom 2015, San Diego, IEEE ICC 2014, Sydney, and IEEE GLOBECOM 2014, Austin. He has been/is serving as an associate technical editor/lead guest editor for IEEE COMMUNICATION MAGAZINE, IEEE WIRELESS COMMUNICATIONS MAGAZINE, and IET Journal on Signal Processing. He has been/is serving as technical co-chair and member of technical program committee of several special sessions/workshops and symposia in flagship conferences, such as ICC, Globecom, and GlobalSIP. Currently, he has been serving as an elected secretary to the IEEE DySPAN 1900.7 working group and the standard representative of ComSoc technical committee on Signal processing and communication electronics. He is an active member of IEEE Standard Association.