Coordinated Multi-Point Clustering Schemes: A Survey

Selcuk Bassoy, Hasan Farooq, Muhammad A. Imran, *Senior Member, IEEE*, and Ali Imran, *Member, IEEE*

Abstract—Mobile data traffic grew by 74% in 2015 and it is expected to grow eight-fold by 2020. Future wireless networks will need to deploy massive number of small cells to cope with this increasing demand. Dense deployment of small cells will require advanced interference mitigation techniques to improve spectral efficiency and enhance much needed capacity. Coordinated multi-point (CoMP) is a key feature for mitigating inter-cell interference, improve throughput and cell edge performance. However, cooperation will need to be limited to few cells only due to additional overhead required by CoMP due to channel state information (CSI) exchange, scheduling complexity, and additional backhaul limitation. Hence, small CoMP clusters will need to be formed in the network. This paper surveys the state-of-the-art on one of the key challenges of CoMP implementation: CoMP clustering. As a starting point, we present the need for CoMP, the clustering challenge for 5G wireless networks and provide a brief essential background about CoMP and the enabling network architectures. We then provide the key framework for CoMP clustering and introduce self organization as an important concept for effective CoMP clustering to maximize CoMP gains. Next, we present two novel taxonomies on existing CoMP clustering solutions, based on self organization and aimed objective function. Strengths and weaknesses of the available clustering solutions in the literature are critically discussed. We then discuss future research areas and potential approaches for CoMP clustering. We present a future outlook on the utilization of big data in cellular context to support proactive CoMP clustering based on prediction modeling. Finally, we conclude this paper with a summary of lessons learned in this field. This paper aims to be a key guide for anyone who wants to research on CoMP clustering for future wireless networks.

Index Terms—Coordinated multi-point, CoMP clustering, 5G.

I. INTRODUCTION

FUTURE wireless cellular networks will be under tremendous pressure with the increasing data demand as the user behaviour changes with popular high bandwidth applications. While smart phones become very popular, high bandwidth hungry applications like video streaming, multimedia file sharing etc becomes more popular. Mobile data traffic grew by

S. Bassoy and M. A. Imran are with the School of Engineering, University of Glasgow, Glasgow G12 8QQ, U.K. (e-mail: s.bassoy.1@research.gla.ac.uk; muhammad.imran@glasgow.ac.uk).

H. Farooq and A. Imran are with the School of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK 74135 USA (e-mail: hasan.farooq@ou.edu; ali.imran@ou.edu;).
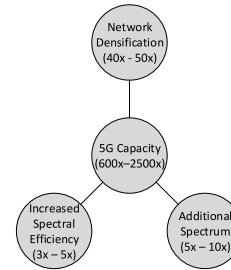
Fig. 1. Proposed Capacity Enhancements for 5G.

74% in 2015 and it's expected to grow 8-fold by 2020 [1]. Moreover, a 1000 fold increase in mobile data traffic is expected for 5G beyond 2020 [2]. To enable 5G to cope with this tremendous increase in data growth, following three development areas in the emerging wireless landscape are proposed [2]–[4].

1) Network Densification - Massive Small cell deployment
2) Increased Spectral Efficiency - CoMP, Multiple Input-Multiple Output (MIMO), Enhanced coding techniques
3) Additional Spectrum

Figure 1 illustrates the potential capacity gains expected from each of the three key capacity enhancement proposed for 5G [2]–[4]. Biggest capacity gains are expected from network densification: a massive deployment of small cells will be required [5], [6] in search for additional capacity. Dense small cell deployment in heterogeneous cellular networks (HetNet) will lead to a severely interference limited network depending on the available frequency spectrum. More advanced inter-cell interference mitigation techniques will need to be deployed to combat interference and improve spectral efficiency. Improved spectral efficiency will lead to much needed capacity enhancement as highlighted above as one of the three key development areas for 5G.

CoMP or Network MIMO is the emerging technology which has been proposed to reduce inter-cell interference and hence improve high data rate coverage and cell edge throughout for future wireless networks. CoMP has been introduced for long term evolution advanced (LTE-A) by the third generation partnership project (3GPP) in Release 11 [7] and it is likely to be a key feature of 5G [2]. However, coordination between all cells in the network is a very complex task, due to precise synchronisation requirement within coordinated cells, additional pilot overhead, additional signal processing, complex beamforming design and scheduling among all base stations (BSs). It will require high bandwidth backhaul links due to CSI and/or user data exchange between all BSs [8], [9].

To reduce this overhead, smaller size cooperation clusters are required where coordination only takes place within the cluster. Optimal CoMP clustering is one of the key challenges for CoMP implementation for future wireless networks. Selecting the right group of BSs for cooperation for a given user profile is key to maximise potential CoMP gains. Trade-off between the overhead and interference cancellation benefits needs to be taken into account for optimum cluster size design. There are multiple objectives for CoMP clustering and the right balance between the various efficiency/overhead indicators is a challenge. For example, maximising spectral efficiency with CoMP clustering can degrade energy efficiency and backhaul limitations may prevent such cluster design. Hence a comprehensive clustering approach should be considered to achieve the right balance between multiple objectives of future networks such as energy efficiency, load balancing and spectral efficiency. Main scope of this article is to provide an extensive survey of CoMP clustering techniques in the literature over the last decade. We provide a novel taxonomy on CoMP clustering techniques, critically discuss the strengths and weaknesses of the available solutions in the literature. The rest of the article is structured as follows.

In Section II, we review the relevant work on CoMP clustering and show our novel contribution with this survey. In Section III, we provide an essential background about CoMP to the reader, main types of CoMP implementation, associated challenges and the enabling network architectures are presented. In Section IV, we introduce a key framework for CoMP clustering challenge and present self organising networks (SON) as a important platform to implement effective dynamic CoMP clustering algorithms. In Section V, a novel self-organisation based taxonomy on CoMP clustering in the literature is introduced. Various CoMP clustering approaches are discussed and criticised based on self organisation, complexity, scalability and practical use. In Section VI, a further taxonomy is introduced based on the aimed objective function of CoMP clustering. An extensive survey of existing clustering approaches based on different objective functions like spectral efficiency, energy efficiency, load balancing and backhaul optimisation are presented and criticised in detail. In Section VII, we discuss open research areas for CoMP clustering and present potential approaches. Big Data empowered prediction based CoMP clustering is identified as an important open research area for much needed low latency in future wireless networks. Big Data aided spatio-temporal channel prediction, user mobility and user profile predictions and their potential use in proactive CoMP cluster decision making is detailed. Furthermore, we present future research directions on dynamic clustering and identify the need for comprehensive multi-objective CoMP clustering in this section. Finally in Section VIII, we conclude with summary of lessons learnt in CoMP clustering. The list of acronyms used in this paper is listed in Table I.

II. RELATED WORK

A number of works have already been conducted for CoMP in general [5], [8], [10] and more specifically for LTE-A

TABLE I
LIST OF ACRONYMS

Acronym	Definition
3GPP	Third Generation Partnership Project
BBU	Baseband Processing Unit
BS	Base Station
CAPEX	Capital Expenditure
CB	Coordinated Beamforming
CCU	CoMP Control Unit
CDR	Call Data Record
CDSA	Control and Data Plane Seperation Architecture
CoMP	Coordinated Multi-Point
C-RAN	Cloud Radio Access Networks
CS	Coordinated Scheduling
CSI	Channel State Information
DAS	Distributed Antenna System
DPS	Dynamic Point Selection
eICIC	Enhanced Inter-cell Interference Coordination
HetNet	Heterogeneous Cellular Network
JT	Joint Transmission
LTE-A	Long Term Evolution Advanced
MDT	Minimisation of Drive Tests
MIMO	Multiple Input Multiple Output
MU	Multi-user
OPEX	Operational Expenditure
QoE	Quality of Experience
QoS	Quality of Service
RAN	Radio Access Network
RRU	Remote Radio Unit
RSRP	Reference Signal Received Power
RSRQ	reference signal received quality
RSSI	Received Signal Strength Indicator
SC	Small Cell
SINR	Signal-to-Interference-plus-Noise Ratio
SON	Self Organising Networks
TP	Transmission Point
UE	User Equipment

implementation in [9] and [11] . Deployment scenarios and brief clustering reviews are presented in these works, however there is no study in literature that extensively surveys clustering challenge for CoMP. In [5], CoMP clustering is reviewed briefly and a subset of static overlapping clusters are presented, however this work lacks a comprehensive survey on all clustering models in literature, especially missing the advanced clustering techniques, i.e., dynamic and/or multi-objective based clustering. CoMP concept and trial results are presented in [8] with a dynamic clustering algorithm trialled in a test network, however the paper again lacks a review of other available clustering models. Lee *et al.* [9] discuss CoMP implementation challanges and various deployment scenarios for LTE-A, however clustering challenge is not exploited in the paper. Backhaul capacity and latency requirement for different CoMP schemes are investigated in [12]. A user-centric CoMP clustering approach is studied to investigate available backhaul capacity/latency impact on CoMP clustering. Wireless cluster feasibility is presented for different cluster size and backhaul capacity. However the paper lacks on an extensive review of other available CoMP clustering algorithms which can be employed to dynamically adapt to available backhaul capacity. Beylerian and Ohtsuki [13] presents a service-aware resource allocation for non-coherent joint transmission (JT) CoMP in cloud radio access networks (C-RAN) architecture where a static and a user-centric clustering approach is presented. Beylerian and Ohtsuki [14] propose a further resource allocation solution combining

non-orthogonal multiple access (NOMA) scheme with CoMP to exploit power and space domain multiplexing and further improve capacity. A static clustering of a fixed cluster size of 2 is employed in this work, however both studies does not intend to cover all clustering solutions available, especially missing the dynamic clustering algorithms which can reduce high complexity on user-centric clustering solution in large clusters of cells. Rao and Fapojuwo [15] presents a survey on energy efficient resource management for cooperative networks however energy efficient cooperative clustering challenge is not reviewed extensively. A comprehensive book is published about CoMP [10], two example clustering techniques, one for static, one for dynamic clustering is presented however it again fails to present an extensive review for CoMP clustering. Coalitional game theory is introduced in [16] as an important analytical tool to form CoMP clusters. An example clustering algorithm is also presented for user equipment (UE) clustering in the uplink, maximising the sum-rate capacity. Nonetheless, the book fails to provide a review of all CoMP clustering approaches available. An extensive survey is provided on control and data plane separation architecture (CDSA) for future networks in [17], however this survey lacks a review on CoMP within the CDSA architecture. Mustafa *et al.* [18] provides a survey on device to device (D2D) CoMP within the CDSA architecture and discuss CoMP clustering briefly with one dynamic clustering example. Both papers [17] and [18] lack a wider review of all CoMP clustering solutions available in literature. In [19], an extensive review for self organising networks (SON) is provided, however CoMP clustering is not discussed in relation to SON framework. To the best of our knowledge, there is no comprehensive survey in the literature about CoMP clustering. This paper aims to fill this gap, providing an extensive survey on the existing CoMP clustering approaches in literature. Two novel taxonomies on CoMP clustering based on aimed objective and self organisation are presented. Strengths and weaknesses of available solutions are critically reviewed and future research directions are identified.

III. CoMP - Essential Background

In this section, we provide an essential background of CoMP to the reader before moving to the main scope of this article, i.e., CoMP clustering.

Network coordination deals with inter-cell interference, reducing the interference especially at the cell edge, resulting in much needed additional capacity and increased UE throughput. By making use of the shared data between coordinating transmission points (CSI/scheduling/user data etc), inter-cell interference can be mitigated or even exploited as meaningful signal at the receiver. Transmission points (TP) are different antenna ports of MIMO enabled cells which may or may not be located at the same place.

CoMP is one of the key features, standardized for LTE-Advanced to uplift the network performance. 3GPP initiated a study item on LTE-Advanced in March 2008 and the requirements for radio interface enhancements are published in [20]. To satisfy these requirements, 3GPP published the physical layer enhancements in [21] where CoMP has been identified as

(a) Joint Transmission (JT)

(b) Dynamic Point Selection (DPS)
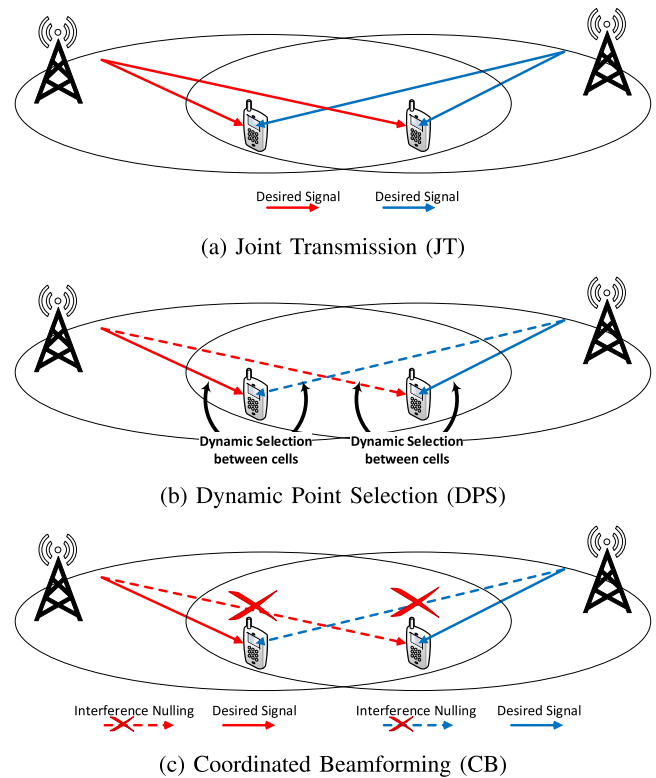

(c) Coordinated Beamforming (CB)

Fig. 2. Main Downlink CoMP Types for LTE-A [7].

one of the key features. A further feasibility study for CoMP in LTE-A is undertaken by 3GPP in Release 11 [7], where physical layer aspects of CoMP is studied. Simulation results from various sources are presented in this study where it is shown that CoMP can offer a significant performance improvement especially at the cell edge for different network deployment scenarios [7].

Lee *et al.* [9] show that more CoMP gains are achievable for cell edge users in scenarios where more interference is experienced. Similarly, more CoMP gains are presented for HetNet scenario where pico cells experience severe interference from macro sites.

Various levels of coordination schemes are studied in [10] but three main downlink coordination categories are identified by 3GPP for LTE-Advanced [7] based on the required backhaul capacity and scheduling complexity. An illustration of downlink CoMP types is given in Figure 2.

1) Joint Transmission (JT): CSI/Scheduling information and also user data is shared between the coordinated TPs. This type of coordination offers better results, however it requires high backhaul bandwidth with low latency due to user data exchange between multiple TPs. Multiple TPs can serve to single user either coherently or non-coherently, converting interference signal to useful signal. Coherent transmission refers to joint precoding design and synchronised transmission to achieve coherent combining. Non-coherent JT does not require joint precoding, user data is received from multiple TPs where data is individually precoded from each cell.

2) Dynamic Point Selection (DPS): This is a special type of JT where user data is transmitted from one TP only and serving TP is changed dynamically in each subframe based on resource availability and channel conditions. Fading conditions are exploited to select the best serving cell at each subframe. User data is available at multiple TPs similar to JT.

3) Coordinated Scheduling/Beamforming (CS/CB): CSI is shared but user data is not shared among the cooperated TPs so user data is only available at one TP but scheduling and beamforming design is coordinated between the TPs. Beamforming vectors are selected such that interfering TPs is steered towards the null space of the interfered user to minimise interference. CS/CB require lower backhaul bandwidth when compared to JT due to reduced data exchange.

There are two main uplink CoMP transmission categories identified by 3GPP in [7].

1) Coordinated Scheduling/Beamforming (CS/CB): User scheduling and precoding design is done by coordination between the TPs however user data is only received by one TP.

2) Joint Reception: User data is received by multiple TPs jointly. Similar to downlink JT, uplink joint reception offers higher gains but with the cost of increased complexity and higher backhaul bandwidth requirement.

A. Enabling Technologies for CoMP

The requirement for network densification for future cellular networks has initiated research on a number of new network architectures to optimise increased energy consumption, signalling and complex mobility management etc. These recently emerging radio access network (RAN) architectures will also help to overcome the challenges for CoMP (i.e., backhaul limitation, complex precoding, signalling etc), enabling CoMP to be one of the main features of future wireless networks.

- *Control/Data Plane Separation Architecture (CDSA):* Motivated by proposed dense HetNet deployment and energy efficiency concerns, a control and data plane separation architecture (CDSA) is proposed for macro BSs to provide coverage layer and handle most of the control signalling and small cell (SC) layer under the macro BS to provide the required data services. Reader is referred to [17] and [18] for two recent extensive surveys for further reading on CDSA. CDSA is one of key enablers of CoMP implementation where macro BSs can be enhanced to function as CoMP control unit (CCU) with strong backhaul links to the SCs within its coverage area. CCU functionality on the macro cell can handle central precoding design, baseband processing and can make intelligent clustering decisions centrally within the SC layer, taking various efficiency metrics into account, i.e., energy efficiency, load balancing, spectral efficiency etc. With all SCs connected to the associated macro BS, there is no need for high bandwith backhaul between the small cells in CDSA.

- *Cloud Radio Access Networks (C-RAN):* Another architecture envisioned for network densification is C-RAN where baseband processing unit (BBU) is decoupled from remote radio unit (RRU). A pool BBU is proposed in the cloud where there is high bandwith front-haul between the cloud and RRUs [22]–[24]. Baseband resource sharing can be maximised and CoMP can easily be realised in this architecture. Cloud can be enhanced to handle CCU function and make intelligent clustering decisions for the connected RRUs. A BBU+RRU based CoMP example has been studied in [25] for LTE-A giving promising spectral efficiency gains as expected. The downside of C-RAN is the requirement for high bandwidth fronthaul. Larger CoMP cluster size in C-RAN can be feasible with ideal fronthaul [26] due to centralised BBUs handling main CoMP functions. Concept of self organising cloud cells is proposed in [27] where SCs within the coverage area of a macro BS are connected to the macro BS. Macro BS then handles the decision making on which SCs to be allocated for user data service to improve blocking probability, energy consumption and handover probability. This setup can also be easily extended to enable CoMP and enhance macro BS to handle CoMP-CCU functionality.

IV. CoMP Clustering and SON

In this section, we first discuss the key challenges in CoMP clustering design and identify the need for dynamic CoMP clustering for maximising CoMP gains by adapting CoMP clustering to changing network and user profile conditions. We then propose SON as the key enabler for dynamic CoMP clustering and give brief introduction on SON.

A. CoMP Clustering Challanges and SON

As discussed earlier, CoMP can only be realised within small cluster of cells due to its complexity which generally increases with the number of coordinating cells. Optimum cooperating cluster selection is key for maximising the benefits of CoMP. An illustration of CoMP clustering in a typical CDSA architecture is provided in Figure 3.

A number of challenges need to be critically evaluated for a comprehensive CoMP clustering approach to maximise the benefits of CoMP:

- *Is it efficient to deploy CoMP?* The first question which need to be answered is, if it's worth deploying CoMP for individual cells in a given network setup. Would the overheads for deploying CoMP be more than the gains it provides? As illustrated in Figure 3, cells closer to each other need to form clusters for cooperation as the CoMP gains would be maximised when there is severe inter-cell interference which can be mitigated. However, isolated cells may need to work without coordination, based on the limited amount of inter-cell interference experienced from other cells. In addition, users close to the cell center may not experience high inter-cell interference, however cell edge users will suffer from high
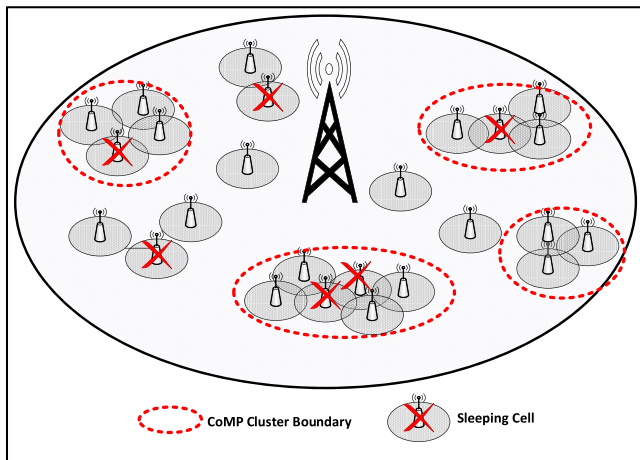
Fig. 3. Dynamic Multi-Objective CoMP Clustering illustration in CDSA Architecture.

interference hence, it can be more efficient to deploy CoMP for cell edge users only. Garcia *et al.* [28] presented a dynamic clustering scheme and suggested no spectral efficiency gain in employing CoMP in high signal-to-interference-plus-noise ratio (SINR) region due to additional pilot signalling required for CoMP, reducing spectral efficiency more than the expected gains. Users are allocated CoMP clusters or CoMP is not used based on their SINR from the local serving BS. It's shown that CoMP gains are maximised when received power levels from coordinating cells are close to the received power levels of the local serving cell. Hence it can be concluded that CoMP gains vary with network density and CoMP may not need to be deployed for some cells based on their location, user profile and the amount inter-cell interference.

- *How many cells in the cluster?* Cluster size is another key parameter for optimal CoMP clustering. Too small clusters will fail to provide full achievable gains from CoMP, on the other hand, big cluster size will lead to increased overhead on CSI feedback and backhaul capacity [29]. Increased cluster size will give better weighted sum rate [30] but with the cost of additional signal processing and increased feedback and signalling. Moreover, increased cluster size can lead to energy inefficiency in terms of achieved bits/joule [31]. As illustrated in Figure 3 for an example CDSA architecture, some clusters will have 6 cells, others will have 5 or 4 and some others will reduce cluster size by switching off some cells within the cluster for energy efficiency. Hence, there is no ideal fixed cluster size, instead, cluster size needs to be a dynamic parameter in the clustering algorithm which needs to change based on channel conditions and user profile.
- *Which cells to switch off for energy efficiency?* As illustrated in Figure 3, some cells can be switched off by forming intelligent CoMP clusters to enhance SINR and make sure minimum SINR is provided while some cells are switched off for energy efficiency. A number of

network objectives will need to be considered for BS switch-off:
  - Can the remaining capacity in the cooperating cluster cope with the traffic demand for a given quality of service (QoS)?
  - Is SINR provided by the cooperating cluster without the sleeping cell over the minimum threshold?
  - Do the cells within the cooperated set have enough backhaul bandwidth to cope with increased traffic when a cell is switched off for energy efficiency?
- *Load Balancing / RAN Capacity/ Backhaul bandwidth:* Cooperation introduce additional capacity in the network by improving spectral efficiency [8]. Intelligent clustering algorithms can be employed to support load balancing by shifting traffic from highly loaded cells to its neighbouring clusters. Increased cluster size can also uplift capacity in hotspot areas based on network topology. However, backhaul bandwidth requirement will also increase with increased cluster size. Hence multiple objectives need to be considered for intelligent CoMP clustering.

Given the challenges for CoMP clustering design as discussed above, static clustering based on a fixed topology will fail to give expected gains for future networks as the network topology will be dynamically changing with on/off sleeping cells, user deployed cells with unknown location etc. Moreover, spatio-temporal distribution of users and service demands dynamically changes. To maximise CoMP gains, clustering algorithms need to be able to accurately respond to these dynamically changing network conditions and user profiles. Self organised CoMP clustering algorithms can be developed to make optimum clustering decisions by reading various network data and making clustering decisions based on the changing conditions, maximising the objectives like spectral efficiency, energy efficiency, load balancing while keeping the fairness between the users.

Dynamic clustering can be implemented in the SON platform which employs autonomous closed-loop changes in the network dynamically. Big Data available from various sources within the cellular network can be exploited as an input for SON platform for proactive CoMP clustering algorithms and other SON functionalities. Accurate prediction of user profiles and mobility based on Big Data can be employed within the SON platform for much needed lower latency on CoMP clustering design. Use of Big Data for proactive CoMP clustering is further discussed in Section VII-A. A brief background for SON is given in the next subsection.

B. Self Organising Networks (SON)

SON is an emerging concept in wireless cellular networks to automate some of the operational tasks in closed loop to overcome the challenges of a complex multi-layer network. Network conditions are monitored dynamically by exploiting Big Data from various sources and intelligent algorithms are employed to effectively manage the network based on the changing local conditions. Dynamic CoMP clustering can also deployed within the SON platform as an enhancement to other SON modules which utilises the Big Data for making

proactive CoMP clustering decisions. SON algorithms can be designed as a distributed or centralised function depending on the requirements of the tasks, especially time and scalability limitations. Given the increasing complexity of the wireless cellular networks, SON will have a strong, enhanced presence in future networks. Future networks will need to deploy effective SON algorithms to improve capacity and QoS and reduce capital expenditure (CAPEX) and operational expenditure (OPEX) by reducing labour costs.

SON has been an important part of 3GPP LTE/LTE-Advanced standardization which has started with Release 8 and enhanced further with most recent Release 12 [32]. An extensive survey on SON has been presented in [19].

SON is mainly categorised in three folds:

1) **Self Configuration:** This group of SON modules aim to manage new entities integrated in the network. A considerable amount of OPEX cost is spent for new site configuration during network rollout and it will increase with proposed massive deployment of small cells. Self configuration algorithms aim to automate new site configuration, initial automated neighbour relations and software updates [33].

2) **Self Optimisation:** This group of SON modules aim to optimise ongoing services in the network. Self optimisation algorithms will monitor network performance data and derive optimisation changes in the network in open and/or closed loop, aiming to reduce OPEX costs and also improve network spectral efficiency, energy efficiency, network capacity and overall QoS. Dynamic CoMP clustering can be incorporated to Self Optimisation module set and implement closed-loop dynamic clustering decisions based on network data already available in the SON platform. Self optimisation is an important part of LTE/LTE-Advanced standardisation [34] and there are already commercialised algorithms deployed in the current LTE networks. Self optimisation tasks can be mainly grouped in three folds [19].
   a) Load balancing
   b) Coverage and Capacity Improvement
   c) Interference Control

3) **Self Healing:** This group of algorithms aim to detect faults in network elements, analyse the fault by gathering relevant information, diagnose and clear the fault. For time consuming fault restoration, self healing also aims to perform compensation actions on neighbour cells until the faulty cell is restored. 3GPP has standardised self healing for LTE/LTE-Advanced as an important feature of SON platform [35].

V. CLUSTERING TAXONOMY BASED ON SELF ORGANISATION

In this section, CoMP clustering algorithms in literature are critically discussed based on self organisation. Three main clustering types are identified:

1) Static Clustering
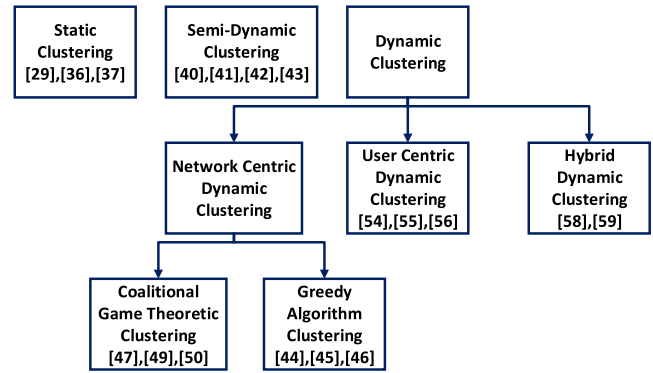2) Semi-Dynamic Clustering
3) Dynamic Clustering

Fig. 4.　CoMP Clustering Taxonomy based on Self Organisation.

A summary of clustering taxonomy based on self organisation is given in Figure 4.

Static clustering method is less complex with less signalling overhead but this method is not responsive to changes in the network nodes or user locations, hence the performance gains are limited. Semi-dynamic clustering is an enhanced version of static clustering where a number of static clusters are formed and employed dynamically to improve the potential gains. Complexity increases with additional signalling but performance is also improved when compared to static clustering. However, this method still lacks on truly responding to the dynamic changes in the network. Dynamic clustering methods are developed to respond to network and user mobility changes, i.e., new sites, sleeping cells, load changes etc. This scheme comes with increased complexity on scheduling and beamforming design but it gives the best results, reducing inter-cluster interference by moving the clusters dynamically. Dynamic clustering can be classified in three main categories within itself based on the approach. In network-centric clustering approach, all users in the same cluster use the same set of cells, however in user-centric clustering, users can be assigned their own clusters which comes with additional complexity. Hybrid approach combines both approaches which can be a good balance of complexity vs. performance.

In the subsequent subsections, we present an extensive literature review for each category and criticise available techniques based on complexity, scalability and potential spectral efficiency gains.

A. Static Clustering

CoMP coordination clusters are formed in a static way, mostly based on topology and don't change according to changes in the network. This method offers a less complex solution which can be a good candidate to deploy in the initial phase of LTE-A deployment. Static clustering within cells in the co-located site is the most basic and practical option which does not require data exchange between the sites, hence not reliant on fast backhaul.

The work presented in [36] propose a static clustering scheme, where sectors looking into each other are clustered to improve SINR. Authors assume a hexagonal grid in deployment which is non-realistic in real network deployments. This

is usually the downside for most static clustering solutions. In [37], static intra-site and inter-site CoMP clustering is considered with orthogonal frequency reuse where antenna bore-sights are shifted to face into each other for extra CoMP gain. Dead-spots would be created with this new topology where small cells are proposed to fill in the dead-spots. CoMP and HetNet deployment are merged in this solution to identify locations for small cell deployment, however an idealistic hexagonal grid is assumed again, which is unrealistic. A disjoint and overlapped static clustering model is presented in [29] where static clusters are formed to maximise mean SINR or to minimise SINR outage at possible user locations. In the overlapped solution, one cell can be in three clusters where system resources are splitted into each of the three clusters. Presented solution is better than the clustering types based on regular patterns as it can apply to realistic network topology. However the proposed work is not scalable as the complexity of the solution increases with the number of possible user locations.

A number of drawbacks for CoMP clustering have been investigated in [38]. Authors have investigated an inter-cell interference model in HetNet scenario with pico-cells to offload macro network. Time-domain resource partitioning is considered between the macro BS and pico layer within the macro BS's coverage area. A static CB-CoMP method is applied with centralised beamforming and scheduling for the cluster of all pico-cells and its connected macro cell. CoMP failed to improve the performance further from enhanced inter-cell interference coordination (eICIC) due to the additional overhead required to implement CoMP, i.e., mainly the UE-RS signal introduced with CoMP in LTE-Advanced. In [39], time synchronisation limitation between coordinated cells is investigated. Authors have shown that time synchronisation will need to be taken into account for a network with large inter-site distance (7km studied), however there is minimal inter-symbol interference (ISI) issues for inter-site distance of < 1 km due to cyclic prefix (CP) length.

In summary, static clustering is an attractive approach with its significantly less complexity for initial CoMP deployment for LTE-A networks. Intra-BS CoMP is a promising solution which eliminates the need for high backhaul bandwidth requirement between the BSs. On the other hand, inter-BS static clustering algorithms are mostly based on the assumption of hexagonal grid layout, which is not applicable to real networks. Furthermore, this method will fail to give the much needed spectral efficiency gains and increased system capacity for future 5G networks. Semi-dynamic and/or fully dynamic solutions are required to respond to changing network/user profile conditions and maximise CoMP gains.

B. Semi-Dynamic Clustering

Semi-dynamic clusters are more advanced than static clusters where several layers of static clusters are designed to avoid inter-cluster interference. More than one static clustering patterns are formed where users are able to select the most suitable cluster. This method also mostly relies on hexagonal grid network topology which is unrealistic in practical networks.

A two layer static clustering, based on regular network topology is proposed in [40] to extend on static clustering. This approach is then extended for several layers for dynamic clustering. It's proposed for users to pick one of the available clusters based on power. While the solution is an improved algorithm compared to static clustering, overlapping nature of the proposed algorithm adds to the scheduling complexity and require increased backhaul bandwidth. A semi-dynamic clustering scheme is introduced in [41] where static clusters are formed based on hexagonal grid topology and multiple shifted cluster patterns are created with different sub-channels allocated for each shifted cluster. A joint, centralised scheduling is developed for this clustering type. In [42], static cluster shift idea from [41] is further enhanced with "full shift" and different frequency bands are allocated on shifted clusters. Static clusters are formed to maximise neighbouring cells in the same cluster for a given hexagonal network layout. Shifted clusters reduce the inter-cluster interference, maximising the CoMP gain, however solution is based on hexagonal grid topology which is not applicable to real networks. In [43], a semi-dynamic clustering scheme is proposed for downlink Time Division Duplex (TDD) JT-CoMP scenario. Solution is based on large size (nine cells) static clustering and creating different static patterns of sub-clusters in each large static cluster. Dynamically selecting sub-clusters achieves almost as good as large cluster spectral efficiency but with reduced complexity. Proposed method is not able to respond to dynamic changes within the static cluster, i.e., new/sleeping cells etc. and also static nature of the big clusters will create inter-cluster interference.

In summary, semi-dynamic clusters are an improved version of static clusters with minimal overhead increase, however most solutions are based on idealistic hexagonal grid topology which is not realistic. Furthermore, majority of semi-dynamic algorithms propose orthogonal frequency allocation from each cell to its assigned static clusters. Based on the utilisation of dedicated bandwidth for each static cluster, proposed algorithms can reduce the overall spectral efficiency. Moreover, static nature of clusters is not able to respond fully to the spatio-temporal changes in user profiles and the network elements. Dynamic clustering algorithms is discussed in the next section which is mostly applicable to real network topology and can dynamically adopt to changing user profile and network conditions.

C. Dynamic Clustering

Dynamic CoMP clustering is more complex with increased signalling overhead but its more responsive to the changes in the network. Inter-cluster interference can be minimised and cluster size for individual users can be optimised dynamically for an optimum balance. Dynamic CoMP clustering can be classified in three groups based on network elements considered for clustering:
1) Network-Centric Clustering
2) User-Centric Clustering
3) Hybrid Clustering

(a) User-Centric Clustering

(b) Network-Centric Clustering
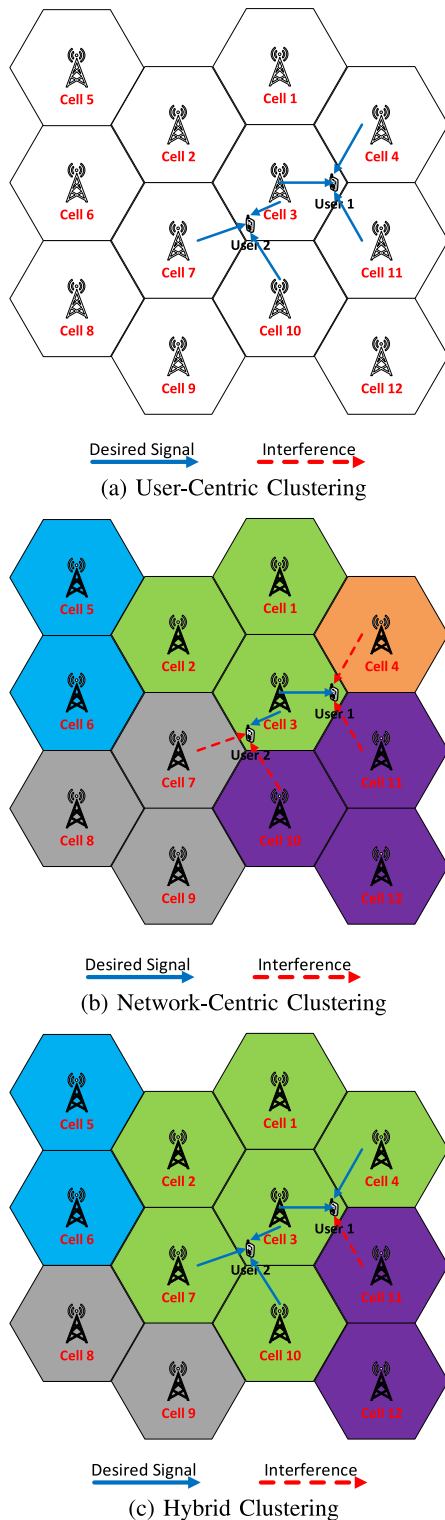

(c) Hybrid Clustering

Fig. 5. Dynamic CoMP Clustering Taxonomy.

An illustration of the three types of dynamic clustering is given in Figure 5. CoMP benefits are illustrated for two sample users for an identical network with different clustering schemes. For example, user-1 in the figure is located at the edge of cell-3, receiving strong interference from cell-4 and cell-11. Network-centric clustering is the most limited scenario where user-1 is located at the edge of the cluster. Its cluster

consists of cell-3 only and there is interference from cell-4 and cell-11. Hybrid clustering employs larger network-centric clusters, which improves user-1's cluster to cell-3 and cell-4. User-1's SINR is improved in this clustering type but there is still interference from cell-11. The most beneficial clustering scheme is the user-centric one where user-1's cluster consists of all three surrounding cells, i.e., cell-3, cell-4 and cell-11. Although user-centric clustering seem to be most beneficial one, it comes with additional scheduling/precoding complexity and increased backhaul requirement. The three types of dynamic clustering are reviewed in detail in the subsequent subsections.

1) Network-Centric Clustering: In network-centric clustering approach, cells are clustered in groups where all users within the serving area of the clustered cells are served by all cells or a sub-group of cells in the cluster. A simple illustration of network-centric clustering is given in Figure 5b. It is less complex when compared to user-centric clustering, especially from scheduling point of view. However cluster edge users suffer from inter-cluster interference. Dynamic network-centric clustering can minimise this effect by moving cluster boundary dynamically.

Two main methodologies are identified in the literature on dynamic network-centric clustering.

a) Greedy algorithms: Greedy algorithms are widely used for cooperation cluster formation in literature. Clusters are formed iteratively, starting from a randomly chosen BS to maximise the main objective, typically spectral efficiency. Best cluster is formed for the randomly chosen BS, maximising the CoMP gains, however the clusters formed in later stages of the algorithm suffer from sub-optimal clusters. It is relatively less complex but may not achieve as good results as the other methods, i.e., game theoretic clusters. A greedy uplink clustering algorithm is studied in [44] aiming to maximise spectral efficiency. It's shown that dynamic clustering with cluster size of two cells outperforms static clustering with much larger cluster size. A predefined fixed cluster size is proposed which is not the optimal solution for some clusters. A similar approach is employed in [45] but a dynamic cluster size is proposed. Authors have designed a dynamic clustering solution for uplink multi-user distributed antenna system (MU-DAS), where one cell has a number of RRUs placed in the cell's coverage area with fast fiber connection to their cell. BSs are merged based on highest interference created to the other users. However, clustering takes only scheduled users into account at any point in time, hence not taking load into account for cluster formation. Starting the iterations from the highly loaded cells can improve the system throughput as the CoMP gains will be maximised for clusters formed in early stages of the algorithm. Also clustering is proposed to change with each scheduling interval in this solution [45] which increases signalling due to high frequency cluster changes. Both proposed algorithms in [44] and [45] offer disjoint clusters where inter-cluster interference is still an important factor reducing spectral efficiency. An overlapping dynamic clustering is proposed in [46] to improve network average sum rate and fairness. A greedy approach is considered starting from a random BS. Authors have shown better

results with cluster size of four with overlap size of two when compared to cluster size of eight with no overlap. The solution lacks scalability where large network size can lead to increased complexity. Overlapping clusters will also require more complex scheduling but overlap and cluster size parameters are introduced in the proposed algorithm to control this complexity.

Greedy algorithms provide lower computational complexity however lack on sub-optimal clusters especially for clusters formed at later stages of the algorithm. Shortcomings of greedy algorithm can be improved by employing coalitional game theory for cluster formation based on merge-split rule for maximising system throughput. Game theory can also provide distributed solutions with reduced signalling overhead as opposed to centralised greedy algorithms, however coalitional game theoretic algorithm's computational complexity is higher than greedy algorithms [47]. Coalitional game theoretic clustering is discussed in detail in the next paragraph.

b) Game theoretic clustering: There is an increasing interest in applying coalitional game theory to design self-organised, distributed cooperative clusters. A utility function is introduced to formulate the cost and CoMP gain trade-off for forming clusters. Proposed utility function can limit the cluster size dynamically based on BS locations and user profiles. Coalitional game theory can provide distributed, stable, converging solutions to maximise CoMP gains. An extensive tutorial on coalitional game theory for wireless communications applications is presented in [48].

Li *et al.* [49] proposed a dynamic network-centric clustering method employing a utility function to maximise the second best servers of the cell edge users in the same cluster. Cluster size is fixed to two only which leads to sub-optimal clustering for varying network conditions. Also network clustering formation is based on exhaustive search for collusion, hence not scalable, i.e., complexity increases with network size. Moon and Cho [50] have studied a dynamic cluster formation algorithm which merges cells into clusters based on the improvement on spectral efficiency, with configurable maximum cluster size and the minimum efficiency gain. This algorithm is semi-distributed where SINR measurements are based on pilot signal measurements but still need a CCU for cluster decision-making. It implicitly takes the number of users into account and hence clusters are formed based on cell load. Although a more flexible cluster size is introduced in [50] when compared to [49], algorithm still lacks on scalability as the complexity increases with the number of BSs involved. Saad *et al.* [51] presented an application of a coalitional formation game for user clustering in the uplink, maximising the sum-rate capacity with a cost function based on power requirements which is dependant on the distance between the users. Inspired by [51], Guidolin *et al.* [47] developed a coalitional game theoretic clustering method where utility function dictates average cluster size and targets for higher spectral efficiency. It's a distributed algorithm which does not need a central entity and reduces signalling overhead. SINR at the cell edge is also significantly improved when compared to a greedy algorithm. On the other hand, solution lacks on scalability where the cluster formation complexity increase with network size. Computational complexity of such algorithms can be reduced by limiting the candidate sites for coalition to neighbour cells only. Utility function for forming coalitional game theoretic clusters play an important role for optimal clusters. Utility function need to include a realistic model for the cost of cluster formation and the relevant CoMP gains. Dynamic cluster size can be self-imposed with accurate implementation of a utility function. Also multi-objective clustering can be implemented by including multiple metrics into the utility function, i.e., energy efficiency, load balancing, spectral efficiency and backhaul bandwidth limitations.

c) Other dynamic network-centric clustering algorithms: A self organising dynamic clustering method is presented in [52] where candidate clusters are formed from reported list of cells from users. CCU is proposed to arrange cluster solution by listing the candidate clusters with minimal cost, where the cost function takes into account the cluster size, number of users and reference signal received power (RSRP). This algorithm is a basic one where cost function can be improved to maximise SINR / spectral efficiency for more optimal solutions. It lacks on scalability with increasing complexity of handling high number of candidate clusters as the network size / number of users increase. Time averaged measurements from users is considered where fast fading is eliminated. Weber's algorithm [52] is further enhanced in [53] by replacing the cost function based on received power levels to a utility function with the aim of maximising the weighted sum rate. Unlike [52], Baracca *et al.* [53] proposed a fast changing cluster design, responding to fast fading channel variations which will lead to increased signalling and possible ping-pong cluster re-selections. To reduce signalling overhead, cluster change frequency can be reduced to a wider time-frame and averaging algorithm can be used for user measurements which can eliminate fast fading variations.

2) User-Centric Clustering: Users are allocated their own cluster of cells individually in user-centric clustering approach. Although this method can give better SINR/throughput gains, it's more complex, especially in terms of scheduling where user clusters overlap with each other. To reduce complexity, user-centric clustering can be implemented in small groups of cells rather than the whole network.

Garcia *et al.* [54] have studied macro diversity CoMP with dynamic user-centric clustering, comparing random network and hexagonal network topologies. It's shown that CoMP gives higher capacity results and bigger cluster size are required in random networks due to the random nature of BSs with more potential for inter-cell interference. Authors had no limitation on user-centric clustering which leads to complex scheduling between the BSs. To reduce complexity, user-centric clustering can be limited to groups of cells for easier scheduling, less signalling overhead and data exchange.

A three-tier clustering approach is presented in [55], wherein it has been proposed that cell center users will not use CoMP, users within the same site will use static clustering between intra-BS cells and a user-centric clustering is proposed for intra-site cluster edge users. Fixed cluster size is assumed which can lead to unnecessary complexity or less efficient coordination, depending on user location and

SINR conditions. Similar complexity arises in works presented in [54] and [55] where no limitation is proposed on user-centric approach to any group of cells which will lead to higher complexity with a large number of BSs cooperating at the same time. In [56], a user-centric inter-cell interference nulling is studied for downlink coordinated beamforming. Interference nulling range is derived from received power levels to form clusters for individual users. A threshold for relative power levels is used to identify BS clusters.

User-centric clustering approach is an ideal scenario to provide an upper limit however it is not realistic due to increased complexity. Hybrid clustering is discussed next, which limits the user-centric approach to a group of BSs only to reduce complexity.

3) Hybrid Clustering: Hybrid clustering approach is the combination of network and user-centric approaches where users are allocated their own preferred cells but limited to a bigger group of cells which can be dynamically changing to adapt to changing network conditions. Hybrid clustering is driven from the complexity/throughput gain trade-off where user-centric clustering is used for better throughout but its complexity is kept at manageable levels by introducing network-centric clustering where users are limited to select cells only within the network-centric cluster.

Zhao *et al.* [57] developed a hybrid clustering method where a pre-defined network-centric clustering is used for cell center users and a number of pre-defined overlapping clusters are used for cell edge users to pick the best overlapping cluster to maximise SINR for the cell edge user. Inter-cluster interference on overlapping clusters is eliminated by orthogonal frequency allocation. Presented solution lacks on self organisation as the pre-defined clusters are static, i.e., can't respond changes in the network (new sites, sleeping cells etc). Although overlapping cluster patterns improve cluster edge user performance, orthogonal frequency use prevents the optimal use of the bandwidth. A simple downlink user-centric clustering is studied in [58] where users coordinate with two best serving cells according to the received power levels under a bigger static cluster. Proposed static network-centric clusters will suffer from high inter-cluster interference and also fixed user-centric cluster size can lead to unnecessary coordination, waste of resources and also possibly not being able to cancel severe interference from third best server for some users. A self-organised, dynamic network-centric clustering can improve inter-cluster interference and also dynamic user-centric cluster size can be employed for better performance. In [59], a hybrid clustering model for downlink SU-COMP is studied. Authors proposed static network-centric clusters and cell edge users are proposed to have user-centric clusters of fixed size of three within each network-centric cluster. Authors also presented a good review of SU-COMP scheduling and a SU-COMP joint scheduling algorithm is provided for the proposed clustering scheme. The presented clustering scheme has low complexity, but further work is required to introduce dynamic network-centric clustering for improved cluster design. Fixed cluster size is also another shortcoming of the algorithm which can generate sub-optimal clusters.

In summary, dynamic CoMP clustering is a promising concept which can improve performance over static/semi-static alternatives. However, increased complexity and performance trade-off need to be evaluated for optimal solutions. User-centric clustering provides a theoretical upper bound for maximum performance gain but it requires complex precoding design, scheduling and increased backhaul bandwidth [54], [55]. To reduce complexity, user-centric clustering solutions need to be limited to smaller network-centric clusters. Main approaches in network-centric clustering in literature are greedy algorithms studied in [44]–[46] and more recently coalitional game theoretic approaches deployed in [47], [49], and [50]. The key balance between additional complexity and the potential CoMP gains can be achieved by hybrid solutions where user-centric clustering is deployed within network-centric clusters [57]–[59]. However, hybrid solutions in current literature focuses either on dynamic user-centric approach with static network-centric clustering or dynamic network-centric clustering with no focus on user-centric clustering. Further research is required to employ dynamic clustering algorithms for both network-centric and user-centric clusters for more optimal solutions. A summary of CoMP clustering approaches based on self-organisation and their shortcomings are provided in Table II.

VI. Dynamic Clustering Taxonomy Based on Objective Function

In this section, a novel CoMP clustering taxonomy is presented based on the main objective function. The main objective of CoMP is to mitigate interference from neighbour cells and improve spectral efficiency in general but a more comprehensive approach is required to include other metrics/limitations for CoMP clustering. Backhaul bandwidth limitations for CoMP implementation and energy efficiency concerns for future wireless networks need to be included in comprehensive CoMP clustering algorithms. Moreover, with exponentially growing mobile data demand, better utilisation of system capacity with load balancing will be a key concept which need to be taken into account for CoMP cluster design. Based on our detailed literature survey, main objective functions studied are:

1) Spectral Efficiency
2) Backhaul Optimisation
3) Energy Efficiency
4) Load Balancing

A summary of CoMP clustering taxonomy based on objective function is given in Figure 6. In the following subsections, each objective function is critically discussed and extensive literature review is presented.

A. Spectral Efficiency

Main objective of CoMP is to mitigate inter-cell interference within the cooperating cluster. Interference cancellation leads to better SINR and improved spectral efficiency. Cluster formation algorithms are designed to maximise spectral efficiency as a common objective, however other utilities such as backhaul bandwidth optimisation, energy efficiency and load

| Clustering Type	Proposed Method	Shortcomings	Reference
Static clustering	Hexagonal grid topology, static clustering pattern	Non-realistic hexagonal grid approach, unable respond to network/user profile changes	[36], [37]
Static clustering	Mean/Outage SINR optimised, overlapping clusters	Increased complexity for bigger size network clustering	[29]
Semi-dynamic clustering	Multiple static clustering patterns	Non-realistic hexagonal grid approach, Increased scheduling/beamforming complexity, increased backhaul bandwidth requirement	[40], [41], [42], [43]
Network-centric dynamic clustering	Greedy algorithm	Sub-optimal clusters which are formed in later stages of the algorithm. Fixed cluster size	[44]
Network-centric dynamic clustering	Greedy algorithm	Sub-optimal clusters which are formed in later stages of the algorithm. Cell load not taken into account	[45],
Network-centric dynamic clustering	Greedy algorithm-overlapping clusters	Sub-optimal clusters which are formed in later stages of the algorithm. Lacking scalability	[46]
Network-centric dynamic clustering	Coalitional game theoretic clustering	Exhaustive search, higher complexity, fixed cluster size	[49]
Network-centric dynamic clustering	Coalitional game theoretic clustering	High computational complexity, not scalable	[50], [47]
User-centric dynamic clustering	User-centric design	Higher complexity for beamforming design and scheduling	[54], [55], [56]
Hybrid dynamic clustering	User-centric clustering within static network-centric clustering	Static network-centric cluster design, not able to respond to dynamic changes in the network/user/service profiles	[58], [59]
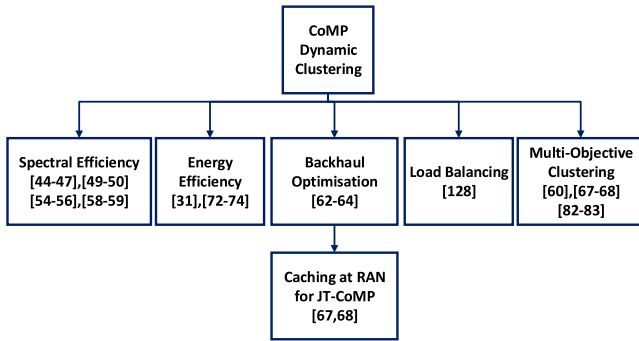
Fig. 6. CoMP Clustering Taxonomy based on Objective Function.

balancing have also been studied in the literature. Trade-off between spectral efficiency and other objectives for optimum clustering has been also in the interest of research community [60].

3GPP identified three CoMP deployment scenarios for LTE-Advanced and released a feasibility study, presenting simulation results from over 20 sources showing significant spectral efficiency improvements by deploying CoMP especially at the cell edge [7]. The most basic, intra-site static clustering is studied as scenario-1 and over 20% increase in spectral efficiency is observed at the cell edge with MU-MIMO JT-CoMP case [7]. Inter-site static clustering solutions are employed in [36] and [37] which is not able to respond to the dynamic changes in the network and user/service profiles, hence limiting the CoMP gains. Semi-dynamic clusters are proposed in [40]–[43] where multiple static clustering patterns are designed to mitigate inter-cluster interference. This type of approach is more responsive to the dynamic changes of the network and user profile, however it still lacks on providing full

spectral efficiency gain. Dynamic network-centric clustering methods can further increase spectral efficiency by dynamically changing CoMP clusters based on the spatio-temporal changes in user profiles and network elements. A game-theoretic, network-centric clustering approach is employed in [47]. Papadogiannis *et al.* [44] used a greedy clustering algorithm for uplink network-centric clustering to maximise spectral efficiency. User-centric dynamic clustering approaches are studied in [54] and [55] which provide an upper bound on spectral efficiency gain but with increased complexity. Hybrid solutions reduce this complexity where user-centric clustering is limited only within a network-centric cluster [57]–[59]. Dynamic clustering solutions require more complex precoding design and scheduling, and increased backhaul. Complexity and additional requirements are reduced in semi-static clustering, and further simplified in static clustering with the cost of reduced spectral efficiency gain. An extensive critical review of CoMP clustering solutions based on static/semi-static/dynamic approaches and the trade-off between complexity and the additional spectral efficiency gains are provided in Section V. A summary of different approaches and their shortcomings are presented in Table II.

B. Backhaul Optimisation & Caching at RAN for JT-CoMP

As discussed in previous sections, one of the key requirement of CoMP is high backhaul bandwidth and low latency. Depending on the type of CoMP, backhaul requirement will vary. JT-CoMP will require more bandwidth due to user data being shared between cooperated cells. Authors studied backhaul bandwidth requirements for network MIMO in [61] and concluded that backhaul requirement for CSI and scheduling information exchange is negligible when compared to

user data sharing. Hence, JT-CoMP require much larger back-haul bandwidth than CS/CB CoMP. Backhaul requirement is also strongly dependant on user SINR and cluster size. Users with high SINR will demand higher throughput which will increase backhaul demand. Biermann *et al.* [62] have studied distributed JT-COMP feasibility in terms of high backhaul bandwidth and latency requirements especially in hotspot scenarios where certain backhaul links are more loaded than others. In the proposed algorithm, all CSI is sent from cooperated cells to the serving cell where it's processed for precoding design and sent back from serving cell to the cooperating cells. Hence serving cell backhaul demand is more than other BSs in the cluster. Based on the backhaul load on each BS, a dynamic serving BS reassignment algorithm is proposed by using "forced handover" to distribute backhaul load evenly. Zhao and Lei [63] have designed a user-centric clustering strategy to minimise the backhaul data transfer for the JT-CoMP scenario where user data exchange between the BSs will be very high. An optimised number of links is proposed for a given CoMP cluster based on minimum SINR requirement of each user. Heuristic approach is used to reduce the links based on channel strength and "Signal to Leakage" (SLR) ratio (i.e., taking signal power and also the interference caused to other users into account). Authors have further improved this design in [64]. An optimisation problem is formulated and approximate results are obtained by convex relaxation. An iterative algorithm is followed to further reduce the number of BSs in each user's cluster. A control unit (CU) is proposed for the semi-distributed solution where each BS is connected and share CSI with CU. Author's approach of further user-centric clustering optimisation in a given network-centric cluster helps reducing wasted network resources. However a trade-off between spectral efficiency and backhaul bandwidth optimisation should be considered for more optimal solutions.

1) Caching at the RAN for JT-CoMP: There is an increasing interest in the research community to explore potential benefits of caching popular multimedia content closer to the user to reduce high backhaul requirement due to duplicate content download. A significant amount of network data usage is due to duplicate downloads of few popular multimedia content from Netflix, Youtube, Facebook etc. Caching the popular content at various points in the network, i.e., RAN, core network or even the user devices can reduce the high backhaul requirement and give opportunity for JT-CoMP deployment, where high backhaul capacity is not available [65]. Furthermore, caching closer to the user can improve overall energy efficiency. A recent study on an operational 4G network shows [66] that 73% of the data volume is cachable and 54% of the cachable data is revisited, so significant gains are possible with caching.

In [67], caching at the BS is proposed and an opportunistic cooperative MIMO is employed without high backhaul requirement. Cells within the same cluster are proposed to cache identical data aiming to be employed for multi-user JT-CoMP. For users where requested data is available at the cache, JT-CoMP is proposed. If the requested data is not available at the cache, CB-CoMP is proposed where user-data

exchange between the BSs is not required but CSI exchange is still employed for joint precoding. Deghel *et al.* [68] presented a JT-CoMP solution in a limited backhaul capacity scenario BS caching is introduced to reduce required backhaul capacity for user data, hence increasing available backhaul capacity for CSI sharing. Improved backhaul availability for CSI sharing improves the accuracy in CSI knowledge at the central node, resulting in better precoding, hence improved interference cancellation.

Realisation of CoMP depends on high backhaul bandwidth availability, hence CoMP clustering algorithms need to take this limitation into account. Caching popular multimedia proves to be one of the ways to reduce backhaul bandwidth requirement for CoMP realisation. Cluster size and type of cooperation are other factors that can change the backhaul bandwidth requirement. Furthermore re-distribution of backhaul data transfer to less-loaded cells can be deployed for better CoMP gains.

C. Energy Efficiency

Energy efficiency has recently become an important topic for wireless networks for both economical and environmental reasons [69], [70]. It has been reported that information and communications technology (ICT) industry contributes 2% of the global carbon footprint, and it's expected to increase to 3% by 2020 [71]. In mobile communications, more than 80% of the power is consumed in RAN, especially BSs [70]. As briefly discussed in the introduction section, network densification is one of the key tools to increase capacity for future wireless networks to meet ever increasing traffic demand which will severely increase energy consumption and OPEX costs. New architectures like CDSA [17] and C-RAN [22]–[24] have been envisioned to enable energy efficiency and reduce OPEX and CAPEX costs in future wireless networks, mainly by providing small cell coverage only when it's required. Enabling CoMP will also improve energy efficiency [15]. It's been in the attention of research to design CoMP clusters to maximise energy efficiency and to optimise the trade-off between spectral efficiency and energy efficiency. On one hand, CoMP can reduce cell/UE output power for a given QoS but there is also additional energy consumed for additional signal processing and backhaul requirement.

CoMP clustering can be optimised for energy efficiency by increasing the number of sleeping BSs and/or their sleeping duration. In [72], BS sleeping with CoMP has been studied for energy efficiency with static clustering and assuming one cell is sleeping on each cluster during off-peak hours. A joint sub-carrier and power allocation algorithm is proposed to minimise the power requirements for coordination and compensate for sleeping cell for a given QoS. Cao *et al.* [73] has compared the energy efficiency gains between CoMP and wireless relaying by maximising the number of sleeping cells. Based on the traffic demand, it's shown that, energy efficiency gains are almost constant in lightly loaded traffic conditions where network is mainly coverage limited. In high traffic load, there is almost no energy gains possible, whereas in "energy efficient region", dynamic energy efficiency algorithms can provide

bigger energy efficiency. As BS density increases, the "energy efficient region" also increases and the region for larger CoMP gains decreases. In [31], user-centric CoMP clustering for all cells within 3dB window is studied for cell switch-off on lightly loaded cells to improve energy efficiency. It's shown that unnecessary increase in cluster size and imperfect channel knowledge can lead to energy inefficiency. Up to 24% more energy efficiency is observed in perfect CSI when compared to imperfect CSI conditions. Authors proposed fast changing clusters responding to fast fading changes which increases signalling overhead and also imperfect CSI knowledge leads to non-optimal clusters. Cluster change frequency can be limited to respond to large scale fading only which can average out the fast scale changes for more reliable cluster formation.

Besides BS switch-off, deployment costs can be reduced and energy efficiency can be maximised by taking network coordination into account at network planning stage. In [74], a BS planning scheme is proposed to reduce the total number of required BSs for a given coverage and traffic quality of service (QoS) by inter-cell cooperation. A single user (SU) MIMO CoMP scheme with user-centric clustering method is employed to choose the optimal BS locations for deployment from a number of candidate BS locations to maximise energy efficiency. A typical example of this work is to reduce the number of BSs required from three to two BSs where some users can't be served without the third BS if CoMP is not employed.

Deployment of CoMP and realisation of future network architectures like CDSA and C-RAN will enable energy efficiency by increasing the number of sleeping cells. However, most studies in the literature are lacking the load conditions in the network but concentrate on coverage requirements only. With predicted mobile data growth, network capacity will be under pressure and will require to be managed more intelligently. BS switch-off with CoMP clustering algorithms will need to include data demand and available capacity in the network. Hence, a more comprehensive approach for dynamic CoMP clustering should optimise energy efficiency and load balancing jointly. We discuss CoMP clustering in relation to load balancing in the next subsection.

D. Load Balancing

Load balancing has always been an important topic for cellular networks due to non-even distribution of user traffic, resulting in some BSs overloaded whereas other BSs not fully utilised. Network planning process takes traffic distribution into account and BS locations are planned accordingly, however unpredictable nature of user activities like traffic accidents, mass events etc still cause overloaded cells. With ever increasing traffic, predicted 1000 fold increase beyond 2020 [2], load balancing becomes even more important in future cellular networks. Various load balancing schemes have already been studied in [75] for traditional networks. A mathematical framework for cell load and a simple load balancing algorithm is presented in [76]. Authors proposed to shift traffic from loaded cell to its unloaded neighbours by changing the handover offset parameter in iterations.

Imran *et al.* [77] presented a distributed self organised load balancing algorithm to reduce reference signal power for the congested cell to make neighbour cells more favourable and hence distribute the traffic onto neighbour cells. Another distributed SON algorithm in [78] focuses on BS antenna tilt optimisation to improve spectral efficiency at hotspots by finding the users centre of gravity and focusing the antenna beam to the hotspots. Kim *et al.* [79] presented a distributed load balancing solution from the idea of each BS periodically sharing its average load with users and users utilise this information alongside with signal quality to make the decision for cell association. A class of user association schemes for HetNet is presented in [80] to achieve load balancing between macro and small cell layer.

Centralised scheduling in emerging technologies like C-RAN [22]–[24] and CDSA [17] can also be utilised for load balancing. Centralised resource management entity (RME) is proposed for CDSA in [81] which will select the most suitable SC for scheduling. Centralised RME can also aim to distribute network load evenly between the SCs [17].

As discussed in previous sections, CoMP will introduce spectral efficiency gain and increased throughput especially at the cell edge [8]. Additional capacity from CoMP can be utilised for load balancing through dynamic CoMP clustering based on cell load. Centralised scheduling function within the CoMP cluster can be located possibly at the macro BS in CDSA, or at the pool BBU cloud in C-RAN. Self organised CoMP clustering algorithms can be developed to dynamically shift traffic from loaded cells to less-loaded neighbours while maintaining a certain level of QoS however there are no studies in the literature to our knowledge where CoMP clustering is used for load balancing.

E. Multi-Objective Clustering

As seen in aforementioned subsections, dynamic CoMP clustering can aim to improve not only spectral efficiency but also other objectives like energy efficiency and load balancing. Recent works on CoMP clustering have focused on multi-objective clustering where two objectives are optimised, trade-off between the objectives have been investigated. Katranaras *et al.* [60] have compared a number of static clustering options for trade-off between throughput and energy efficiency in sparse, medium and dense deployment scenarios. They have identified transmit power, inter-site distance and SINR service demands as the main inputs for this trade-off. Li *et al.* [82] proposed a dynamic CoMP clustering algorithm with BS sleeping to maximise energy efficiency while maintaining high achievable rate for all users. Candidate clusters are formed by all possible combinations of groups of cells with predefined cluster size and each BS selects a suitable cluster from the candidate clusters by maximising achievable rate for its users. Developed algorithm then looks for cell load and moves users from cells with low load onto other clusters to increase the number of sleeping cells and hence better energy efficiency. Proposed clustering algorithm lacks on scalability as number of candidate clusters increase with network size, leading to high computational complexity.

TABLE III
SUMMARY OF COMP CLUSTERING APPROACHES BASED ON OBJECTIVE FUNCTION

Clustering Objective	Proposed Approach	Impact / Shortcomings	Reference
Spectral Efficiency	Dynamic clustering as summarised in Table II	As summarised in Table II	[44], [45], [46], [49], [50], [47], [54], [55], [56], [58], [59]
Backhaul Bandwidth Optimisation	Dynamic serving BS reassignment	Re-distribute backhaul load from serving cell to cooperating cells	[62]
Backhaul Bandwidth Optimisation	Minimise cluster size based on min. SINR requirement	Reduced backhaul requirement, however spectral efficiency is sacrificed.	[63], [64]
Backhaul Bandwidth Optimisation	Caching at the BS. Switch between CB/JT CoMP based on backhaul availability	Reduced backhaul bandwidth requirement by caching popular multimedia at the BS	[67], [68]
Energy Efficiency	CoMP clustering to maximise BS switch-off	CoMP clustering to switch-off lightly loaded cells for better energy efficiency. Only coverage constraints are considered, network load constraints need to be jointly optimised	[72], [73], [31]
Energy Efficiency	Minimise number of BS deployment by employing CoMP	CoMP clustering to reduce the number of BSs required for deployment for better energy efficiency and cost saving. Only coverage constraints are considered, network load constraints need to be jointly optimised	[74]
Multi-Objective Clustering	Energy efficiency without BS switch-off and Spectral Efficiency jointly optimised	Energy efficiency by deploying CoMP without BS switch-off. Comparing different CoMP static clustering schemes for energy efficiency/spectral efficiency trade-off.	[60]
Multi-Objective Clustering	Energy efficiency with BS switch-off and Spectral Efficiency jointly optimised	Energy efficiency by deploying BS Switch-off with CoMP while maximising spectral efficiency	[82]
Multi-Objective Clustering	Backhaul Bandwidth and Spectral Efficiency jointly optimised	Effect of backhaul channel reliability on spectral efficiency for CB/JT CoMP. Reduced backhaul requirement by caching at the RAN.	[83], [67], [68]

Moreover, proposed algorithm fails to look at total system capacity and load balancing aspects, i.e., BS load need to be looked at for any congestion and reduce the number of sleeping BSs for much needed capacity in the network. Hence, energy efficiency and load balancing will need to be jointly optimised for an improved multi-objective CoMP clustering algorithm.

Available backhaul capacity is one of the biggest limitations for CoMP, especially JT-CoMP. A number of research works in literature focus on CoMP clustering where spectral efficiency and backhaul requirement is jointly optimised. Mayer *et al.* [83] presented the implications of backhaul channel reliability on spectral efficiency of the clusters. It's shown that, both JT and CB-CoMP scenarios give better spectral efficiency results with strong backhaul reliability. However, spectral efficiency improvement reduces sharply when backhaul reliability goes down. As discussed in Section VI-B1, caching popular multimedia content at the BS is an emerging research area for reducing backhaul requirement and hence enabling JT-CoMP in limited backhaul scenarios [67], [68].

Existing literature focuses on one limiting objective and investigates the trade-off against spectral efficiency gains. However, a more comprehensive CoMP clustering approach need to take all limiting factors in the same algorithm for intelligent clusters which jointly optimise backhaul bandwidth, energy efficiency, load balancing and spectral efficiency. For example, BS switch-off is a widely studied concept in literature as part of CoMP clustering, however only the SINR

constraints are taken into account to make sure there is enough coverage for BS switch-off. However, other constraints like RAN capacity, load balancing, backhaul bandwidth availability also need to be considered in a realistic network for BS switch-off decision. In this context, more research is required for multi-objective CoMP clustering algorithms with above mentioned constraints. A comparison of CoMP clustering algorithms based on aimed objective and their shortcomings are provided in Table III.

VII. FUTURE RESEARCH DIRECTIONS

In this section, we present open research areas for CoMP clustering challenge. Potential use of Big Data in proactive CoMP clustering is reviewed in the next subsection. Its followed by open research areas in dynamic clustering approaches, reviewing the challenges on complexity/gain trade-off of dynamic clustering and the need for comprehensive CoMP clustering solutions to maximise not only spectral efficiency but also other system objectives like load balancing, energy efficiency and backhaul optimisation.

A. Big Data Empowered Proactive CoMP Clustering

As mentioned in aforementioned sections, CoMP has the capability to significantly improve spectral efficiency and cell edge throughput through cooperation of limited number of BSs referred to as CoMP clusters. The state-of-the-art research on dynamic CoMP clustering in general have a reactive line

of action, i.e., CoMP clustering are designed/optimized with respect to current network conditions. For example load balancing targeted CoMP clustering will kick in when congestion is observed or diagnosed. However, in light of emerging 5G future cellular networks personified with ambitious QoS requirements of almost infinite capacity or zero latency [3], this approach will not be able to meet stringent performance requirements of 5G. This is because in classic dynamic CoMP clustering, certain time is required to observe the current conditions, find optimum clustering with respect to the objective function and current conditions and then trigger the appropriate clustering action. The resultant intrinsic delay is not compatible with 5G targeted QoE levels. Therefore for 5G, CoMP clustering paradigm requires proactive or predictive approach such that spatio-temporal future network state in terms of channel variation, mobility behaviour and capacity requirements can be predicted beforehand. This is possible through exploitation of the cognition of context of application as well as state of the network by inferring network-level intelligence from the massive amount of control, signalling, and contextual data known as Big Data as proposed in [4]. Key elements and sources of Big Data for mobile networks have been identified in [4]. By leveraging a dexterous combination of advanced techniques of machine learning, statistics and optimization, Big Data can be tapped to enable and empower CoMP clustering algorithms to achieve true performance gains of CoMP. Endowed with predictive capabilities, -thanks to Big Data- CoMP clustering algorithms can track, learn and dynamically build user mobility and demand profiles as well as channel characteristics models to predict future user locations coupled with service requests and channel state information. This can lead to timely efficient CoMP clustering as well as can help to alleviate high backhaul requirements. Another advantage of exploiting Big Data in CoMP clustering is that, it can represent the global state of the network which enables the global optimal CoMP clustering with respect to the defined objective functions such as energy efficiency, spectral efficiency or load balancing as opposed to relying only on the local information that may lead to only locally optimal CoMP clustering solutions. As the current research on CoMP clustering lacks this proactive perspective and to the best of our knowledge, currently no existing work targets "Proactive CoMP Clustering" in general and "Big Data empowered Proactive CoMP Clustering" in particular, therefore the goal of this section is to give future outlook of Big Data enabled proactive CoMP clustering. In subsections to follow, we briefly explain how Big Data can empower prediction based proactive CoMP clustering algorithms in terms of channel prediction, mobility prediction and user profiling.

1) Big Data in Cellular Networks: In the context of cellular networks, Big Data refers to the massive amounts of control, signalling and contextual data that is being routinely produced during day-to-day operation of cellular network. The potential constituents of Big Data in cellular networks are [4]:

a) Subscriber level data: The subscriber level data comprises of key performance indicators obtained from a voice or a data session initiated by the subscriber to give an indication of the accessibility, retainability and integrity performances of the network. Several metrics including blocked call rates, access failure rates, setup times, success rate, and hand-over failure rates project accessibility of the network. The dropped call rates, completion times, packet data protocol context and success rate together define the retainability of the network. The metrics like speech and data streaming quality, throughput, packet jitter and delay give an idea about user perceived quality of experience (QoE).

b) Cell level data: It refers to the measurements that are reported by a BS and all users within the coverage of that BS. Examples of useful cell level data streams are measurements reporting uplink noise floor in terms of reference interference power, channel based power information, physical resource block usage per cell, no. of active users per cell and minimization of driving test (MDT) measurements. MDT reports consist of the RSRP and reference signal received quality (RSRQ) values of the serving and neighbouring cells reported by the users to their serving BS [84], [85].

c) Core network level data: The core network data includes signalling information, historical alarm logs, equipment configuration lists and service and resource utilization accounting records (Call Data Records - CDRs and Extended Data Records XDRs) as well as aggregate statistics of network performance metrics.

d) Miscellaneous data: It consists of the structured information already stored in the separate databases including customer relationship management, customer complaint center and spectrum utility maps. This also includes unstructured information such as social media feeds, specific application usage patterns, and data from smart phone built-in sensors and applications.

2) Role of Big Data in Proactive CoMP Clustering: In perspective of proactive CoMP Clustering, channel maps built upon collected MDT reports and unified information of handover traces and CDRs are potential Big Data constituents that can be harnessed to predict future network state through machine learning algorithms and statistical techniques. Specifically, they can be utilized to predict future spatio-temporal rate requirements along with the associated channel state information as explained in subsequent subsections. This can pave the way for enabling timely efficient prediction based proactive user-centric dynamic CoMP clustering.

a) CoMP with big data aided channel prediction: Accurate and timely channel estimation is one of the most vital requirements of CoMP system. The coordinating BSs in a CoMP cluster are typically assumed to be connected to a centralized CCU through backhaul links [86]–[88]. In FDD systems, each user within CoMP cluster needs to estimate and predict the CSI from all BSs of the cluster and feed it back to the serving BS which is then forwarded to CCU. Based on the available CSI, joint transmission, user scheduling or coordinated beamforming schemes are designed. The quality of CSI has large impact on the performance of CoMP systems and clustering decisions. Restricted feedback and backhaul links induce different degrees of latencies resulting outdated measurements [89]. The X2 latencies observed in 3GPP-LTE deployed networks is of order 10 to 20 milliseconds [8], [90].

The outdated CSI leads to severe performance degradation in CoMP systems even when the users exhibit low mobility [7]. Channel estimation at the terminals as well as compression and quantization of CSI are further sources of inaccuracy.

In time-varying wireless channels, channel prediction is a popular approach to provide up-to-date channel information and it is shown in [91] that the performance of CoMP systems is improved significantly even with the large backhaul latency when channel prediction is applied. State of the art prediction techniques like Kalman and Weiner filtering make CoMP links more robust for CSI delays of few milliseconds and at moderate mobility [5]. Recently Doppler-delay-based prediction has been proposed wherein the channel for each link between a transmitter and a receiver antenna can be modelled by a number of multi-paths with their individual complex amplitude, delay and Doppler frequency. These parameters can be estimated for each path based upon the recent channel history embedded in Big Data and the future condition of the channel can be predicted by inserting the estimated parameters into the channel model. Both Doppler delay and Kalman prediction lead to significant improvement in Mean Squared Error (MSE) for the CSI that leads to better performance. The powerful Big Data aided CSI prediction can be an important enabler for CoMP clustering decisions.

Big Data can also play crucial role in proactive selection of BSs for cluster formation. One of the vital sources of Big Data in mobile communications are MDT reports consisting of RSRP and other channel quality related metrics reported by the users to their serving BS [84], [85]. The averaged RSRP values of the BSs, as reported by the UEs, can be compared to a threshold to determine which of these BSs should cooperate. Based on current MDT reports, future channel conditions can be predicted through conventional time-series forecasting methods. In case of sparse MDT reports due to small number of users like in small cells, light-weight Grey modelling techniques [92] that are useful for short term forecasting can be utilized as done in [93]. The grey model can predict the next RSRP value from data points obtained in the database. Therefore, instead of waiting for actual MDT reports, the predicted RSRP measurements can be fed to the channel estimation and subsequently to the cluster optimization algorithm that proactively adapts the cell clustering in CoMP perspective.

b) CoMP with big data aided mobility prediction: Big Data aided mobility prediction can play important role in proactive CoMP clustering decisions. Mobility prediction utilizes persons mobility history, i.e., a series of locations and corresponding dwell times to predict this persons next location, as well as his/her dwell time in that location [94]–[100]. In this way, CoMP clustering algorithms can plan in advance the clustering decisions thereby meeting the strict latency requirements of 5G networks. Big Data as identified in [4] also contains handover reports which contain Cell IDs and corresponding timestamps whenever user is handed over to new cell. Several techniques such as mobility pattern matching using mobility database, periodicity and multi-class classification and bio-inspired approaches as presented in [94]–[96] can be used to predict user mobility behaviour. Markov and hidden markov models have been commonly used for temporal-spatial prediction purposes as in [97] and [98]. Received signal strength indicator (RSSI) available in MDT reports can also be utilized to predict future location as has been done in [99] and [100]. The identified future location of the users along with the corresponding time stamps can be fed to the CoMP dynamic clustering algorithms (both user-centric as well as network-centric) for computing optimal clusters.

Mobility behaviour of the users directly affects the CoMP clustering decisions as CSI has small validity period for high speed users and clustering decisions needs to be performed frequently leading to high computational overheads. One solution can be that low speed or static users can be served by CoMP cluster BSs, however, high speed mobile users continue to be served by single BS. By utilizing RSSI and the cell sizes information embedded in Big Data and predicted future user locations, CoMP clustering algorithms can be executed beforehand leading to significant reduction in latency and bandwidth requirements.

c) CoMP with big data aided user profiling: Call Data Records (CDRs) are one of the key elements of the Big Data that can be harnessed from a cellular network. CDRs reflect mobile users behaviour and give out clues on how the users utilize the network resources. CDRs contain information about the voice calls and data usage pattern and are important markers of temporal-spatial capacity requirements across the deployed network [101]–[103]. CDRs can be utilized to profile the network usage behaviour of the mobile users which in turn can be utilized for user-centric or behaviour-centric CoMP clustering. By applying machine learning and statistical tools on CDRs, we can determine the capacity requirements of the users at different time periods and can utilize this profile information to cluster the CoMP enabled BSs to satisfy the expected QoS requirements of the users.

Social media feeds are another element within Big Data that give helpful insights about the interaction of the users and expected temporal-spatial demand of network resources across the network. Among many online social networks, Twitter is one of the popular ways users share information and experience socially on the Web. Twitter data can be mined through application program interface (Twitter APIs) wherein each timestamped tweet contains number of useful information like location, number of re-tweets, number of favourites, message itself and hashtags. Twitter data can be utilized to estimate traffic demand as number of tweets is highly correlated with the number of people in confined places [104]. It can also be utilized to assess networks QoE from subscriber's perspective [105]. The social media feeds together with the CDRs can be taped to accurately model the user behaviour and can be utilized to optimize user-centric CoMP clustering algorithms.

3) Tapping Big Data for CoMP Clustering: Multifaceted and multifarious Big Data can help to enable and optimize the proactive CoMP clustering algorithms. Big Data consists of big pool of training datasets that is of significant advantage for prediction techniques based on advanced supervised machine learning algorithms like deep learning methods [106]. However, Big Data comes with its own set of challenges like how to efficiently tap the potential of this Big Data in real
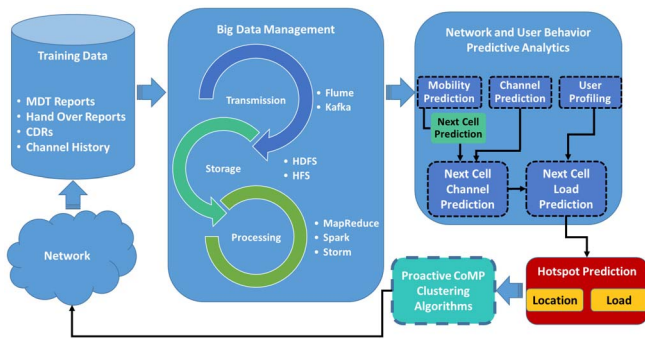
Fig. 7. Big Data Empowered Proactive CoMP Clustering Framework.

time that is hindered by four inherent characteristics of Big Data, i.e., Volume, Variety, Velocity and Veracity [4]. Big Data management tools under umbrella of Hadoop ecosystem are potential enablers to deal with the acute dynamicity of the Big Data. The main components of Big data processing platform consist of [107]:

1) *Trasmission Module* consisting of Flume [108] and Kafka [109] that uploads network data in real time with stable transmission to the cloud platform.

2) *Storage Module* consisting of Distributed File System (HDFS) [110] and HBase [111] with high fault tolerance capability.

3) *Processing Module* comprising of MapReduce [112] for parallel distributed processing, Spark [113] for cyclic data flow and in-memory computing and Storm [114] for enabling real-time analysis.

4) *Management Module* to monitor the whole platform with Flume collecting the monitored data and Zookeeper [115] to modify configuration parameters of each machine and equipment.

The processed statistics from Big Data can then be fed to advanced machine learning methods to model network and user behaviour and predict future spatio-temporal network states. By knowing probable future user locations, their expected rate requirements and estimated channel state information, CoMP clustering algorithms can proactively adapt themselves on the fly to cope with acute dynamics of cellular networks resulting in seamless quality of perception. This framework is depicted in Figure 7.

4) Relevant Work on Big Data Driven Proactive CoMP Clustering: Although no existing work target proactive CoMP clustering leveraging Big Data explicitly yet, there exist certain works, wherein dynamic CoMP clustering is performed targeted at hotspots, assuming hotspot location are already somehow known by the network. The Big Data processing framework presented above cannot only identify the future hotspots but it can also predict future load, e.g., using data of mobility traces and past CDR records [107], [116], [117]. Once a hotspot is characterized, the appropriate CoMP algorithm can be leveraged to cope with high capacity demands for hotspots. Examples of work which can leverage this idea include study in [118]. Rezagah *et al.* [118] have proposed a novel cell structuring and clustering algorithm to dynamically transfer network resources from sparse cells to crowded cells

or hotspots wherein optimal large cooperative clusters, performing joint transmission (JT), are formed around hotspots and the coverage of BSs are transferred to hotspots by dynamically changing the antennas beam angles. With the proposed big data framework in place, this process can be done proactively, instead of reactively, thereby further improving the QoE CoMP can offer.

Another work [119] has proposed a cross-tier cooperation in non-uniform HetNets wherein cell edge hotspot users are served by CoMP BSs. The location of clustered users or hotspots present in the network have also been utilized in [78] and [120]–[122] wherein network configuration parameters (antenna parameters) are optimized w.r.t the identified hotspots. The underlying phenomenon is inherently the same as in case of dynamic CoMP Clustering since network parameters optimization is done based on hotspot location. The aforementioned algorithms initiate reactively assuming hotspots have already formed into the network and their location is cent percent known accurately. However, with Big Data Predictive Analytics, formation of hotspots can be predicted beforehand as explained above and dynamic CoMP clustering can be performed well in time to minimise QoE degradation time. This is where Big Data comes into the picture.

Very recently, some works have emerged that leverage Big Data driven predictive analytics in mobile networks for predicting hotspot formation using CDRs. The work in [107] has performed Big Data collection, storage, and pre-processing of CDRs and has proposed:

i The rules for extracting location data, and constructing people trajectories

ii The methods for solving data noise (i.e., cell tower oscillations)

iii The algorithms for discovering common mobility patterns in densely populated area

iv Identifying hotspots.

Similarly, in [116], [117], and [123]–[125], Big Data technologies and analytical algorithms have been used for predicting hotspot formation or forecasting pedestrian destinations with satisfactory accuracy.

In a nutshell, Big Data driven predictive analytics predicting the future spatio-temporal state of the network accurately and using this knowledge for dynamic CoMP clustering well in time is the future of the CoMP clustering that can truly unleash the real potential of CoMP and can be instrumental in improving user experience in future 5G cellular networks. Presence of more data (Big Data) results in better and accurate models as it allows the data to tell for itself,İ instead of relying on assumptions and weak correlations since a weak assumption coupled with complex algorithms is far less efficient than using more data with simpler algorithms. This fact has been captured by many studies, e.g., [126] and [127] wherein results suggests for a given problem, adding more examples to the training set monotonically increases the accuracy of the model. However with aid of Big Data, as the ability to generate better predictions continues to improve, it is noteworthy that the accuracy of these predictions is only as good as the accuracy of the underlying data (garbage-in, garbage-out). Leveraging Big Data of poor quality for proactive CoMP clustering might

produce erroneous predictions, counter-productive clustering decisions and poor performance than that achievable with conventional reactive dynamic CoMP clustering.

B. Dynamic CoMP Clustering Challenges

1) Complexity/Gain Trade-off for Dynamic CoMP Clustering: Dynamic clustering has more potential for better performance gains due to its ability to respond to network and user/service profile changes. Inter-cluster interference can be mitigated with dynamically changing cluster boundaries. Both user-centric, network-centric and hybrid algorithms have been studied in the literature. CoMP clustering research on user-centric approaches lacks scalability and suffer high scheduling/precoding design complexity. Network-centric approaches mainly fail to provide full CoMP gains when compared to user-centric cluster design. Hybrid clustering provide a balance between complexity and CoMP gain trade-off. However, existing works fail to provide fully dynamic hybrid solutions. The challenge with fully dynamic solution is the increased complexity especially with increased scheduling and precoding design and additional overheads. More rigorous research is required on novel hybrid solutions where dynamic user-centric clustering is employed within a dynamic network-centric clustering algorithm and the gains of such algorithms against the complexity and additional overhead costs.

2) Multi-Objective CoMP Clustering: CoMP is envisioned for mitigating inter-cell interference and hence increasing spectral efficiency. Hence the primary aim for CoMP clustering is to maximise spectral efficiency, however other limitations like load balancing, backhaul bandwidth availability, system capacity and energy efficiency are also taken into account for improved clustering solutions. Existing literature focuses on maximising spectral efficiency along side with one more objective, mostly focusing on backhaul bandwidth and energy efficiency constraints. However, a more comprehensive approach is required to take all constraints into account for a realistic CoMP clustering solution. We outline the potential research directions in multi-objective clustering as below.

a) Load balancing: As discussed in Section VI-D, load balancing is an increasingly important concept for mobile networks due to the exponential increase in data demand [1]. CoMP is likely to be deployed in interference limited networks where there is high data demand. An interesting research area is to develop CoMP clustering algorithms to support load balancing while spectral efficiency is maximised. A load-aware, user-centric CoMP clustering approach is presented in our previous work [128], however further research is required to analyse fully dynamic CoMP clustering techniques and the trade-off between load balancing gains and potential losses on spectral efficiency.

b) Backhaul optimisation: A number of research are conducted for CoMP clustering which takes backhaul bandwidth limitation into account. The main contributors for high backhaul bandwidth requirement such as cluster size [63], [64] and type of cooperation (i.e., CB or JT) [67], [68] are dynamically changed to adapt to limited backhaul bandwidth availability.

RAN caching is employed in [67] and [68] to reduce high backhaul bandwidth dependency for cooperation. However, backhaul bandwidth limitation is studied in isolation, not in relation to other objective functions like spectral efficiency and load balancing. An open research area is to develop backhaul-aware CoMP clustering algorithms which aim to maximise spectral efficiency and user satisfaction in relation to backhaul limitations and load balancing.

Another open research area is to utilise Big Data for RAN caching to compensate for the high backhaul requirement. Big Data aided proactive caching can play significant role in JT mode of CoMP wherein user data is shared among cooperating BSs. Such proactive caching can relax high backhaul requirements of JT CoMP.

c) Energy efficiency: CoMP deployment and intelligent clustering solutions can improve energy efficiency especially with increasing the number of sleeping BSs [31], [72], [73]. BS sleeping has been employed in most works to improve energy efficiency, however only SINR constraints are taken into account for BS sleeping to make sure there is coverage available for all users. As discussed in Section VI-E, other constraints like system capacity and backhaul bandwidth will need to be taken into account for BS sleeping. For example, a more realistic approach should consider load balancing while making decision for BS switch-off with the aim of maximising energy efficiency. Sleeping cells may need to be switched on to handle additional load in the network, however it comes with the additional signalling cost overhead and degradation on energy efficiency.

Furthermore, extensive research is required on comprehensive multi-objective clustering algorithms to include all limitations/objectives, i.e., spectral efficiency, energy efficiency, backhaul bandwidth and load conditions into algorithm design and analyse the trade-off between multiple objective gains and associated costs. Analytical tools like coalitional game theory can be utilised for merging multiple objectives into a single payoff function for exhaustive multi-objective CoMP clustering design. Trade-off between different objectives and optimum balance between these metrics is an area worth exploring further. Moreover, Big Data aided predictive models need to be explored further for novel proactive multi-objective proactive CoMP clustering design to support much faster response rates required for future networks.

VIII. Conclusion

This article provides an extensive survey on the CoMP clustering methods for future cellular networks. We first give the motivation for CoMP for future wireless networks and briefly provide an outline of CoMP implementation challenges and the need for CoMP clustering. We then provide a section to give brief tutorial about different types of cooperation, associated challenges and propose network architectures like CDSA and C-RAN which will enable CoMP implementation. The core of the article provides an extensive survey on CoMP clustering techniques available in the literature and introduce two novel taxonomies for CoMP clustering algorithms based on self-organisation and aimed objective function.

Firstly, we provide a CoMP clustering taxonomy based on self organisation, and critically discuss static, semi-static and dynamic CoMP clustering works in literature. Dynamic clustering algorithms are further divided based on their approach, network-centric and user-centric approaches, their benefits and shortcomings are highlighted.

Secondly, we present a novel CoMP clustering taxonomy based on the objective function. CoMP clustering algorithms aiming for spectral efficiency, energy efficiency, backhaul optimisation and load balancing are extensively discussed. More focus is given on comprehensive multi-objective clustering, available works in literature are presented, shortcomings are identified in detail.

We then outline open research areas for CoMP clustering and propose potential approaches for solutions. Proactive CoMP clustering is envisioned to accommodate much faster response rates required for 5G. We highlight the potential use of Big Data to empower prediction based CoMP clustering algorithms. Big Data in cellular networks context is identified, and use of Big Data for channel prediction, mobility prediction and user profiling prediction is discussed. We identify Big Data aided prediction models to form a future outlook in prediction based CoMP clustering. We then discuss further future research areas on dynamic CoMP clustering complexity/gain trade off and multi-objective CoMP clustering algorithms to optimise load balancing, backhaul limitation, energy efficiency and spectral efficiency simultaneously.

REFERENCES

[1] "Cisco visual networking index: Global mobile data traffic forecast update, 2015–2020," White Paper, Cisco, San Jose, CA, USA, Feb. 2016.

[2] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 71–78, Mar. 2014.

[3] "Looking ahead to 5G," White Paper, Nokia Netw., Espoo, Finland, Jul. 2014.

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

[5] V. Jungnickel *et al.*, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

[6] D. López-Pérez, M. Ding, H. Claussen, and A. H. Jafari, "Towards 1 Gbps/UE in cellular systems: Understanding ultra-dense small cell deployments," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2078–2101, 4th Quart., 2015.

[7] 3GPP, "Coordinated multi-point operation for LTE physical layer aspects," 3rd Gener. Partnership Project (3GPP), TR 36.819 R11 v11.2.0, Sep. 2013.

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

[9] D. Lee *et al.*, "Coordinated multipoint transmission and reception in LTE-advanced: Deployment scenarios and operational challenges," *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.

[10] P. Marsch and G. P. Fettweis, *Coordinated Multi-Point in Mobile Communications: From Theory to Practice*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[11] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated multipoint transmission/reception techniques for LTE-advanced [Coordinated and Distributed MIMO]," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 26–34, Jun. 2010.

[12] T. Biermann, L. Scalia, C. Choi, W. Kellerer, and H. Karl, "How backhaul networks influence the feasibility of coordinated multipoint in cellular networks [Accepted From Open Call]," *IEEE Commun. Mag.*, vol. 51, no. 8, pp. 168–176, Aug. 2013.

[13] A. Beylerian and T. Ohtsuki, "Multi-point fairness in resource allocation for C-RAN downlink CoMP transmission," *EURASIP J. Wireless Commun. Netw.*, vol. 2016, no. 1, p. 12, 2016.

[14] A. Beylerian and T. Ohtsuki, "Coordinated non-orthogonal multiple access (CO-NOMA)," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, Washington, DC, USA, Dec. 2016.

[15] J. B. Rao and A. O. Fapojuwo, "A survey of energy efficient resource management techniques for multicell cellular networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 154–180, 1st Quart., 2014.

[16] E. Hossain, D. I. Kim, and V. K. Bhargava, *Cooperative Cellular Wireless Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[17] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Control-data separation architecture for cellular radio access networks: A survey and outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 446–465, 1st Quart., 2015.

[18] H. A. U. Mustafa, M. A. Imran, M. Z. Shakir, A. Imran, and R. Tafazolli, "Separation framework: An enabler for cooperative and D2D communication for future 5G networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 419–445, 1st Quart., 2016.

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

[20] 3GPP, "Requirements for further advancements for evolved universal terrestrial radio access (E-UTRA) (LTE-advanced)," 3rd Gener. Partnership Project (3GPP), TR 36.913 V9.0.0, Dec. 2009.

[21] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects," 3rd Gener. Partnership Project (3GPP), Tech. Rep. 36.814, Mar. 2010.

[22] C. Mobile, "C-RAN: The road towards green RAN," White Paper, 2011.

[23] K. Sundaresan, M. Y. Arslan, S. Singh, S. Rangarajan, and S. V. Krishnamurthy, "FluidNet: A flexible cloud-based radio access network for small cells," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, Miami, FL, USA, 2013, pp. 99–110.

[24] I. Chih-Lin *et al.*, "Toward green and soft: A 5G perspective," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 66–73, Feb. 2014.

[25] Q. Wang *et al.*, "Application of BBU+RRU based CoMP system to LTE-advanced," in *Proc. IEEE Int. Conf. Commun. Workshops*, Dresden, Germany, 2009, pp. 1–5.

[26] A. Davydov, G. Morozov, I. Bolotin, and A. Papathanassiou, "Evaluation of joint transmission CoMP in C-RAN based LTE-A HetNets with large coordination areas," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Atlanta, GA, USA, 2013, pp. 801–806.

[27] T. Alsedairy, Y. Qi, A. Imran, M. A. Imran, and B. Evans, "Self organising cloud cells: A resource efficient network densification strategy," *Trans. Emerg. Telecommun. Technol.*, vol. 26, no. 8, pp. 1096–1107, 2015.

[28] I. D. Garcia *et al.*, "Impact of base station cooperation on cell planning," *EURASIP J. Wireless Commun. Netw.*, vol. 2010, no. 1, p. 5, 2010.

[29] P. Marsch and G. Fettweis, "Static clustering for cooperative multi-point (CoMP) in mobile communications," in *Proc. IEEE Int. Conf. Commun.*, Kyoto, Japan, 2011, pp. 1–6.

[30] A. S. Rodriguez and A. G. Armada, "Analysis of the cluster size in coordinated multipoint transmission," in *Proc. IEEE 19th Signal Process. Commun. Appl. Conf.*, 2011, pp. 1197–1200.

[31] G. Cili, H. Yanikomeroglu, and F. R. Yu, "Cell switch off technique combined with coordinated multi-point (CoMP) transmission for energy efficiency in beyond-LTE cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Ottawa, ON, Canada, 2012, pp. 5931–5935.

[32] 3GPP, "Evolved universal terrestrial radio access network (E-UTRAN); self-configuring and self-optimizing network (SON) use cases and solutions," 3rd Gener. Partnership Project (3GPP), TR 36.902 R9 v9.3.1, Mar. 2011.

[33] 3GPP, "Telecommunication management; self-configuration of network elements; concepts and requirements," 3rd Gener. Partnership Project (3GPP), TS 32.501 R12 v12.1.0, Dec. 2013.

[34] 3GPP, "Telecommunication management; self-organizing networks (SON) policy network resource model (NRM) integration reference point (IRP); requirements," 3rd Gener. Partnership Project (3GPP), TS 32.521 R11 v11.1.0, Dec. 2012.

[35] 3GPP, "Telecommunication management; self-organizing networks (SON); self-healing concepts and requirements," 3rd Gener. Partnership Project (3GPP), TS 32.541 R12 v12.0.0, Oct. 2014.

[36] S. S. Ali and N. Saxena, "A novel static clustering approach for CoMP," in *Proc. IEEE 7th Int. Conf. Comput. Converg. Technol. (ICCCT)*, Seoul, South Korea, 2012, pp. 757–762.

[37] H. Shimodaira *et al.*, "Diamond cellular network—Optimal combination of small power basestations and CoMP cellular networks," in *Proc. IEEE 24th Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC Workshops)*, London, U.K., 2013, pp. 163–167.

[38] A. Barbieri *et al.*, "Coordinated downlink multi-point communications in heterogeneous cellular networks," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, 2012, pp. 7–16.

[39] A. M. Hamza and J. W. Mark, "A clustering scheme based on timing requirements in coordinated base-stations cooperative communications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, 2013, pp. 3764–3769.

[40] F. Huang, Y. Wang, J. Geng, M. Wu, and D. Yang, "Clustering approach in coordinated multi-point transmission/reception system," in *Proc. IEEE 72nd Veh. Technol. Conf. Fall (VTC Fall)*, Ottawa, ON, Canada, 2010, pp. 1–5.

[41] S. A. Ramprashad, G. Caire, and H. C. Papadopoulos, "A joint scheduling and cell clustering scheme for MU-MIMO downlink with limited coordination," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Cape Town, South Africa, 2010, pp. 1–6.

[42] V. Pichapati and P. Gupta, "Practical considerations in cluster design for co-ordinated multipoint (CoMP) systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Budapest, Hungary, 2013, pp. 5860–5865.

[43] H. Sun, X. Zhang, and W. Fang, "Dynamic cell clustering design for realistic coordinated multipoint downlink transmission," in *Proc. IEEE 22nd Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Toronto, ON, Canada, 2011, pp. 1331–1335.

[44] A. Papadogiannis, D. Gesbert, and E. Hardouin, "A dynamic clustering approach in wireless networks with multi-cell cooperative processing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, 2008, pp. 4033–4037.

[45] J. Liu and D. Wang, "An improved dynamic clustering algorithm for multi-user distributed antenna system," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Nanjing, China, 2009, pp. 1–5.

[46] S. Feng, W. Feng, H. Mao, and J. Lu, "Overlapped clustering for CoMP transmissions in massively dense wireless networks," in *Proc. IEEE Int. Conf. Commun. Syst. (ICCS)*, 2014, pp. 308–312.

[47] F. Guidolin, L. Badia, and M. Zorzi, "A distributed clustering algorithm for coordinated multipoint in LTE networks," *IEEE Wireless Commun. Lett.*, vol. 3, no. 5, pp. 517–520, Oct. 2014.

[48] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar, "Coalitional game theory for communication networks," *IEEE Signal Process. Mag.*, vol. 26, no. 5, pp. 77–97, Sep. 2009.

[49] H. Li, H. Tian, C. Qin, and Y. Pei, "A novel distributed cluster combination method for CoMP in LTE-A system," in *Proc. IEEE 15th Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Taipei, Taiwan, 2012, pp. 614–618.

[50] J.-M. Moon and D.-H. Cho, "Formation of cooperative cluster for coordinated transmission in multi-cell wireless networks," in *Proc. IEEE Consum. Commun. Netw. Conf. (CCNC)*, Las Vegas, NV, USA, 2013, pp. 528–533.

[51] W. Saad, Z. Han, M. Debbah, and A. Hjorungnes, "A distributed coalition formation framework for fair user cooperation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4580–4593, Sep. 2009.

[52] R. Weber, A. Garavaglia, M. Schulist, S. Brueck, and A. Dekorsy, "Self-organizing adaptive clustering for cooperative multipoint transmission," in *Proc. IEEE 73rd Veh. Technol. Conf. (VTC Spring)*, Budapest, Hungary, 2011, pp. 1–5.

[53] P. Baracca, F. Boccardi, and V. Braun, "A dynamic joint clustering scheduling algorithm for downlink CoMP systems with limited CSI," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Paris, France, 2012, pp. 830–834.

[54] V. Garcia, Y. Zhou, and J. Shi, "Coordinated multipoint transmission in dense cellular networks with user-centric adaptive clustering," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4297–4308, Aug. 2014.

[55] C. Zeng, P. Ren, C. Zhang, and G. Lv, "A hybrid clustering approach in coordinated multi-point transmission system," in *Proc. 7th Int. ICST Conf. Commun. Netw. China (CHINACOM)*, Kunming, China, 2012, pp. 627–631.

[56] C. Li, J. Zhang, M. Haenggi, and K. B. Letaief, "User-centric intercell interference nulling for downlink small cell networks," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1419–1431, Apr. 2015.

[57] J. Zhao, T. Zhang, Z. Zeng, Q. Gao, and S. Sun, "An overlapped clustering scheme of coordinated multi-point transmission for LTE-A systems," in *Proc. IEEE 14th Int. Conf. Commun. Technol. (ICCT)*, Chengdu, China, 2012, pp. 479–484.

[58] W. Xu, J. Huang, F. Yang, and H. Zhang, "Dynamic cell clustering for CoMP in LTE-A and its calibrated system level performance evaluation," in *Proc. IEEE 5th Int. Symp. Microw. Antenna Propag. EMC Technol. Wireless Commun. (MAPE)*, Chengdu, China, 2013, pp. 71–77.

[59] M. Liu, B. Li, and Z. Tong, "Cell clustering and resource allocation scheme for CoMP SU-MIMO," *Wireless Pers. Commun.*, vol. 69, no. 4, pp. 1399–1411, Apr. 2013.

[60] E. Katranaras, M. A. Imran, and M. Dianati, "Energy-aware clustering for multi-cell joint transmission in LTE networks," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC)*, Budapest, Hungary, 2013, pp. 419–424.

[61] D. Samardzija and H. Huang, "Determining backhaul bandwidth requirements for network MIMO," in *Proc. 17th Eur. Signal Process. Conf. (EUSIPCO)*, Glasgow, U.K., 2009, pp. 1494–1498.

[62] T. Biermann, L. Scalia, C. Choi, H. Karl, and W. Kellerer, "Improving CoMP cluster feasibility by dynamic serving base station reassignment," in *Proc. IEEE 22nd Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Toronto, ON, Canada, 2011, pp. 1325–1330.

[63] J. Zhao and Z. Lei, "Clustering methods for base station cooperation," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Paris, France, 2012, pp. 946–951.

[64] J. Zhao, T. Q. S. Quek, and Z. Lei, "Semi-distributed clustering method for CoMP with limited backhaul data transfer," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Shanghai, China, 2013, pp. 3776–3781.

[65] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.

[66] B. A. Ramanan et al., "Cacheability analysis of HTTP traffic in an operational LTE network," in *Proc. Wireless Telecommun. Symp. (WTS)*, Phoenix, AZ, USA, 2013, pp. 1–8.

[67] A. Liu and V. K. N. 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.

[68] M. Deghel, E. Bastug, M. Assaad, and M. Debbah, "On the benefits of edge caching for MIMO interference alignment," in *Proc. 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Stockholm, Sweden, 2015, pp. 655–659.

[69] G. Auer et al., "Enablers for energy efficient wireless networks," in *Proc. 72nd Veh. Technol. Conf. VTC*, Ottawa, ON, Canada, 2010, pp. 1–5.

[70] V. Mancuso and S. Alouf, "Reducing costs and pollution in cellular networks," *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 63–71, Aug. 2011.

[71] G. Rittenhouse, S. Goyal, D. T. Neilson, and S. Samuel, "Sustainable telecommunications," in *Proc. Tech. Symp. ITU Telecom World*, Geneva, Switzerland, 2011, pp. 19–23.

[72] S. Han, C. Yang, G. Wang, and M. Lei, "On the energy efficiency of base station sleeping with multicell cooperative transmission," in *Proc. 22nd Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Toronto, ON, Canada, 2011, pp. 1536–1540.

[73] D. Cao, S. Zhou, C. Zhang, and Z. Niu, "Energy saving performance comparison of coordinated multi-point transmission and wireless relaying," in *Proc. Glob. Telecommun. Conf. (GLOBECOM)*, Miami, FL, USA, 2010, pp. 1–5.

[74] Z. Niu, S. Zhou, Y. Hua, Q. Zhang, and D. Cao, "Energy-aware network planning for wireless cellular system with inter-cell cooperation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 4, pp. 1412–1423, Apr. 2012.

[75] A. Sharma, A. Roy, S. Ghosal, R. Chaki, and U. Bhattacharya, "Load balancing in cellular network: A review," in *Proc. 3rd Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Coimbatore, India, 2012, pp. 1–5.

[76] I. Viering, M. Dottling, and A. Lobinger, "A mathematical perspective of self-optimizing wireless networks," in *Proc. Int. Conf. Commun. (ICC)*, Dresden, Germany, 2009, pp. 1–6.

[77] A. Imran, E. Yaacoub, M. A. Imran, and R. Tafazolli, "Distributed load balancing through self organisation of cell size in cellular systems," in *Proc. 23rd Int. Symp. Pers. Indoor Mobile Radio Commun. (PIMRC)*, Sydney, NSW, Australia, 2012, pp. 1114–1119.

[78] A. Imran, M. A. Imran, A. ul Quddus, and R. Tafazolli, "Distributed spectral efficiency optimization at hotspots through self organisation of BS tilts," in *Proc. GLOBECOM Workshops*, Houston, TX, USA, 2011, pp. 570–574.

[79] H. Kim, G. De Veciana, X. Yang, and M. Venkatachalam, "Alpha-optimal user association and cell load balancing in wireless networks," in *Proc. INFOCOM*, San Diego, CA, USA, 2010, pp. 1–5.

[80] Q. Ye et al., "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[81] I. Gádor. (Jul. 2012). "EARTH Project Deliverable D3.3: Final Report on Green Network Technologies." [Online]. Available: http://bscw.ict-earth.eu/pub/bscw.cgi/d70472/EARTHWP3D3.3.pdf

[82] Y. Li, Y. Ma, Y. Wang, and W. Zhao, "Base station sleeping with dynamical clustering strategy of CoMP in LTE-advanced," in *Proc. IEEE Int. Conf. Green Comput. Commun.*, Beijing, China, Aug. 2013, pp. 157–162.

[83] Z. Mayer, J. Li, A. Papadogiannis, and T. Svensson, "On the impact of backhaul channel reliability on cooperative wireless networks," in *Proc. Int. Conf. Commun. (ICC)*, Budapest, Hungary, 2013, pp. 5284–5289.

[84] S. Hämäläinen, H. Sanneck, and C. Sartori, *LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency*. Hoboken, NJ, USA: Wiley, 2012.

[85] 3GPP, "Universal Terrestrial Radio Access (UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Measurement Collection for Minimization of Drive Tests (MDT); Overall

Description; Stage 2," 3rd Gener. Partnership Project (3GPP), TS 37.320 R12 v12.2.0, Sep. 2014.

[86] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, Jr., "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.

[87] A. Papadogiannis and G. C. Alexandropoulos, "The value of dynamic clustering of base stations for future wireless networks," in *Proc. Int. Conf. Fuzzy Syst. (FUZZ)*, Barcelona, Spain, 2010, pp. 1–6.

[88] J. Li, C. Botella, and T. Svensson, "Resource allocation for clustered network MIMO OFDMA systems," *EURASIP J. Wireless Commun. Netw.*, vol. 2012, no. 1, pp. 1–19, 2012.

[89] 3GPP, "CoMP performance evaluation under low-capacity/high-latency backhaul," 3rd Gener. Partnership Project (3GPP), TR R1-112339, Aug. 2011. [Online]. Available: http://www.3gpp.org/ftp/tsg_ran/WG1_RL1/ TSGR1_66/Docs/R1-112339.zip

[90] 3GPP, "Evolved universal terrestrial radio access network (E-UTRAN); X2 general aspects and principles," 3rd Gener. Partnership Project (3GPP), TS 36.420 R12 v12.1.0, Dec. 2014. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36420.htm

[91] L. Su, C. Yang, and S. Han, "The value of channel prediction in CoMP systems with large backhaul latency," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4577–4590, Nov. 2013.

[92] E. Kayacan, B. Ulutas, and O. Kaynak, "Grey system theory-based models in time series prediction," *Expert Syst. Appl.*, vol. 37, no. 2, pp. 1784–1789, 2010.

[93] O. Onireti *et al.*, "A cell outage management framework for dense heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 4, pp. 2097–2113, Apr. 2016.

[94] H. Ge, X. Wen, W. Zheng, Z. Lu, and B. Wang, "A history-based handover prediction for LTE systems," in *Proc. Int. Symp. Comput. Netw. Multimedia Technol. (CNMT)*, Wuhan, China, 2009, pp. 1–4.

[95] J. Wang and B. Prabhala, "Periodicity based next place prediction," in *Proc. Mobile Data Challenge Nokia Workshop 10th Int. Conf. Pervasive Comput.*, Urbana, IL, USA, 2012.

[96] M. Daoui, A. M'zoughi, M. Lalam, M. Belkadi, and R. Aoudjit, "Mobility prediction based on an ant system," *Comput. commun.*, vol. 31, no. 14, pp. 3090–3097, 2008.

[97] A. Ulvan, M. Ulvan, and R. Bestak, "The enhancement of handover strategy by mobility prediction in broadband wireless access," in *Proc. Netw. Electron. Comm. Res. Conf. (NAEC)*, Dallas, TX, USA, 2009, pp. 266–276.

[98] S. Gambs, M.-O. Killijian, M. N. del, and P. Cortez, "Next place prediction using mobility Markov chains," in *Proc. 1st Workshop Meas. Privacy Mobility*, Bern, Switzerland, 2012, pp. 1–6.

[99] A. S. Sadiq, K. A. Bakar, K. Z. Ghafoor, and A. J. Gonzalez, "Mobility and signal strength-aware handover decision in mobile IPv6 based wireless LAN," in *Proc. Int. MultiConf. Eng. Comput. Sci.*, vol. 1. Hong Kong, 2011.

[100] T.-H. Kim and J.-W. Kim, "Handover optimization with user mobility prediction for femtocell-based wireless networks," *Int. J. Eng. Technol.*, vol. 5, no. 2, pp. 1829–1837, 2013.

[101] V. Frias-Martinez, C. Soguero, and E. Frias-Martinez, "Estimation of urban commuting patterns using cellphone network data," in *Proc. ACM SIGKDD Int. Workshop Urban Comput.*, Beijing, China, 2012, pp. 9–16.

[102] M. Ficek and L. Kencl, "Inter-call mobility model: A spatio-temporal refinement of call data records using a Gaussian mixture model," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 469–477.

[103] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in Rome," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 141–151, Mar. 2011.

[104] F. Botta, H. S. Moat, and T. Preis, "Quantifying crowd size with mobile phone and Twitter data," *Roy. Soc. Open Sci.*, vol. 2, no. 5, 2015, Art. no. 150162.

[105] T. Qiu *et al.*, "Listen to me if you can: Tracking user experience of mobile network on social media," in *Proc. 10th ACM SIGCOMM Conf. Internet Meas.*, Melbourne, VIC, Australia, 2010, pp. 288–293.

[106] M. M. Najafabadi *et al.*, "Deep learning applications and challenges in big data analytics," *J. Big Data*, vol. 2, no. 1, pp. 1–21, 2015.

[107] Y. Qiao, Y. Cheng, J. Yang, J. Liu, and N. Kato, "A mobility analytical framework for big mobile data in densely populated area," *IEEE Trans. Veh. Technol.*, to be published.

[108] *Apache Storm*. Accessed on Feb. 8, 2017. [Online]. Available: https://flume.apache.org

[109] *Apache Storm*. Accessed on Feb. 8, 2017. [Online]. Available: https://kafka.apache.org

[110] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Proc. IEEE 26th Symp. Mass Stor. Syst. Technol. (MSST)*, 2010, pp. 1–10.

[111] L. George, *HBase: The Definitive Guide*. Sebastopol, CA, USA: O'Reilly Media, 2011.

[112] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, pp. 107–113, 2008.

[113] *Apache Storm*. Accessed on Feb. 8, 2017. [Online]. Available: https://spark.apache.org

[114] *Apache Storm*. Accessed on Feb. 8, 2017. [Online]. Available: https://storm.apache.org

[115] P. Hunt, M. Konar, F. P. Junqueira, and B. Reed, "ZooKeeper: Wait-free coordination for Internet-scale systems," in *Proc. USENIX Annu. Tech. Conf.*, vol. 8. Boston, MA, USA, 2010, p. 11.

[116] Q. Lv, Y. Qiao, N. Ansari, J. Liu, and J. Yang, "Big data driven hidden Markov model based individual mobility prediction at points of interest," *IEEE Trans. Veh. Technol.*, to be published.

[117] A. Nika *et al.*, "Understanding and predicting data hotspots in cellular networks," *Mobile Netw. Appl.*, vol. 21, no. 3, pp. 402–413, 2016.

[118] R. E. Rezagah *et al.*, "Large scale cooperation in cellular networks with non-uniform user distribution," in *Proc. Globecom Workshops (GC Wkshps)*, Atlanta, GA, USA, 2013, pp. 43–48.

[119] K. Yang, J. Zhang, X. Zhang, and W. Wang, "Edge aware cross-tier base station cooperation in heterogeneous wireless networks with non-uniformly-distributed nodes," *IET Commun.*, vol. 10, no. 15, pp. 1896–1903, Oct. 2016.

[120] M. S. Parwez, H. Farooq, A. Imran, and H. Refai, "Spectral efficiency self-optimization through dynamic user clustering and beam steering," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, 2015, pp. 1–7.

[121] N. Bitar, A. Imran, and H. Refai, "A user centric self-optimizing grid-based approach for antenna steering based on call detail records," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Doha, Qatar, 2016, pp. 1–6.

[122] T. V. Santana, A. Galindo-Serrano, B. Sayrac, and S. M. López, "Dynamic network configuration: Hotspot identification for virtual small cells," in *Proc. Int. Symp. Wireless Commun. Syst. (ISWCS)*, Poznań, Poland, 2016, pp. 49–53.

[123] A. Alahi, V. Ramanathan, and L. Fei-Fei, "Socially-aware large-scale crowd forecasting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, 2014, pp. 2211–2218.

[124] R. O. Sinnott and W. Chen, "Estimating crowd sizes through social media," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Sydney, NSW, Australia, 2016, pp. 1–6.

[125] H.-W. Chang, Y.-C. Tai, and J. Y.-J. Hsu, "Context-aware taxi demand hotspots prediction," *Int. J. Bus. Intell. Data Min.*, vol. 5, no. 1, pp. 3–18, 2009.

[126] M. Banko and E. Brill, "Scaling to very very large corpora for natural language disambiguation," in *Proc. 39th Annu. Meeting Assoc. Comput. Linguist.*, Toulouse, France, 2001, pp. 26–33.

[127] R. Bar-David and M. Last, "Context-aware location prediction," in *Revised Selected Papers From the 5th International Workshop on Big Data Analytics in the Social and Ubiquitous Context*, vol. 9546. New York, NY, USA: Springer-Verlag, 2016, pp. 165–185.

[128] S. Bassoy, M. J. Jaber, M. A. Imran, and P. Xiao, "Load aware self-organising user-centric dynamic CoMP clustering for 5G networks," *IEEE Access*, vol. 4, pp. 2895–2906, 2016.

Selcuk Bassoy received the B.E degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 1999. He is currently pursuing the Ph.D. degree with the University of Glasgow, researching on 5G radio access network coordination and clustering techniques. Since 1999, he has been working in the wireless telecommunications industry, providing consultancy on radio access network design, planning, and optimization for global telecommunications companies like Ericsson, T-Mobile, Telefonica, and Vodafone. He is currently with Telefonica U.K. Ltd., as a Principal Engineer, specialized in self-organizing networks and other smart automation tools for GSM/UMTS/LTE radio access networks. His research interests include wireless communications, network MIMO, self-optimization, coordinated multi-point, game theory, radio resource management, and load balancing.

Hasan Farooq received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2009, and the M.Sc. degree in information technology from Universiti Teknologi Petronas, Malaysia, in 2014, wherein his research focused on developing ad-hoc routing protocols for smart grids. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Oklahoma, USA. His research area is big data empowered proactive self-organizing cellular networks focusing on intelligent proactive self-optimization and self-healing in HetNets utilizing dexterous combination of machine learning tools, classical optimization techniques, stochastic analysis, and data analytics. He has been involved in multinational QSON project on self-organizing cellular networks and is currently contributing to two NSF funded projects on 5G SON. He was a recipient of Internet Society First Time Fellowship Award toward Internet Engineering Task Force 86th Meeting, USA, 2013.

Muhammad A. Imran (M'03–SM'12) received the M.Sc. (Distinction) and Ph.D. degrees from Imperial College London, U.K., in 2002 and 2007, respectively. He is a Professor of Communication Systems, University of Glasgow, the Vice Dean of Glasgow College UESTC, and the Program Director of Electrical and Electronics with Communications. He is an Adjunct Associate Professor with the University of Oklahoma, USA, and a Visiting Professor with the 5G Innovation Centre, University of Surrey, U.K., where he has worked previously from 2007 to 2016. He has led a number of multimillion-funded international research projects encompassing the areas of energy efficiency, fundamental performance limits, sensor networks, and self-organizing cellular networks. In addition to significant funding from EPSRC, RCUK, Qatar NRF, and EU FP7/H2020, he has received direct industrial funding from leading industries in communications. He also led 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 over 25 successful Ph.D. graduates and published over 300 peer-reviewed research papers. He was a recipient of the Award of excellence in recognition of his academic achievements conferred by the President of Pakistan, the IEEE Comsoc's Fred Ellersick Award 2014, the Sentinel of Science Award 2016, the FEPS Learning and Teaching Award 2014, twice nominated for Tony Jean's Inspirational Teaching Award, shortlisted finalist for the Wharton-QS Stars Awards 2014, the Reimagine Education Awards 2016 for innovative teaching, and the VC's Learning and Teaching Award in University of Surrey. He is a senior fellow of Higher Education Academy, U.K.

Ali Imran (M'15) received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2005, the M.Sc. (with Distinction) degree in mobile and satellite communications, and the Ph.D. degree from the University of Surrey, Guildford, U.K., in 2007 and 2011, respectively. He is an Assistant Professor of Telecommunications with the University of Oklahoma, USA, where he is the Founding Director of Big Data and Artificial Intelligence Enabled Self Organizing Research Center. He has been leading several multinational projects on Self-Organizing Cellular Networks such as QSON, for which he has secured research grants of over $2 million in last four years, as a Lead Principal Investigator. He is currently leading two NSF funded projects on 5G SON amounting to over $750K. He has authored over 60 peer-reviewed articles and presented a number of tutorials at international forums, such as the IEEE International Conference on Communications, the IEEE Wireless Communications and Networking Conference, the European Wireless Conference, and the International Conference on Cognitive Radio Oriented Wireless Networks, on his topics of interest. His research interests include self-organizing networks, radio resource management, and big data analytics. He is an Associate Fellow of the Higher Education Academy, U.K. and a member of the Advisory Board to the Special Technical Community on Big Data of the IEEE Computer Society.