On the Benefits of RAN Virtualisation in C-RAN Based Mobile Networks

Alexander Dawson, Mahesh K. Marina
The University of Edinburgh
Email: alex.dawson@ed.ac.uk, mahesh@ed.ac.uk

Francisco J. Garcia
Agilent Technologies
Email: frankie_garcia@agilent.com

***Abstract*—With ever growing data traffic the traditional mobile network architecture is struggling to cope. Network densification using heterogeneous networks supported by Cloud-RAN is one of the core concepts in terms of physical resources. The system achieves increased capacity by reducing the number of devices (commonly refered to as user equipment - UE) connected to any individual cell. Cloud-RAN decouples the baseband processing from the radio units, allowing the processing power to be pooled at a central location thus reducing the required redundancy. The decoupling also supports innovation in many other RAN technologies by simplifying intercell coordination. While Cloud-RAN differs significantly from traditional base station architectures, interactions with the core network do not reflect these differences. We argue that there is a strong need for an intermediate stage that will reconcile the core network and Cloud-RAN. In this paper we propose a virtual network architecture for Cloud-RAN base stations that will allow us to present the core network with an abstracted view of the physical network. By logically grouping macro cells with collocated small cells we can provide the core network with a simplified overview, reducing signalling overhead. Meanwhile, low latency decisions, such as cell load balancing and interference management, can be made entirely within the Cloud-RAN base station. We present practical applications of the proposed scheme and assess its interoperability with other improvements to the wider infrastructure proposed in related works. The principles presented in this paper lend themselves to evolving key concepts and themes for future 5G networks and beyond.**

I. Introduction

As traffic in mobile networks continues to grow, traditional base station (BS) architectures and coverage schemes are becoming increasingly overwhelmed. The densification of coverage cells eases congestion in the radio access network (RAN) by increasing the number of base stations for user equipment (UE) to connect to [1]. Denser deployments bring new challenges in interference management and intercell coordination that need new approaches to manage. Cloud-RAN (CRAN) and heterogeneous network design has come to the fore as the key architecture concept to provide this for future 5G networks and beyond [2]. In the wider system, changes to the core network (Evolved Packet core or EPC) look to improve the services provided by traditional BSs. Together these technologies present a cost effective approach to increasing the coverage density of the radio access network (RAN). Given the significant changes proposed in each area it is important to ensure that these new technologies are also considered as a whole.

Recent work regarding C-RAN has focussed on its role as a supporting technology for improvements in the RAN with mobility management and intercell coordination pushed out to the edges:

1) Matsuo et al. propose the connection of remote radio heads (RRHs) to multiple BSs at the edge of C-RAN coverage areas to allow all multi-cell operation to be handled within a single C-RAN BS [3].
2) Sundaresan et al. make the case for dynamic pairing of RRHs with virtual BSs to improve resource use in the C-RAN BS [4].
3) Costa-Perez et al. focuses on the virtulisation of the RAN to allow multiple operators to share the same infrastructure [5].

Much of the work discussing the modernisation of the EPC looks to SDN to provide improvements:

1) Gudipati et al. introduce an SDN controller at the network edge to allow traditional BSs to be grouped together emulating a C-RAN BS with RRHs [6].
2) Naudts et al. present an economic analysis of SDN architecture for the EPC [7].
3) Jin et al. present an in-depth design for an SDN based EPC focussed on improving mobility and load balancing core network functions [8].

While C-RAN offers lower costs and easier deployment of these new technologies, the network's overall operation efficiency will continue to be limited if we do not update the signalling with the EPC to reflect the differences between C-RAN and traditional BSs. A formalised interface is required to ensure interaction between C-RAN BSs and the EPC is correctly bounded.

Network heterogeneity and associated improvements are also challenging the traditional view of the cell as the central concept of mobile network design [9]. Relaying, coordinated multipoint (CoMP) and distributed antenna systems (DAS) all rely upon coordination between cells and UE devices not acting in a client capacity (eg. UE relay). 5G networks thus need to look toward a future in which the cell is no longer the core design principle.

In this paper we outline a C-RAN virtualisation architecture. The virtualised interface will present the EPC with an abstracted view of the physical resources associated with a C-RAN BS. The abstraction provides a level of isolation to both systems, allowing improvements to either to be safely implemented without the risk of negatively impacting the other. The EPC retains the required overview of the network infrastructure for routing and other core functionality. Decisions regarding the RAN are made at the network edge to ensure that changes are made with low latency with no signalling to the EPC required.

The authors of [10] touch on the idea of user-centric network architecture. Their proposal centres around a centralised RAN controller similar to that put forward in [6], which groups traditional BSs to move intelligence to the network edge. They then build upon this by redefining mobility using a virtual eNodeB per UE that moves between BSs. While enabling user-centric network design is important to this work our main goal is to abstract the physical resources of the C-RAN BS to enable efficient operation and ensure safe isolation of changes between systems.

After providing some relevant background concepts we present the design of our virtualisation framework. We then proceed with discussion of the practical applications of the proposed architecture and conclude with an assessment of its the interoperability with related works.

II. Background

We present here a primer on the core concepts of modern mobile networks upon which our proposal builds. While many of these are no doubt familiar to the reader it is important to have them in mind moving forward.

A. Heterogeneous Networks

In heterogeneous networks existing macro cell deployments are augmented with smaller cells to provide higher coverage density in high traffic environments [11]. Figure 1 shows a common layout of such a network.

Alongside the advantages offered by network heterogeneity there are of course also disadvantages. Increasing the coverage density and overlapping coverage areas results in an increase in signalling traffic as handovers occur more regularly and cells must coordinate with each other to minimise interference. In a RAN composed of individual BSs this this results in numerous latency issues; such deployments also represent a large investment. It is here that C-RAN offers a solution.

B. Cloud-RAN

C-RAN [12] replaces the self contained BSs at each radio mast with shared processing and distributed radio elements. The core components represented in Figure 2 are:

- **Base Station Pool** - A centralised pool of computing resources to provide the signal processing and coordination functionality required by all cells within the area;
- **Optical Fronthaul** - Optical fibre links carrying digitised representations of the baseband data ready for transmission in the RAN;
- **Remote Radio Heads** - RRHs are light weight radio units and antennae that user equipment connects to via the RAN. The design of the units means that they can be located almost anywhere. Where a traditional base station requires a mast and housing for the baseband processing unit, RRHs need only the space for the antenna and access to some form of fronthaul. RRHs can be used in place of any size of cell from macro down to femto and pico.

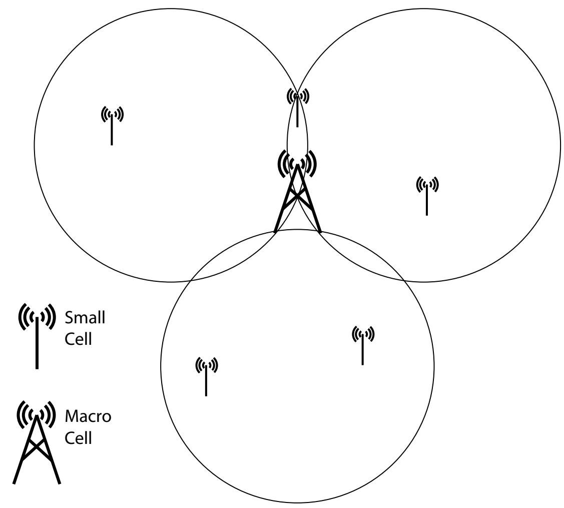


Fig. 1. A heterogeneous network with small cells providing additional capacity within the coverage area of the macro cell.

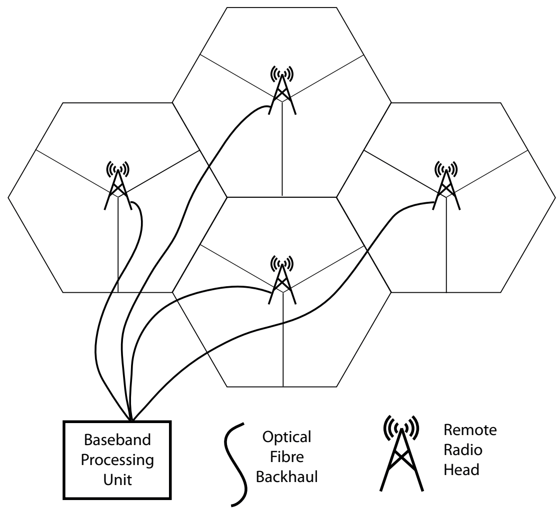


Fig. 2. Cloud-RAN architecture with centralised baseband processing and remote radio heads.

As management of all remote radio heads (RRH) in a given area is handled by a single BS pool, inter cell coordination is made significantly easier as communication occurs directly within the pool. C-RAN principles and the use of lightweight antennas is also seen as a key concept to facilitate massive MIMO technologies in future 5G networks [2].

C. Evolved Packet Core (EPC)

The EPC is the core network behind LTE. Unlike its predecessors it is entirely packet switched with all data sent using IP. The EPC also features an entirely flat architecture consisting of four main elements [13]. The relationship between these

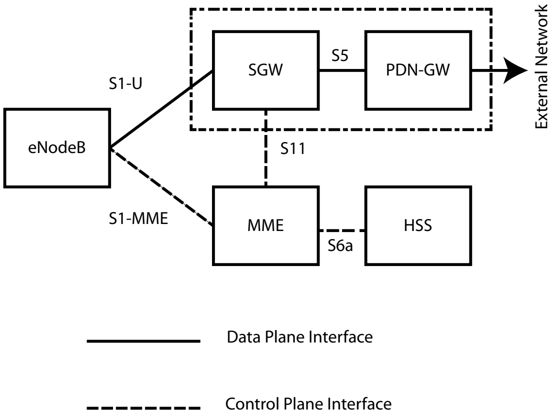


Fig. 3. The existing EPC architecture. The eNodeB here may represent either a traditional or C-RAN BS. Grouping of SGW and PDNGW indicates that they are commonly deployed as a single device.

elements and the interfaces that connect them can be seen in Figure 3 and are listed below:

- **HSS** - The home subscriber server contains all information of relating to users and subscribers, as well as providing support functions for mobility management.
- **MME** - the mobility management entity handles mobility related signalling in the control plane;
- **SGW** - The serving gateway is the dataplane entity that connects the RAN and EPC. In the case of inter BS handover it serves as the mobility anchor;
- **PDNGW** - The packet data network gateway is also located in the dataplane. It connects the EPC to external networks such as the internet and corporate intranets. It is logically linked to the SGW and is commonly physically collocated with it.

The EPC is currently facing a significant challenge in terms of signalling load. Compared to 2G and 3G/HSPA, LTE results in a significantly higher signalling requirement per subscriber increasing by 42% to approximately 120 transactions per busy hour. While a portion of this new signalling is required for new services and new devices types, over 50% of the signalling is related to mobility and paging. This increase is in part due to architectural changes such as heterogeneous networks and greater node density [14].

D. Network Function Virtualisation (NFV) and Network Programming

While not unique to LTE networks NFV is gaining ground in the mobile networking realm, particularly in conjunction with C-RAN [15], [4]. C-RAN baseband processing units offer the greatest efficiency benefits when their processing power is reconfigurable. A significant body of work has gone into virtualising radio access technology so that it can be implemented in software running on generic processing hardware. With this available we can also consider the virtualisation of edge functions of the core network without incurring additional hardware costs.

With virtualised network functions we can also consider the possibility of their re-programmability. The authors of [16] present a system in which an software defined RAN (SD-RAN) controller and a Python SDK provide a framework for accessing network resource information and scheduling transmissions independent of access technology.

Jointly, the ideas presented in this section are enablers for dominant themes in wireless evolution towards 5G networks. Themes such as network densification [1] and moving away from cell-based coverage, resource management and signal processing, facilitating user or device centric 5G architectures [2], [9].

III. Virtualisation Framework

The concepts discussed above and the sizeable body of work that has been published in recent years present a considerable challenge to the EPC. The increases in signalling and more stringent latency requirements for inter cell cooperation are placing pressure on a network that was never designed to support these technologies. To support the continued growth in mobile data traffic we must either reduce the strain on the core network or rebuild it to support these new demands.

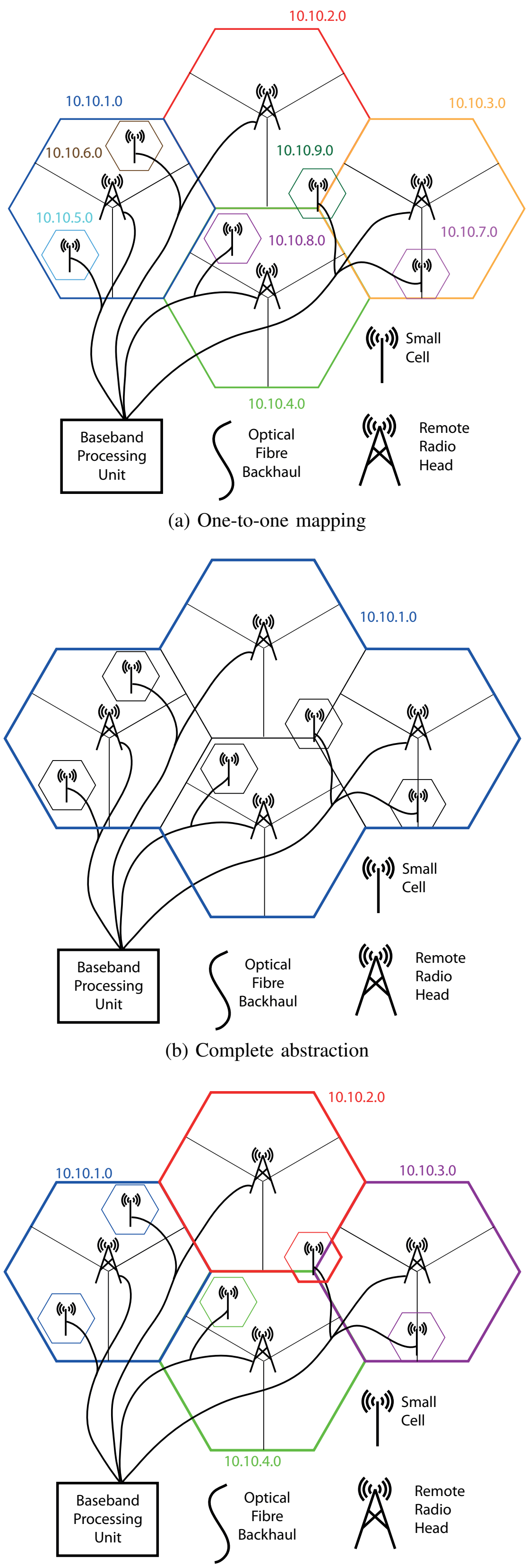
We present here the argument for the first option. By presenting the EPC with an abstracted view of the available infrastructure we can reduce the flow of information to only the required signalling, reclaiming bandwidth for the transmission of user data. We will focus on the use of C-RAN as the basis for this abstraction, however we will later discuss how technologies such as SoftRAN [6] could be built upon to use existing BSs.

A. Overview

Our goal is to isolate the EPC from the RAN to allow innovation to continue in both areas without negatively impacting intercommunication between the two. Many works in the area implicitly require some level of abstraction but none present a formal solution. RAN improvements are heavily reliant on intercommunication between cells for services such as CoMP, DAS and interference management. Such low latency decisions are best made at the BS and require little or no information from the wider network. In contrast, improvements to the EPC are largely RAN agnostic focussing on improving data transport and signalling between BSs and to and from external networks. As the EPC is designed to support a traditional cellular architecture a simplified cell representation is the logical choice of abstraction.

Figure 4a shows how physical resources will be seen by the EPC if traditional cellular network design logic is followed. Physical resources are mapped to directly, with all signalling being passed to the EPC. This presents a significant load if the UE regularly moves between small cells or requires CoMP to provide sufficient coverage at the cell edge.

Figures 4b and 4c show two possible abstraction scenarios that could be presented to the EPC. In (b) the C-RAN BS presents itself to the network as if it were a traditional



(a) One-to-one mapping

(b) Complete abstraction

(c) Macro overview

Fig. 4. Example abstractions of the physical resources

macro cell, albeit one with a larger coverage area. In such a case the EPC would see mobility signalling only when UEs moved between C-RAN BSs. In (c) we present an intermediate solution, the C-RAN BS presents multiple macro cells to the EPC with small cells visible only to the BS. Mobility signalling overhead is reduced as rapid transitions between the macro cell and small cells are handled at the BS but the EPC still maintains overall vision of user mobility.

B. Architecture

The abstraction of physical resources is a well explored topic in many other areas of networking. Virtual local area networks (V-LANs) and network address translation (NAT) have long been accepted features of the modern internet. In mobile networks we face additional challenges as decisions on resource allocation must be based on the physical location of independently mobile UEs.

Figure 5 shows the proposed architecture of the virtualisation framework. On the EPC side of the C-RAN BS we have a VLAN controller that is responsible for grouping cells together as virtual cells and a NAT that will represent the virtual cells to the EPC as a single macro cell. The controller will route data traffic and any required signalling on to the EPC over the existing interfaces.

The core of the new C-RAN BS will perform three main functions. On the dataplane we require a mobility anchor analogous to the SGW in the EPC which provides a static endpoint for communications to and from the UE as it transfers between cells within the virtual cell. Handovers between cells within virtual cells or between virtual cells will be handled by the mobility manager in the BS. As an intermediate stage, this will likely take the form of a virtualised EPC MME. The mobility manager would replicate the existing handover protocols and intercept standard handover signalling from the UE to the MME in the EPC, relaying modified signals for UEs moving between virtual cells. Moving forward we expect to see completely new protocols that focus on user-centric design and completely redefine mobility, with the VLAN allowing the EPC to continue to function without requiring knowledge of these RAN changes.

The final element introduced is an SD-RAN building upon the one proposed by Riggio et al. in [16]. The SDK described provides methods for polling all available resources available at connected cells (termed wireless termination points in the paper - WTP) and associating these resource blocks with a light virtual access point (LVAP) specific to a given UE. These LVAPs provide an ideal data structure scheduling baseband processing in the pool. By processing per LVAP rather than per cell we can ensure that only the minimum required resources are active at any time. Extending the system to allow an LVAP to have multiple resource blocks from different WTPs assigned to it would allow scheduling of CoMP and carrier aggregated transmissions. The SD-RAN controller would also provide slicing of the network resources to allow for RAN sharing between operators.

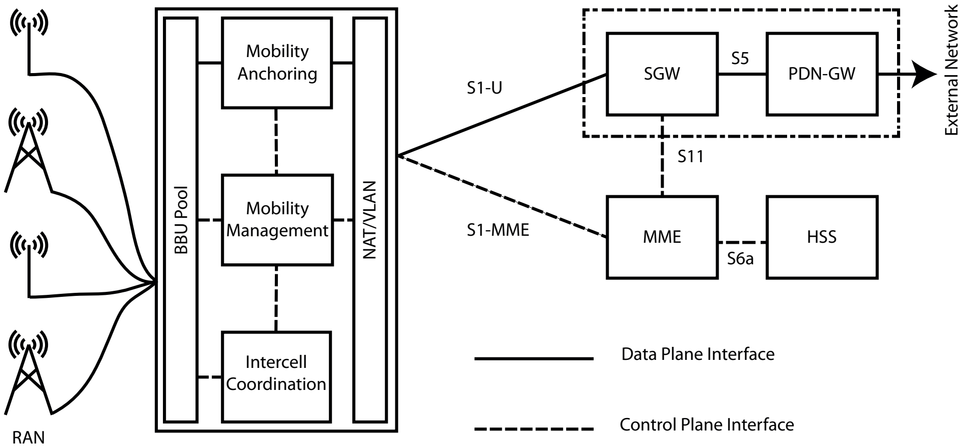


Fig. 5. Visualisation of new architecture elements required in C-RAN BS

C. Benefits

The core benefits of the abstraction layer in terms of system load are seen in the area of mobility and intercell coordination.

The latency benefits of moving mobility to the BS in C-RAN scenarios have been extensively explored [17], however as previously discussed signalling load in the core network is also a major issue. [14] gives us a figure of 31,000 transactions per second per million subscribers in busy hours. Taking only those transactions related to mobility and extracting the messages to the MME, this translates to 290,000 messages per second. As small cells are placed in areas of high traffic, we will assume that handovers between small cells or between small cells and macro cells represent 40%. In the scenario presented in Figure 4 this would translate directly to a 40% decrease in signalling to the MME or 116,000 less messages per second.

In the area of inter cell coordination, the reduction in signalling load will be proportional to coverage area of C-RAN BS as only inter site coordination would require EPC signalling. If we introduce the idea of RRH multi-homing presented in [3], whereby RRHs are connected to multiple C-RAN BSs, we can remove support for intercell cooperation from the EPC completely.

IV. Practical Applications

Although the proposals made here represent a relatively small change to the architecture of the network, formalising the abstraction of resources has a number of practical implications in various aspects of the network's operation.

A. Enabling a User Centric Network Architecture

By representing EPC level mobility management with an abstract view of the physical resources we pave the way for the implementation of user centric networks. In such a network control signalling is provided by the nearest macro cell while uplink and downlink data flows are sent via the best available heterogeneous cells [2], [18]. In our proposed architecture the entire system appears to the EPC as a single macro cell. The advantages of user centric design are improved uplink and downlink data rates as they are decoupled and thus optimization is done separately, and improved system power usage as unneeded small cells can be deactivated as they are data only.

B. Differentiation of Traffic Types

With a virtual network in place on the BS, operators can differentiate between traffic types in their network. Taking the example of machine to machine (M2M) communication we have a very different type of traffic with far lower service level requirements than UE generated traffic. M2M devices are often stationary or if mobile do not require real-time communication and can delay transmission until they have a better connection. An operator could create secondary groupings of cells containing only macro cells as this is sufficient to meet the traffic demand. For stationary devices the C-RAN macro cell proposed in Figure 4b would be sufficient. This configuration would also allow the EPC to identify M2M traffic from IP alone, as the set of virtual cells the addresses refer to will be unique to that traffic type.

In network sharing scenarios we can extend this traffic differentiation to offer separate V-LANs for each operator. Each operator can separately control their own cell virtualisation, potentially implementing different mobility standards. The infrastructure owner can also more readily offer different service tiers depending on what infrastructure the secondary operators require.

V. Related Works

We present here a selection of works that propose improvements to the RAN that we believe can be used in combination with the system set out here.

A. SoftRAN and V-Cell

SoftRAN [6] and V-Cell [10] both approach RAN virtualisation from the direction of traditional base stations. They introduce RAN controllers and control APIs to group existing cells into C-RAN like big base stations. Both APIs offer mapping of resources based on time, frequency and end point as well as monitoring interference between end points and devices. SoftRAN focusses on balancing the loading of cells to improve utility and power usage while V-Cell focusses on smoother mobility between cells. The API used in V-Cell has been published separately [16] and is the one we propose to use in this work.

In both proposals communication is assumed to be between a single UE and BS which is likely to become less and less common as 5G technologies are introduced. V-Cell in particular relies on the instantiation of a virtual cell at a physical base station for mobility. By starting with C-RAN we can more readily support these new multi cell technologies, leveraging the innate intercell communication advantages to ease development. With the API extensions proposed above the big base station concept could be revisited later to provide a basis for bringing complete network virtualisation.

B. FluidNet

FluidNet proposed in [4] is a logically reconfigurable fronthaul for C-RAN deployments. By enabling the reconfiguration of the fronthaul, baseband unit (BBU) resources can be used more efficiently as physical resources can be mapped to the minimum number of baseband units necessary. The abstraction layer presented here will lie between the EPC and the BBUs, with virtual cells representing groupings of BBUs. Without FluidNet these groupings would potentially contain sets of RRHs controlled by multiple separate BBUs. If FluidNet is adopted, energy can be saved by using a single BBU to control multiple RRHs when lower throughput is required.

VI. Conclusions

Current trends in the development of the RAN cannot be supported by the existing core infrastructure. Despite the deployment of C-RAN, high signalling overhead and increasing demands for low latency are still affecting the network as intelligence remains centralised. Other works have shown the potential gains of moving intelligence to the network edge, but no clear options were put forward for redefining the interface with the EPC to account for this movement. We presented in this paper the arguments for abstracting the EPC's view of the physical hardware connected to a C-RAN BS. This abstraction isolates RAN changes and minimises their impact on the wider network. We have outlined the requirements of the virtualisation framework and proposed an architecture for the system.

The virtualisation framework is currently still in the conceptual stage. Moving forward we would like to evaluate the system and assess the projected system benefits in simulation and in a real implementation. We also wish to extend the concepts presented here in combination with the SDN based big BSs presented in SoftRAN and V-CELL.

References

[1] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 82–89, Feb 2014.

[2] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 74–80, Feb 2014.

[3] D. Matsuo, R. Rezagah, G. Tran, K. Sakaguchi, K. Araki, S. Kaneko, N. Miyazaki, S. Konishi, and Y. Kishi, "Shared remote radio head architecture to realize semi-dynamic clustering in comp cellular networks," in *Globecom Workshops, 2012 IEEE*, Dec 2012, pp. 1145–1149.

[4] 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 *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, ser. MobiCom '13. New York, NY, USA: ACM, 2013, pp. 99–110.

[5] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," *Communications Magazine, IEEE*, vol. 51, no. 7, pp. 27–35, Jul 2013.

[6] A. Gudipati, D. Perry, L. E. Li, and S. Katti, "Softran: Software defined radio access network," in *Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking*, ser. HotSDN '13. New York, NY, USA: ACM, 2013, pp. 25–30.

[7] B. Naudts, M. Kind, F. Westphal, S. Verbrugge, D. Colle, and M. Pickavet, "Techno-economic analysis of software defined networking as architecture for the virtualization of a mobile network," in *Software Defined Networking (EWSDN), 2012 European Workshop on*, Oct 2012, pp. 67–72.

[8] X. Jin, L. E. Li, L. Vanbever, and J. Rexford, "Softcell: Scalable and flexible cellular core network architecture," in *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT '13. New York, NY, USA: ACM, 2013, pp. 163–174.

[9] I. Chih-Lin, C. Rowell, S. Han, Z. Xu, G. Li, and Z. Pan, "Toward green and soft: a 5G perspective," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 66–73, Feb 2014.

[10] L. Goratti, R. Fedrizzi, T. Rasheed, R. Riggio, and K. M. Gomez, "V-Cell: Going beyond the cell abstraction in 5G mobile networks," *IEEE / IFIP International Workshop on SDN Management and Orchestration (part of IEEE NOMS 2014 Conference)*, May 2014.

[11] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 10–21, Jun 2011.

[12] C. Kullin and D. Ran, "C-ran the road towards green ran," China Mobile Research Institute, White Paper, Oct 2011.

[13] "ETSI TS 123 002 V11.6.0 (2013-06) Digital cellular telecommunications system (Phase 2+) Universal Mobile Telecommunications System (UMTS); LTE Network architecture," European Telecommunications Standards Institute, Standards Documentation, 2013.

[14] "Liquid net : Signalling is growing 50% faster than data traffic," Nokia Siemens, Marketing Material, 2012.

[15] M. Chiosi et Al., "Network functions virtualisation : An introduction, benefits, enablers, challenges & call for action," in *SDN and OpenFlow World Congress*, Oct 2012.

[16] R. Riggio, M. Marina, and T. Rasheed, "Programming software-defined wireless networks," in *ACM Mobicom (The Annual International Conference on Mobile Computing and Networking)*, Hawaii, September 2014.

[17] L. Liu, F. Yang, R. Wang, Z. Shi, A. Stidwell, and D. Gu, "Analysis of handover performance improvement in cloud-ran architecture," in *Communications and Networking in China (CHINACOM), 2012 7th International ICST Conference on*, Aug 2012, pp. 850–855.

[18] J. Andrews, "Seven ways that hetnets are a cellular paradigm shift," *Communications Magazine, IEEE*, vol. 51, no. 3, pp. 136–144, March 2013.