On the Performance Evaluation of Enabling Architectures for Uplink and Downlink Decoupled Networks

Maria A. Lema, Toktam Mahmoodi, Mischa Dohler,
Centre for Telecommunications Research
Department of Informatics, King's College London
Email:{maria.lema_rosas, toktam.mahmoodi, mischa.dohler}@kcl.ac.uk

***Abstract*—The road to 5G is posing challenging requirements to the cellular network to introduce more applications from several industry verticals. Low delay, high scalability, ultra-reliability and device-centric procedures are some of these requirements. Decoupled Uplink (UL) and Downlink (DL), DUDe, is a key enabler of the device-centric network, and provides a good solution to the UL and DL imbalance problem in heterogeneous networks, improving the UL reliability and load balancing. However, the direct applicability of this technique in 4G networks is subject to either very low backhaul latency between both cooperative base stations, or assisting UL and DL connections that can carry the user plane control signals. This article does a comprehensive study of the enabling architectures for DUDe; the proposed architectures are based on two well-known techniques, Dual Connectivity and Cloud Radio Access Networks. The impact of high latency fronthaul and X2 interfaces is studied and compared to the upper bound UL reliability and throughput obtained with regular round trip time (RTT) values. Results show that even if the radio access network RTT is doubled, DUDe provides an improvement in the UL reliability compared to the classical DL received power cell association.**

***Index Terms*—DUDe, UL Reliability, RAN architecture, C-RAN, Dual Connectivity**

I. Introduction

The architectural evolution of Long Term Evolution - Advanced (LTE-A) towards a 5G technology requires a solution that enables user (or device) centric technologies that lets the system improve capacity, latency, reliability and scalability. One of the key issues when moving towards a user or service driven network is to provide the mobile network with sufficient flexibility to select the the serving cell that better suits the device or service requirements. In this sense, Downlink (DL) and Uplink (UL) decoupling (DUDe) goes one step further, and allows the user to independently transmit and receive to and from different evolved Node Bs (eNBs). Essentially, DUDe breaks the hard and classical constraint of cell selection based on downlink received power, and provides the system with flexibility to associate users to different eNBs in the DL and in the UL.

The third generation partnership project (3GPP) organization has introduced the concept of Dual Connectivity (DC) in Heterogeneous Networks (HetNets) in Release 12 [1], defined as the simultaneous use of radio resources from two eNBs connected via non-ideal backhaul link over the X2 interface. In particular, DUDe is recognised as a solution to reduce the UL and DL imbalance problem caused by the eNB transmission power disparities in the context of HetNets. A UE is said to be in an imbalance situation if the best UL cell and the best DL cell are different based on received power metrics. This topic has generated interest lately in the research community and several works show how effective this technique is to improve the UL capacity, reliability and load balancing [2], [3], [4], from both simulation and analytical perspectives.

DUDe introduces new challenges in the architecture due to the tight delay requirements in the access network, in the main related to the hybrid automatic repeat request (HARQ) process. Ideally, when a UE is transmitting in a complete decoupled mode, holding parallel UL and DL connections with the corresponding eNBs can contribute negatively to the decoupled performance. Holding more than one UL connection is less power efficient for users that are placed near the cell edge, which are the ones more likely to decouple [5], [6]. Having no User-Plane (U-Plane) control information being signaled back to the target eNB through the user interface, the delivery of Layer 1 and Layer 2 (L1 and L2) signaling and radio resource control (RRC) relies on the X2 interface between both serving cells. The 3GPP has proposed several architectural alternatives for DL DC in [1], where it is also highlighted that the architecture needed to support the U-Plane aggregation from different eNBs is expected to be very similar to those proposed for DC, based on the bearer split concept. In this context, work in [7] studies the performance of DUDe considering a bearer split architecture, where the split is done at the packet data convergence protocol (PDCP) layer. However, no specific architecture is proposed to support the UL and DL split with no UL or DL assisting connections.

The main contribution of this work is to study and evaluate possible architectural solutions that enable the use of DUDe. With the aim of fully decoupling UL and DL L1/L2, two architectural schemes are studied. The first, follows the LTE-A architecture of distributed cells and it is based on the bearer split concept presented in [1]; this work proposes some modifications in order to support L2 retransmissions, not considered in [1]. The second architecture that supports DUDe is based on the eNB function centralization in the cloud radio access network (C-RAN), where shared base band processing units (BBUs) reduce the challenge of cooperation among eNBs. The performance of both architectures is studied and compared by means of realistic system level simulations. In particular, different latency scenarios for the X2 and fronthaul interfaces are considered to evaluate the impact over the UL reliability and throughput. Both architectures are compared to a solution that considers assisting UL and DL connections and to the upper bound DUDe performance.

This document is organized as follows, next section is devoted to explain the RAN architecture schemes that can support the use of DUDe, section III explains the simulation conditions for the performance evaluation, and section IV shows the results obtained with the system level simulations. Finally, the paper is concluded in section V.

II. RAN Architecture Discussion

Those architectures that enable a full UL and DL decoupling should support a feasible cooperation among both serving cells while not jeopardising the improvements in the UL in terms of reliability and capacity. To achieve this it is necessary to assure the delivery of L1 and L2 control signals while maintaining the RAN latency requirements.

A. RAN Control Signals

Both UL and DL control signals are crucial to support the U-Plane data transmission. The inclusion of DUDe is challenging in terms of control signalling handling, since delay in transmission and processing can reverse the potential improvements. Apart from L1 and L2 control signals, RRC messages that carry information on the number of resources that are consumed for control signalling need to be exchanged; the number of resources used dynamically change depending on the cell traffic type and number of UEs being served [8].

In the RAN, control signals that are of paramount importance to handle scheduling and medium access control (MAC) layer procedures are:

- Uplink control information carried in the physical uplink control channel (PUCCH) in charge of transmitting the DL channel quality indicator (CQI), buffer status reports, scheduling requests and power headroom reports.
- Downlink control information (DCI) in charge of indicating, among others, both UL and DL physical resource blocks for transmission (UL-SCH and DL-SCH), as well as link adaptation forms and transmit power for the uplink. DCI is carried in the physical downlink control channel (PDDCH)
- Downlink HARQ acknowledgment messages carried in the Physical Hybrid-ARQ Indicator Channel (PHICH).
- RRC messages that configure the UE connection and release, as well as the PUCCH position and resources and sounding reference signals (SRSs) configuration.

From these, the most stringent ones in terms of latency are the HARQ process messages, which affect directly in the UL Round Trip Time (RTT). The HARQ in the UL follows a synchronous process, with a periodicity of 8 sub-frames so the eNB knows exactly which HARQ process comes at each sub-frame; there is no explicit information being forwarded about the process ID. This means there is a strict relation between the HARQ process identification to the sub-frame, and if a delay occurs in the HARQ transport, the next suitable sub-frame needs to be awaited.

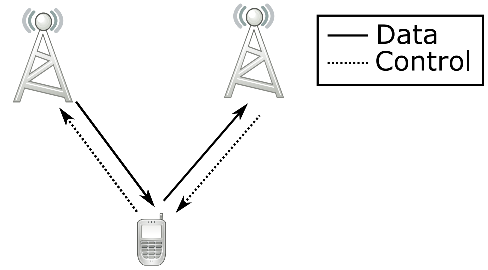

Fig. 1: Example of DUDe configuration with assisting connections

B. DUDe with Assisting Connections

The most straightforward way of implementing DUDe is to assume supporting UL and DL connections to transmit the control signals. The terminal is connected to both cells and aggregates the data flows. Also, L1/L2 signalling can be handled locally. Architecture alternatives in this line where presented in [9].

A simple way to support this is with the use of inter-site carrier aggregation (CA), where each carrier component is configured to carry a shared and a control channel. This configuration allows to keep the RTT at desired levels since no further delay is introduced. However, potential disadvantages of this configuration are: first, power limited UEs may struggle to hold two UL connections since power availability in the cell edge is lower, and second, this configuration does not maximize capacity over the available spectrum, since one component carrier is exclusively used to handle control information. Figure 1 shows the inter-site CA configuration.

C. User Plane Bearer Split for DUDe

Latest releases of LTE-A (Release 12 and 13) consider new architecture alternatives for DC [1]. One of the new advances is the introduction of the bearer split concept, which facilitates the UE having two or more simultaneous transmissions in different eNBs, known as Master eNB and Secondary eNB, MeNB and SeNB respectively. Several alternatives have been proposed for the DL bearer split and some studies argue the UL bearer split feasibility in terms of power consumption [5]. Having two simultaneous UL connections may lead to a degradation of the UL performance in terms of UE energy efficiency. However, from a received power perspective the UL traffic is preferred to be directed to the eNB which suffers the smallest path-loss. Potential complexity is associated to this, since packet data units (PDUs) need to be forwarded to/from the MeNB and SeNB, according to the bearer split architecture. Similarly, in [10] several options are proposed and compared for the UL while assuming DL bearer split. Conclusions highlight that UL bearer split should not be

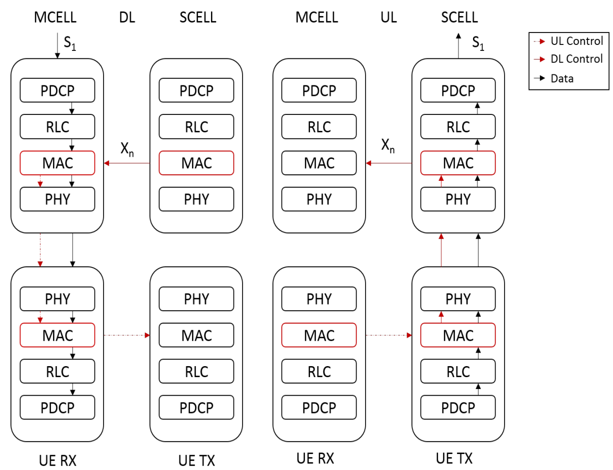

Fig. 2: Protocol stack information flow for radio bearer managed at SGW

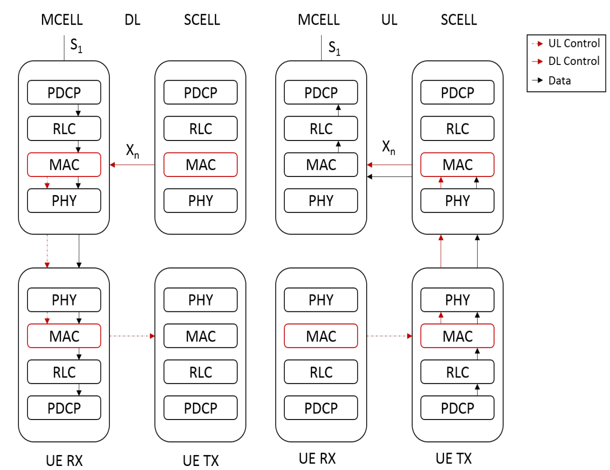

Fig. 3: Protocol stack information flow for radio bearer managed at MeNB

supported and UL data should be either transmitted directly to the MeNB, or forwarded to the MeNB by the SeNB. Based on these limitations, in a DUDe context, assisting PUSCH or PUCCH connections may not be carried out in the MeNB; and PDSCH and PDCCH connections may not be carried out in the SeNB, and therefore the RAN control signals need to be forwarded through the non ideal backhaul connection.

In light of this, there are two possible architecture alternatives that suits best the DUDe transmission based on the DC solutions:

1) The Radio bearer is managed at the serving gateway (SGW), and the DL flows though the S1 from the SGW to the MeNB and the UL flows through the S1 to the SGW from SeNB. This option has reduced flow control among both serving eNBs, since only part of the control signals, for example HARQ acknowledgements (ACKs), needs to be forwarded through the X2. Figure 2 shows a diagram of the information flow for DL and UL.
2) The Radio bearer is managed at the MeNB, so the configuration is the master/slave MAC for UL and DL control feedback. This means that real-time MAC PDUs need to be forwarded to the corresponding eNB via the X2 backhaul interface, while respecting the 8 ms HARQ round trip time (RTT) requirement. Master/slave configuration is for UL and DL, and the processing of each MAC PDU is done on the corresponding cell. Figure 3 shows a diagram of the information flow for DL and UL.

Current heterogeneous networks architecture pose a big challenge to accomplish this, since delays that range from 5ms to 30ms are expected in the X2 interface [1]; in this sense in the best case the UL RTT is doubled, and when latencies reach 20 ms the RTT quadruplicated. This is due to the strict relationship between the HARQ process id with respect to the sub-frame number, and the user is forced to interleave the HARQ process.

D. Alternative Solution: Use of C-RAN

The architecture proposals to allow for DC with distributed eNBs has been shown to support DUDe without considering assisting UL and DL connections, with the counterpart of having an increased RTT in the UL access, essentially caused by the delay in ACK forwarding from one serving cell to another. Also, considering assisting connections with the use of CA may impair the capacity maximization. From the RRC connection perspective, the previous architecture is challenging in terms of complexity for the UE side, since parallel RRC procedures need to be handled [11].

Based on this, we can draw the hypothesis that, if both eNBs, MeNB and SeNB, are able to share the same base band processing unit (BBU), then complete DUDe can be handled. This solution can mitigate most of the DC architecture drawbacks, and treat both UE connections as one. The concept under the BBU sharing is the C-RAN based architecture. C-RAN breaks the static relationship between BBU and remote radio head (RRH), and each RRH does not belong to any specific BBU. In particular, the radio signals from/to a RRH can be processed by a centralized eNB, which supports real-time cooperation among them. Given this, virtualization technology (network function virtualization, NFV) will maximize the flexibility in C-RAN [12]. Figure 4 [1] shows the C-RAN architecture for DUDe.

One of the most important issues when working with C-RAN is the potential increase in latency brought by the fronthaul interface. Works in [13], [14], [15] study the different layer split options in terms of delays and capacities of the common public radio interface (CPRI); it is shown that there is a benefit in locating a portion of the base station signal processing functions near the RRH, the bandwidth and latency requirements are brought to a level that can be fulfilled by cost-effective transport networks and at the same time enable the

[1]EPC: Evolved Packet Core; S1: Interface between RAN and EPC

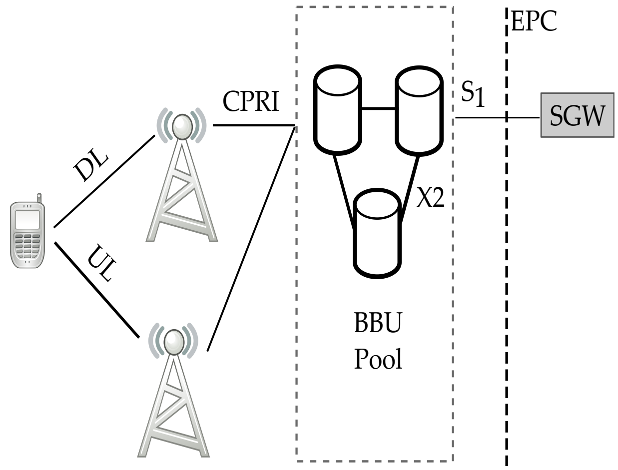

Fig. 4: C-RAN Architecture with DUDe

possibility of having enhanced inter-node cooperative radio resource management procedures. Studies agree that high-latency fronthaul will introduce delays that can range from 250 μs to 4 ms, which may cause the UL RTT to double in the worst case [13].

III. Simulation Conditions

The performance of the discussed architectures is evaluated by means of system level simulations. The following cases have been implemented and compared:

- Downlink received power (DLRP) association. CA is considered and UL transmissions are done in two component carriers based on the UE power availability, following the guidelines in [16].
- DUDe with 8 ms RTT latency and all frequency resources are available for U-Plane data transmission with the use of CA, this case is the baseline for DUDe analysis since shows the upper bound performance.
- DUDe with assisting connections with the use of CA. This case uses one carrier exclusively to forward the control data to the corresponding eNB; hence, only one carrier is used for U-Plane data transmission.
- DUDe with C-RAN with double RTT. Both carriers are used for U-Plane data transmission.
- DUDe with bearer split and high latency on the X2 interface. Both carriers are used for U-Plane data transmission.

Channel state information (CSI) acquisition is modeled with the use of SRS which are sent in the last single carrier frequency division multiple access (SC-FDMA) symbol, which is specifically reserved for this purpose. UEs are code and frequency division multiplexed to be able to have more sounding signals. The interval between two consecutive SRS reports depends on the cell load, the more UEs are connected to the eNB the less spectrum is sounded on each TTI to allow more users to be part of the sounding process.

The scenario is a 3GPP based, urban macro-case that follows the guidelines in [17], [18] where small cells are

TABLE I: Parameters common to all studies

Parameter	Value
Carrier frequency	2 GHz
Bandwidth	2x20 MHz
Power delay profile	Extended pedestrian B
Doppler model	Young and Beaulieu [20] 3 km/h
Shadowing correlation distance	Macro: 50 m Small: 25 m
Shadowing deviation	8 dB, 10 dB [21]
Target BLER	10 %
SRS periodicity	2 TTI
SRS information expiration	10 TTI
Maximum UE transmission power	23 dBm
Termal noise power(σ^2)	-174 dBm/Hz
Distance dependent path-loss Macro	128.1+37.6log(d)[2]
Distance dependent path-loss Small	140.7+36.7log(d)[2]

[2] d: distance in km

located in certain hotspot areas of the scenario. The inter-site distance (ISD) is considered to be 500 m between macro-eNBs. Realistic long and short term fading is considered. Spatially correlated log-normal variations are introduced, based on the two dimensional correlated shadowing model presented in [19]. An extended pedestrian B power delay profile is implemented considering a UE speed of 3 km/h based on the guidelines of [20].

The simulation tool has been fully designed following guidelines in [21] and has been calibrated with the 3GPP performance curves in [21]. The simulator is dynamic, which means that the system is evaluated during a certain observation time and with a time resolution of 1 transmission time interval (TTI). Users are scattered in the simulation area following both random and hotspot distributions. Unless specified otherwise, the wireless access network is considered to have a RTT of 8 ms in the uplink, considering processing times in both eNB and UE. File transfer protocol (FTP) communication is assumed following the model presented in [21] and, as soon as the buffer is entirely transmitted, the UE is automatically reconnected in another position, as a new user. This keeps a constant number of interference sources during the simulation time. The scheduler is proportional fair based and the link to system level abstraction follows the guidelines in [22]. General specifications of the scenario are detailed in table I.

IV. Performance Evaluation

UL reliability and throughput is highly improved when users are allowed to decouple the UL to the SeNB. The improvements mainly come due to the increased received power at the eNB and also the more even distribution of users among the different cells in the scenario. Figure 5[2] shows the upper bound improvement in the block error rate (BLER) with respect to the DLRP UL association; there is a 20% improvement of the BLER at the first attempt, which represents the UEs that are successful in the first transmission

[2]BLER is measured as the percentage of successful transmissions over the UE connection time

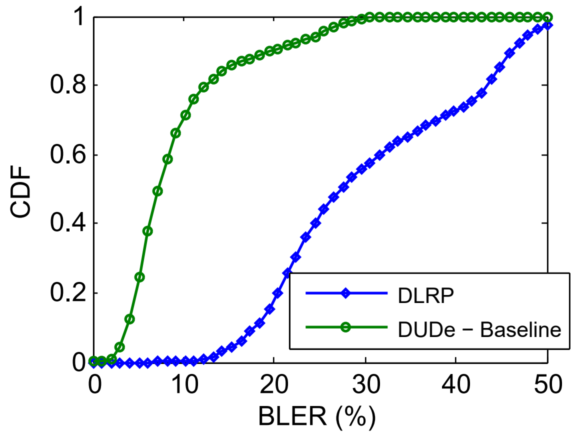

Fig. 5: BLER at first attempt improvement with DUDe

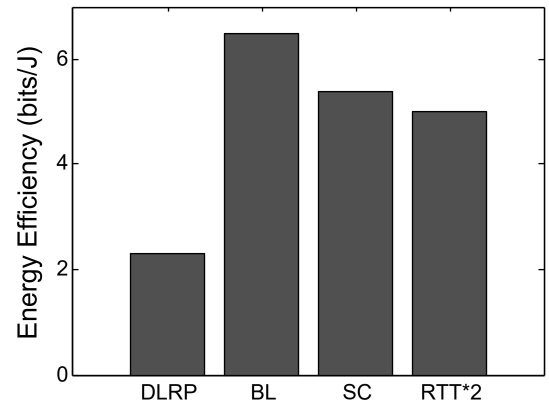

Fig. 7: Energy efficiency performance for the different cases

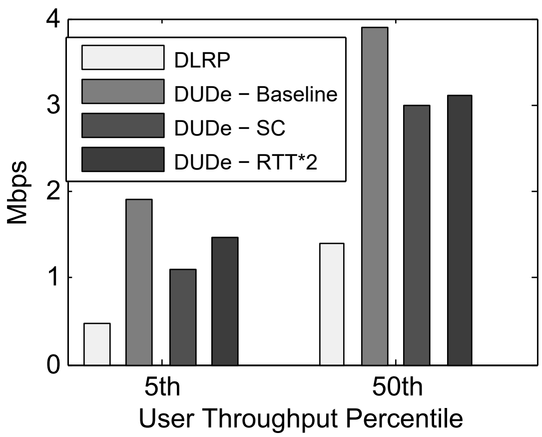

Fig. 6: UE UL Throughput performance for the different cases

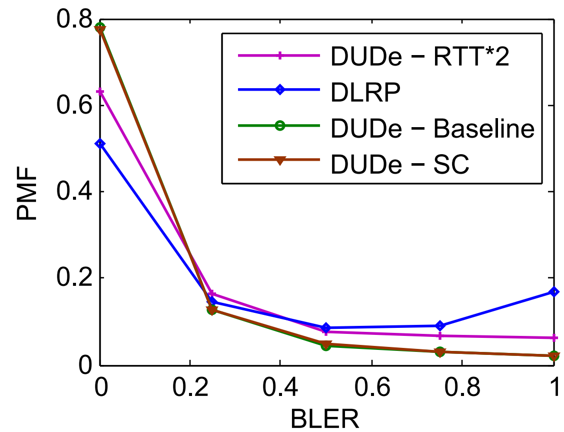

Fig. 8: Probability mass function of instantaneous BLER at first attempt

without the need of any HARQ retransmission. This leads to a subsequent improvement in the UL throughput as well.

When considering a realistic architecture that enables the use of this technology, the upper bound performance may be impaired. Figure 6 compares the throughput performance of the different architectural strategies with respect to the DLRP association and the DUDe baseline performance. In both architectures, DC with bearer split or C-RAN, a delay in the control flow can cause the UL RTT to double, given the stringent requirements of the synchronous HARQ process. On the other hand, if assisting connections are carried out and one component carrier is exclusively used to forward the control signals, and the other carrier is used for U-Plane data, the RTT duration remains equal (i.e., 8 ms) and the price to pay is less available spectrum for U-Plane allocations. To compensate the loss in throughput, non-power limited UEs can transmit an increased bandwidth, raising the total transmit power; as a consequence, the total energy efficiency drops as less bits are sent per power unit, as shown in figure 7.

Regarding the UL reliability, when the RTT is doubled the BLER increases. This is because the CSI available in the reception instant, 16 ms later, is more likely to be outdated. The delay in the X2 or fronthaul interfaces increase the misalignment between the signal to interference noise ratio (SINR) measured from the sounding signals and the actual SINR experienced in the reception instant. In spite of this, the performance of a higher RTT is still better than with DLRP association; figure 8 shows the probability mass function of the instantaneous BLER at the first transmission attempt derived from the link to system level abstraction model.

As remarked in the architecture discussion, the X2 delay may range from 5 to 30 ms. In this sense, the best case for the DC architecture is to double the RTT, whereas in the C-RAN solution, the CPRI delay may range from 250 μs to 4 ms, meaning that in the worst case the RTT is doubled. Figure 9 represents again the probability mass function of the instantaneous BLER, now comparing the performance of a higher RTT, three times higher, to the DUDe baseline case and the DLRP association case. As the RTT increases the performance of DUDe gets closer to the DLRP.

C-RAN worst case (i.e., double RTT) is close to the

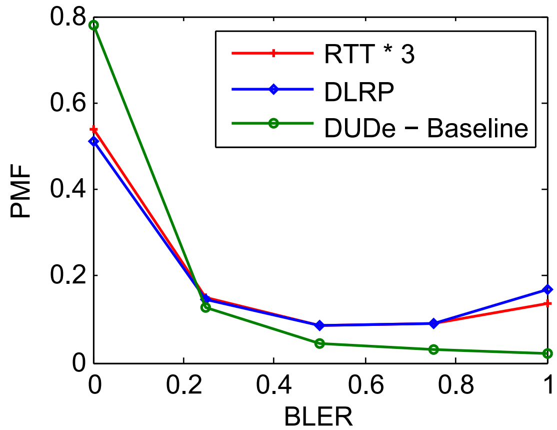

Fig. 9: Probability mass function of instantaneous BLER at first attempt

CA solution with assisting connections in terms of energy efficiency and throughput. Is worth highlighting that the probability of this worst case event to happen has a huge impact on the system performance. If low latency fronthaul link is considered (i.e., 250 μs) the performance of C-RAN, in terms of BLER, energy efficiency and throughput, is closer to the upper bound DUDe, when all carriers are used for U-Plane data transmission and RTT is maintained in 8 ms.

V. Conclusions

This paper has proposed and studied architectural solutions that can enable the use of DUDe in heterogeneous networks. The proposed schemes are based on the use of CA for assisting UL and DL connections, UL control flow sharing at a MAC level using DC architecture and a C-RAN solution where RRH can share the same BBU. All three solutions are evaluated and compared in terms of UL reliability, energy efficiency and throughput.

Whilst the CA with assisting connections lets maintain the BLER improvements brought by DUDe as there is no potential increase in the RTT, it entails a hard constraint in the use of the frequency resources, as one component carrier is exclusively used for control feedback. In this sense, the C-RAN solution worst case, with double RTT, performs equal than the assisting connections scheme in terms of throughput and energy efficiency. Nonetheless, when C-RAN fronthaul latency is the smallest, the performance of this architectural scheme equals the upper bound DUDe, since all carriers are used for U-Plane data transmission and there is no increase on the UL RTT. DUDe supported by DC architecture with control flow through the X2 interface considering delays that range from 5-10ms shows the worst performance.

Acknowledgment

This work is supported by the Ericsson 5G and Tactile Internet industry grant to King's College London.

References

[1] 3GPP, "Study on Small Cell Enhancements for E-UTRA and E-UTRAN; Higher Layer Aspects," 3rd Generation Partnership Project (3GPP), TR 36.842, Sep. 2014. [Online]. Available: http://www.3gpp.org/dynareport/36842.htm

[2] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and Uplink Decoupling: A Disruptive Architectural Design for 5G Networks," in *Global Communications Conference (GLOBECOM), 2014 IEEE*, Dec 2014, pp. 1798–1803.

[3] ——, "Load & Backhaul Aware Decoupled Downlink/Uplink access in 5G Systems," in *Communications (ICC), 2015 IEEE International Conference on*, June 2015, pp. 5380–5385.

[4] S. Singh, X. Zang, and J. Andrews, "Joint Rate and SINR Coverage Analysis for Decoupled Uplink-Downlink Biased Cell Associations in HetNets," 2014, Available online at http://arxiv.org/abs/1412.1898.

[5] Huawei, "Handling of UL Traffic of a DL Split Bearer," 3GPP TSG-RAN, Tech. Rep. R2-140054, 2014.

[6] K. Smiljkovikj, P. Popovski, and L. Gavrilovska, "Analysis of the Decoupled Access for Downlink and Uplink in Wireless Heterogeneous Networks," 2014, Available online at http://arxiv.org/abs/1407.0536.

[7] A. Ratilainen and S. Wager, "Protocol Performance of UL/DL Separation in LTE Heterogeneous Networks," in *The Twelfth International Symposium on Wireless Communication Systems 2015*, August 2015.

[8] 3GPP, "LTE Radio Access Network (RAN) Enhancements for Diverse Data Applications," 3rd Generation Partnership Project (3GPP), TR 36.822, Sep. 2012. [Online]. Available: http://www.3gpp.org/dynareport/36822.htm

[9] S. Ericsson, "Further Discussions on UL/DL Split," 3GPP TSG-RAN, Tech. Rep. R2-131678, May 2013.

[10] CATT, "Analysis on UL bearer split," 3GPP TSG-RAN, Tech. Rep. R2-140181, 2014.

[11] ZTE, "Comparison of CP Solution C1 and C2," 3GPP TSG-RAN, Tech. Rep. R2-132383, 2013.

[12] M. Arslan, K. Sundaresan, and S. Rangarajan, "Software-Defined Networking in Cellular Radio Access Networks: Potential and Challenges," *Communications Magazine, IEEE*, vol. 53, no. 1, pp. 150–156, January 2015.

[13] NGMN, "Further Study on Critical C-RAN Technologies," NGMN, Deliverable, 2015.

[14] U. Dtsch, M. Doll, H.-P. Mayer, F. Schaich, J. Segel, and P. Sehier, "Quantitative analysis of split base station processing and determination of advantageous architectures for lte," *Bell Labs Technical Journal*, vol. 18, no. 1, pp. 105–128, 2013. [Online]. Available: http://dx.doi.org/10.1002/bltj.21595

[15] A. Maeder, M. Lalam, A. De Domenico, E. Pateromichelakis, D. Wubben, J. Bartelt, R. Fritzsche, and P. Rost, "Towards a flexible functional split for cloud-ran networks," in *Networks and Communications (EuCNC), 2014 European Conference on*, June 2014, pp. 1–5.

[16] M. A. Lema, M. Garcia-Lozano, S. Ruiz, and D. G. Gonzalez, "Improved Component Carrier Selection Considering MPR Information for LTE-A Uplink Systems," in *2013 IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, Sept 2013, pp. 2191–2196.

[17] 3GPP, "Physical Layer Procedures," 3rd Generation Partnership Project (3GPP), TS 36.213, Sep. 2009. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36213.htm

[18] ——, "User Equipment (UE) Radio Transmission and Reception (Release 11)," 3rd Generation Partnership Project (3GPP), TS 36.101, Sep. 2012. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36101.htm

[19] R. Fraile, O. Lazaro, and N. Cardona, "Two Dimensional Shadowing Model," COST 273, TR available as TD(03)171, 2003.

[20] D. Young and N. Beaulieu, "The Generation of Correlated Rayleigh Random Variates by Inverse Discrete Fourier Transform," *IEEE Transactions on Communications*, vol. 48, no. 7, pp. 1114 –1127, jul 2000.

[21] 3GPP, "Further Advancements for E-UTRA; Physical Layer Aspects," 3rd Generation Partnership Project (3GPP), TS 36.814, Mar. 2010. [Online]. Available: http://www.3gpp.org/ftp/Specs/html-info/36814.htm

[22] J. Olmos, M. Ruiz, S.and Garcia, and D. Martin, "Link Abstraction Models Based on Mutual Information for LTE Downlink," Cost 2100, Tech. Rep. available as TD(10)11052, 2010.