
Chapter 8

Decoupled Uplink and Downlink Access in Heterogeneous Networks

Hisham Elshaer, Maria A. Lema, Toktam Mahmoodi, and Mischa Dohler

8.1 Introduction

Data centric UEs are currently not only demanding more capacity from wireless networks, but also service oriented Quality of Service (QoS) and comparable Quality of Experience (QoE) in both Downlink (DL) and Uplink (UL). With the introduction of internet of things, machine to machine communications, cloud services and the widely-used social media, UEs and devices in general are increasingly more content generators than they were before. As a consequence, the new generations of mobile communications must devise strategies that improve the device's overall experience. Adding intelligence into the Media Access Control (MAC) and Radio Resource Control (RRC) layers has shown significant improvements in terms of data rate and fairness. However, the non-constant QoE of UEs along the cell area motivates to increase the number of serving eNBs by adding cells of shorter coverage, bringing the network closer to the UE. This change in the deployment allows to further improve the network capacity by enabling better load balancing among the cells and eliminating coverage holes. However, this new paradigm of system design brings with it significant open challenges to ensure a correct operation such as: backhaul improvement, mobility and interference management, cell association and UL/DL relationship.

One of the key issues when moving towards a user or service centric network is to provide the mobile network with sufficient flexibility to select the serving cell that better suits the device or service requirements. To this direction, Downlink and Uplink Decoupling (DUDe) goes one step further, and allows the user to independently transmit and receive to and from different base stations. Essentially, DUDe breaks the hard and classical constraint of cell selection based on downlink received power, and provides the network with the flexibility to associate users to different eNBs in the DL and UL.

This chapter describes in detail the DUDe technique. First, Section 8.2 goes through the main challenges HetNets must face in terms of radio planning and interference management, and reviews the potential solutions to address these problems.

Thereupon, decoupling UL and DL is recognised as a new technique that can effectively improve the HetNet performance, as shown in Section 8.4, and also can impact positively the performance of other radio access technologies, such as Carrier Aggregation (CA), Coordinated Multipoint transmission and reception (CoMP) or millimeter wave (Section 8.5). The enabling radio access network architectures are discussed in Section 8.6 and higher layer opportunities and challenges are addressed in Section 8.7.

8.2 HetNets Challenges in 4G Networks

LTE-A has to face a variable QoE within cells because of the difference in spectral efficiency depending on the UE position. Adding spectrum or improving the link adaptation provides faster connectivity, however, no homogeneous performance is actually met. One of the challenges 4G and future technologies, such as the forthcoming 5G, must meet are the new requirements of area spectral efficiency and user rate distribution. Further improvements to achieve a fair per-user performance along the cell are possible by increasing the eNB deployment density. Nowadays, mobile networks are shifting from a single-tier homogeneous network approach to a multi-tier heterogeneous one, the so-called HetNets. It has become a popular approach in the past few years as an efficient and scalable solution to improve the network capacity in hot-spots; it is also a viable solution to improve fairness, since the network gets closer to the UE.

A HetNet is a network deployment consisting of Macro Cells or MCells and overlaid low-power nodes such as pico-cells, femto-cells, RRHs or relays, referred to more generally as Small cells or SCells. The MCells are high power eNBs typically located along the geographical area with purposes like coverage maximization or interference reduction. SCells are in charge of eliminating coverage holes or improving capacity by off-loading the MCells at hot-spot areas. Figure 8.1 shows a typical HetNet deployment. Based on the frequency deployment, two solutions arise: co-channel deployment, where both MCells and SCells share the same frequency band; and dedicated deployment, where each type of cells transmits at different frequencies. While the former allows to maximize the spectrum utilization and capacity, interference can become a major concern in the system design. On the other hand, dedicated deployments may solve the interference problem, at the expense of over provisioning SCells with frequency resources. In this sense, this new paradigm of network deployment potentially increases the spectral efficiency; however, significant challenges are introduced to ensure the correct operation of HetNets.

8.2.1 *Radio Planning Challenges in HetNets*

The trend of cell hyperdensification with multiple tiers require different approaches in network planning and design to meet the fundamental objectives of maximizing user rate and empowering fair per-user spectral efficiency, among others.

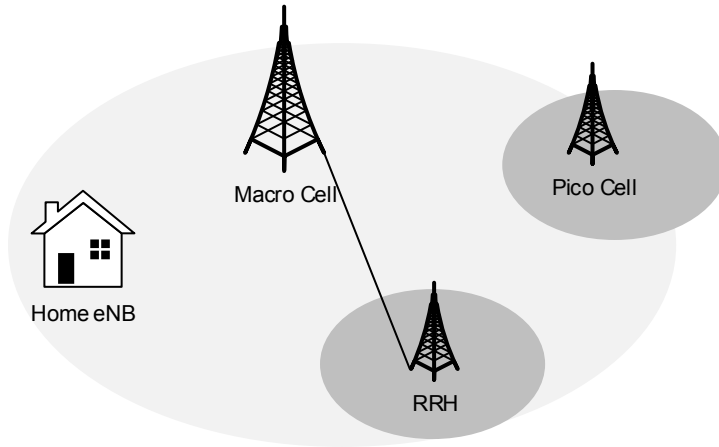


Figure 8.1: Example of HetNet deployment

An obvious challenge in the HetNets network planning is the frequency deployment: if maximizing capacity is the main driver, or access to spectrum is limited, MCell frequencies should be reused by the other tiers. In co-channel deployments the inter-cell interference is an important constraint. In both UL and DL, there is a higher number of interfering nodes compared to classical MCell deployments; a large number of SCells reusing the same carriers may generate high levels of interference. In such cases, sophisticated radio resource management procedures shall be considered to minimize the interference generated. On the other hand, dedicated deployments are attractive in large bandwidth availabilities, which will also allow to better control the interference.

By definition SCells are low power nodes. Very low DL transmit power may result in low SCell coverage, providing very little user migration from the MCell. On the other hand, higher transmit powers increase the SCell size, which may increase the SCell load. The DL transmit power disparities among the different nodes and cell selection based on the DL received power result in imbalance problems between both UL and DL connections. The DL/UL imbalance problem has been recognised by 3GPP in [4, 15]. A UE is said to be in this situation if the best serving cell from an UL received power perspective and the best DL serving cell are different. The UE optimal connection in the UL is to the SCell as it is closer, but in the DL it would be connected to the MCell. Classical cell association rules have a huge impact in the load imbalance and the UL performance. Figure 8.2 shows a graphical example of the HetNet imbalance problem.

SCells need to be provided with energy efficient and low cost backhaul, and this has quite often been proven tricky in SCells deployments. The lack of high capacity backhaul can limit the SCell availability, which may impair the potential improvements brought by cell densification. Hence, cell association procedures should consider the availability of such resources which provide the user with sufficient quality of experience.

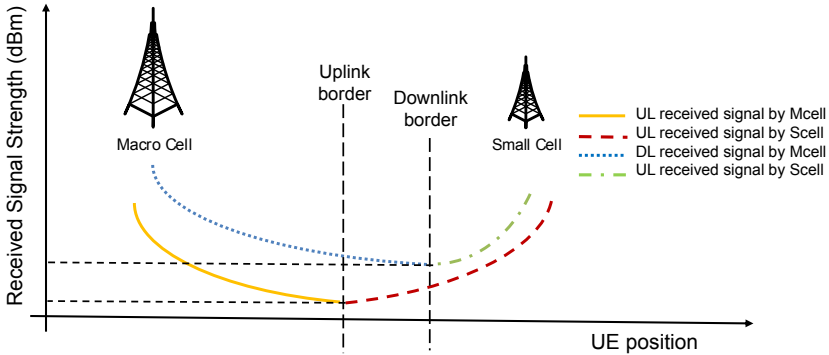


Figure 8.2: Example of HetNet imbalance problem

8.2.2 Strategies for Improvements in HetNets

Cell selection based on the reference signal received power (RSRP) result in imbalance problems, since the DL coverage of the MCell is much larger than that of the SCell. One strategy that can address this problem and bring some fairness to the UL is the cell Range Extension (RE), which allows the UEs to associate to cells that do not provide the highest DL RSRP, increasing the SCell coverage region. The SCell range is expanded by adding a cell selection offset to the RSRP measured from the SCell, and the UEs in the range expanded area are associated to the SCell in both UL and DL. In this situation, the UL of UEs is fairly improved, while the DL is associated in a suboptimal way, since the maximum received power is still provided by the MCell. Although expanding the range of low power nodes partially compensates UL interference issues, DL interference on RE UEs is substantially increased. Studies in [20, 10] have shown that using high offsets (greater than 3-6 dB) increases the DL interference levels. For this reason, eICIC mechanisms have also been investigated in this context. These techniques are mostly variations of the same idea: frame muting and coordinated scheduling. Indeed, 3GPP introduced the possibility of using almost blank subframes (ABS) since LTE-A Release 10.

Eventually, the range extension technique is limited to moderate offset values and the adjustment is not trivial for very heterogeneous coverage footprints as it will occur in ultra dense deployments. In addition, current research work assumes slow updates (scale of seconds) for interference coordination [32, 19]. The reasons are an increased RTT (due to X2 signalling delays) but more importantly, the need to guarantee the system stability.

One of the major design goals for HetNets is the use of radio resources across MCells and SCells to achieve figures of per-user throughput and system capacity similar to ideal backhaul deployments. The 3GPP has introduced the *Dual Connec-*

tivity concept in Release 12, where the user consumes radio resources provided by at least two different network points. Dual connectivity is one of the 3GPP potential solutions to improve user performance by combining the benefits of the MCell coverage and the SCell capacity [4]. Work in [37] addresses this topic in a DL scenario where MCells share resources with other cells, where a CA window is proposed to determine if CA-capable UEs should be included in inter-site CA. A dedicated frequency deployment is considered. The benefit of aggregating resources from both cells is verified for different traffic patterns, as well as for different load situations. This topic has been addressed as well in the UL in [38] where results show improvement of UL throughput with the use of inter-site CA in low load situations due to larger bandwidth accessibility. The improvements in UE performance with the use of shared resources provide a strong indication that cooperative techniques are becoming mandatory to maximize resource utilization and meet the requirements for innovative and more demanding applications.

One step further in the optimization of HetNet, is the relationship between UL and DL and how the association policies affect the performance on both links. Both UL/DL power and MCell/SCell load and power imbalance motivates the decoupling of both links, which is particularly beneficial in co-channel heterogeneous deployments. The basic decoupling scheme between a SCell and a MCell is shown in figure 8.3. In Release 12, 3GPP provided an initial evaluation of the HetNet performance when including UL and DL split, results show improvements particularly at the cell edge for both low and medium load scenarios [4, 14]. The literature has tackled the power and load imbalance problem recently and some relevant references can be identified. Authors in [12] present the path-loss cell association solution to the power imbalance problem. Results in terms of gain that can be achieved in the UL capacity are very promising. A detailed analysis of the decoupled access in terms of association probability, coverage and capacity are presented in [35], where prior work is extended by adding the analytical evaluation using stochastic geometry and architectural considerations. Results show same trend between the stochastic geometry analysis and the real world experimental data. Work in [13] introduces cell load and the backhaul limitation into the cell association process. SINR variance is reduced with the enhanced DUDe solution presented; also, the interference-aware UL power control applied allows a further improvement in the UL throughput. Finally, [34] contributes to the topic with the analysis of the UL SINR and rate distributions as a function of the association rules considering UL power control design parameters. Results show that minimum path-loss association leads to identical load distribution across all cells, which is also optimal in terms of rate, irrespective of power control parameters. When both UL and DL joint coverage must be maximized, the decoupled association is the optimal solution. It is beneficial because it reduces the QoS imbalance between both links.

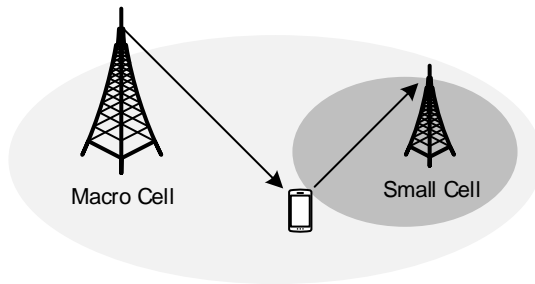


Figure 8.3: Example of HetNet with DL and UL decoupling

8.2.3 *Decoupling as a solution*

The current technologies that address the main HetNet challenges in the context of radio planning and interference management can be classified into three main groups:

- **Dual connectivity.** The literature has verified the UL improvements in dedicated deployments. However, there is yet no study that includes dual connectivity as a solution to the UL/DL imbalance problem, the association rules proposed in the prior art are based on the DL RSRP or RSRQ.
- **Cell RE with eICIC.** The literature has verified the UL improvement in co-channel deployments. This strategy helps to reduce the UL/DL imbalance, however, while the UL improvement is maximized the DL inter-cell interference is also increased. Therefore RE techniques must be always accompanied by eICIC solutions for the DL such as ABS. Nevertheless, the RE technique is limited to moderate offset values due to the harsh interference in the DL. Cell selection rules are based on DL RSRP and RSRQ with a RE offset added.
- **UL/DL decoupling.** There are verified improvements in the UL in co-channel deployments with the use of DUDe. The decoupled association policies in the UL and DL can cater for the different requirements of both links and, in turn, can successfully solve the UL/DL imbalance problem in terms of coverage, load and interference.

Based on this comparison, it is clear that DUDe brings the benefits of having very high RE offsets in the UL without the interference effects in the DL, since both links are separated, and connected to the best serving cell.

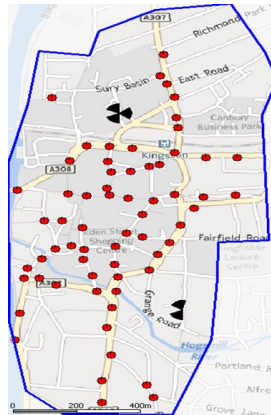


Figure 8.4: Vodafone SCell LTE test network in London.

8.3 Simulation setup

In this section, the simulation setup used in the performance evaluation in Section 8.4 is presented. The simulation setup follows largely from [12] where the simulation scenario is based on the Vodafone LTE SCell test network in the London area shown in Figure 8.4. The test network covers an area of approximately one square kilometre and consists of two sectored Macro sites and sixty four SCells illustrated by the black shapes and small circles respectively. The users distribution is based on traffic data extracted from the live test network. Additionally, a high resolution 3D ray tracing pathloss prediction model is used which takes into account clutter, terrain and building data. This guarantees a realistic and accurate propagation model. The UL power control is based on the open-loop fractional pathloss compensation algorithm as specified by 3GPP [3].

8.4 Performance Evaluation

As explained in Section 8.1, DUDe is considered to be a major paradigm shift from the cell-centric approach used up to now in designing cellular networks into a more flexible device-centric approach, which is envisioned to be one of the main building blocks of future 5G networks. DUDe can offer substantial gains in the UL in terms of coverage, throughput, reliability, load balancing and interference behaviour. Improving the UL performance has become more and more important due to the emergence and exponential growth of the Internet of Things (IoT) where the traffic is often UL centric and also the increasing popularity of symmetric traffic applications such as social networking, video calls and real time video gaming. In this section, simulation results that confirm the gains in terms of the previously mentioned criteria will be presented and discussed in details.



Figure 8.5: Uplink coverage of the Femto-Baseline (left), Pico-Baseline (middle) and DUDe (right) cases where black and grey represent the MCells and SCells coverage respectively.

The simplest form of DUDe is considered in this section where the UL association is based on the UL received power whereas the DL association is based on the DL received power.

DUDe is compared with conventional LTE operation where both UL and DL associations are coupled and based on the DL received power. Two LTE baseline cases are considered where SCells are treated as Pico cells and Femto cells and these two cases are termed as Pico-Baseline and Femto-Baseline respectively. The transmit powers of Macro, Pico, Femto cells are 46, 30, 20 dBm respectively. Subsequently, a set of results illustrating the performance gains from DUDe will be presented.

8.4.1 Coverage and Capacity

In a HetNet, the UL and DL coverages are quite different as discussed in Section 8.2 and shown in Figure 8.2, therefore basing the UL and DL associations on the same criterion, which is the DL received power, is highly suboptimal. In this part, the great difference between the UL coverage in the LTE baseline cases and DUDe is illustrated and the resulting gains in capacity are highlighted.

Figure 8.5 illustrates the UL coverage of the Mcell and Scell layers for the three cases in comparison, where Mcell and Scell coverages are shown in black and grey respectively. The UL coverage of Scells is shown to be very small in the Femto-Baseline and Pico-Baseline cases. However, in the DUDe case Scells have a much larger UL coverage, which is shown to be much less dominated by Mcells than in the LTE baseline cases. This effect results in a more homogeneous distribution of UEs between the nodes which, in turn, leads to a much more efficient use of resources as will be demonstrated in the following results.

The homogeneous coverage has a large effect on the throughput and specially the cell edge UEs throughput, which is represented by the 5th percentile through-

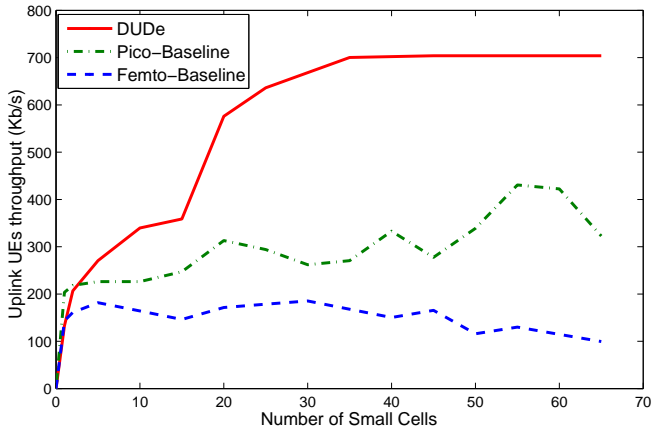


Figure 8.6: Uplink cell edge (5th percentile) throughput evolution with the number of SCells.

put. Figure 8.6 shows the effect of increasing the number of SCells in the simulated scenario on the 5th percentile throughput. In the DUDe case, it can be noticed that the 5th percentile throughput is improving quickly with the number of SCells, this is partly due to the increased UL coverage of SCells in DUDe as shown in Figure 8.5. As the number of SCells increase, the 5th percentile UEs throughput starts to saturate as these UEs become limited by the channel quality and transmit power and the gains start to be more pronounced in the 50th and 90th percentiles. Looking at the Femto-Baseline and Pico-Baseline cases, it can be noticed that adding SCells has little effect on the 5th percentile throughput due to the very limited coverage of SCells in both cases, which makes them more effective for the 50th and 90th percentile UEs. In addition, in these two cases the throughput is fluctuating as the number of SCells is increased. This is due to the high interference that the SCells UEs create to the Mcell cell edge UEs, as these UEs are closer to the SCells so they suffer from a high level of interference. This effect is emphasised more in the Femto-Baseline case as the 5th percentile throughput starts to decrease after a certain point whereas in the DUDe case the throughput increases in a more stable way since UEs always connect to the node to which they have the best UL channel which guarantees a lower interference level. This effect on interference is further addressed later on in this section.

Figure 8.7 shows the 5th, 50th and 90th percentile throughput results for the three cases in comparison. The 5th percentile UL throughput in DUDe is increased by more than 200% and 100% compared to Femto-Baseline and Pico-Baseline respectively. Similarly, the 50th percentile throughput is improved in the DUDe case by 600% and 100% compared to Femto-Baseline and Pico-Baseline. The gains result from the improved coverage of SCells in the DUDe case which results in a better distribution of UEs between the nodes giving way to a more efficient resources utilization. In addition, the fact that UEs connect to the node to which they have the best

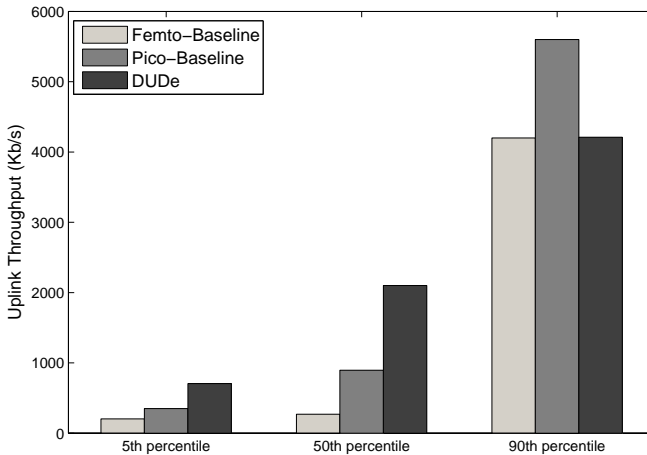


Figure 8.7: Uplink 5th, 50th and 90th percentile throughput.

UL channel results in an improved Signal to Interference and Noise Ratio (SINR), which allows UEs to use a higher modulation and coding scheme and in turn achieve a better utilization of resources and a higher throughput. As for the 90th percentile throughput, DUDe does not achieve the highest throughput, which makes sense since DUDe aims at improving the load balancing effect that will naturally result in a reduction of the peak data rate.

Additionally, it can be seen that Pico-Baseline achieves the highest 90th percentile throughput, which seems counter intuitive since it would be expected that the Femto-Baseline would be the one achieving the highest peak data rate. However, looking at the 98th percentile throughput, the throughput reaches 15 Mb/s and 10 Mb/s in the Femto-Baseline and Pico-Baseline cases respectively. This shows that the effect of Scells in the Femto-Baseline case is limited to a small number of UEs.

8.4.2 *Reliability and load balancing*

Reliability is becoming one of the most important requirements of future cellular systems due to the proliferation of the Internet of Things (IoT) which is an umbrella for several applications such as sensor networks, factory automation and many more. The reliability of wireless systems can be affected by several factors including congestion and coverage loss, where a device could have its connection dropped due to lack of available resources or due to very low link quality resulting from the device being out of coverage or being subject to high level of interference. DUDe plays a crucial role in improving the UL reliability by reducing the outage rate as will be shown next.

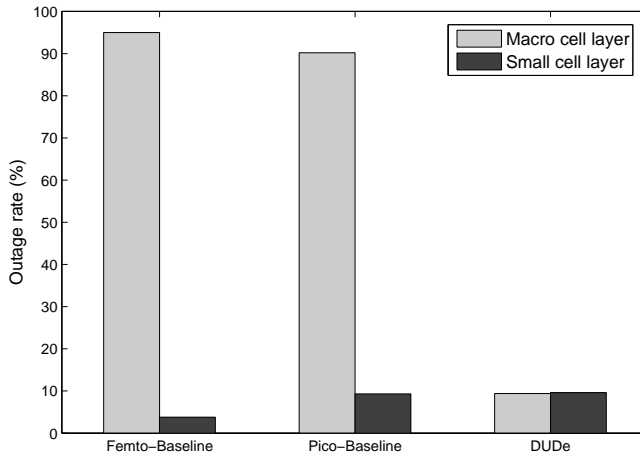


Figure 8.8: Outage probability.

Figure 8.8 illustrates the average outage rate for the Mcell and Scell layers for the three cases where a high traffic scenario is simulated by setting the minimum requested throughput for each UE to 1 Mb/s. The outage is defined as the percentage of UEs that cannot achieve the 1 Mb/s minimum throughput. Since this scenario is considered to be a high traffic congestion scenario it requires a very efficient use of resources in order to satisfy the high throughput requirements of the UEs. As can be noticed from the figure, the Macro layer has a very high outage rate in the LTE baseline cases, which is explained by the fact that the Macro layer is very congested in the UL and MCells cannot serve all the UEs with the required throughput resulting in a high number of dropped UEs. However, in the DUDe case, UEs are distributed more evenly among the nodes so the outage rate can be drastically reduced to less than 10% on the MCell and SCell layers.

Another trend that affects reliability is outage caused by the loss of coverage or poor channel conditions. Figure 8.9 shows the outage rate against the maximum transmit power of the UE. Outage here is defined by the fraction of UEs whose UL signal quality is lower than the one needed to access the lowest LTE modulation and coding scheme (MCS) and therefore are dropped. This result shows that for the same maximum UE transmit power DUDe can reduce the UL outage rate by up to 25% compared to baseline LTE. Alternatively, for the same outage rate the maximum transmit power can be reduced by up to 10 dB using DUDe. These results are very crucial for IoT applications, where either link reliability or battery life or both are of paramount importance and, it can be seen from the figure, that both can be drastically improved using DUDe.

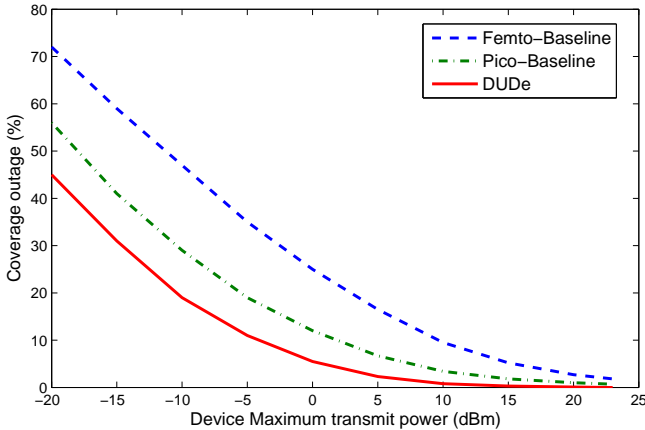


Figure 8.9: *Outage probability versus the maximum uplink transmit power.*

The load at a given BS in the UL maybe different than the load in the DL at the same BS. This implies that it is not optimal, in terms of load balancing, to have the same set of UEs connected to the same BS in the UL and DL.

In addition, DUDe has been shown in the previous results to improve the UL coverage of SCells, which results in a much better distribution of the UEs among the different tiers of the cellular network. This load balancing effect is illustrated in Figure 8.10 where the average number of UEs per cell in each tier is shown for the three cases in comparison. DUDe results in a much more homogeneous distribution of UEs between the different node types, which is directly translated into a drastically improved spectral efficiency and outage reduction as shown in the previous results.

8.4.3 *Interference behaviour*

Interference is one of the main limiting factors in cellular networks. DUDe has the ability to decrease the UL interference due to multiple complementary effects.

First, and as a result of the UL association that tends to connect the UEs in the UL to their closest node or the node to which they have the best UL received power, this can be translated into a reduction in UEs transmit power as shown in Figure 8.9 and emphasized in [8]. This has the effect of reducing the UL interference to other base stations, which is quite significant especially for UEs with low UL SINR.

Secondly, DUDe provides the ability to independently select the association that minimizes the interference at the UE as well as the BS. The UL interference depends on the UEs position relative to the interfered BS, the UL power control algorithm used and the UL precoding weights. In contrast, the DL interference towards a certain UE depends on the transmit power of the BS, the DL beamforming weights and the distance from the different BSs. On top of this, the UL and DL have different traffic and scheduling behaviours that can be considered nearly independent. For all

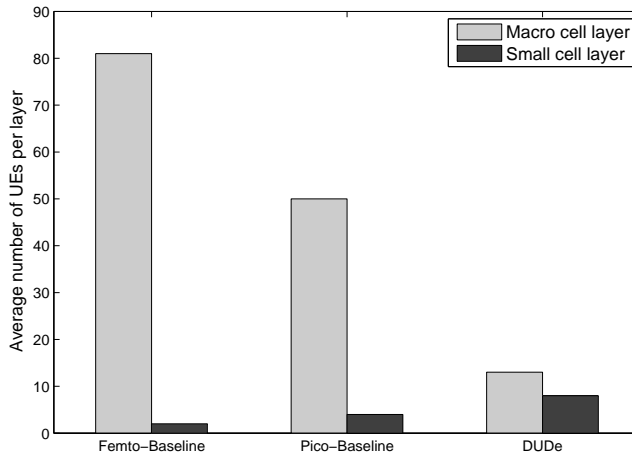


Figure 8.10: Average number of UEs per base station for MCell and Scell layers.

the above reasons, a decoupled UL and DL association that allows the UE or BS to seek the best UL or DL connection that reduces the interference on the respective link makes a lot of sense and is expected to outperform the conventional coupled association.

In addition to reducing the average amount of interference, DUDe allows a reduction in the standard deviation of the UEs UL SINR over time as shown in Figure 8.11. Reducing the variance of the SINR means that the channel is more stable and predictable, which has a substantial impact on reducing the complexity of radio resource management (RRM) and self-organizing network (SON) functions.

8.5 Interoperability of DUDe with 4G and 5G features

The deployment of a reliable high speed spectral efficient network needs the inclusion of a variety of innovative features, provided that link level solutions have evolved to near Shannon limit capacity with advanced Modulation and Coding Schemes (MCS). Given this, 4G and the forthcoming 5G, must offer pioneering solutions or improved versions of earlier releases in spectrum management and cooperative communications.

In particular, the interoperability of DUDe with other radio access technologies can lead to an overall improvement of these features. The reduced interference variability, the enhanced network flexibility or the reduced transmit power are some of the advantages that help to make the most out of the radio management techniques.

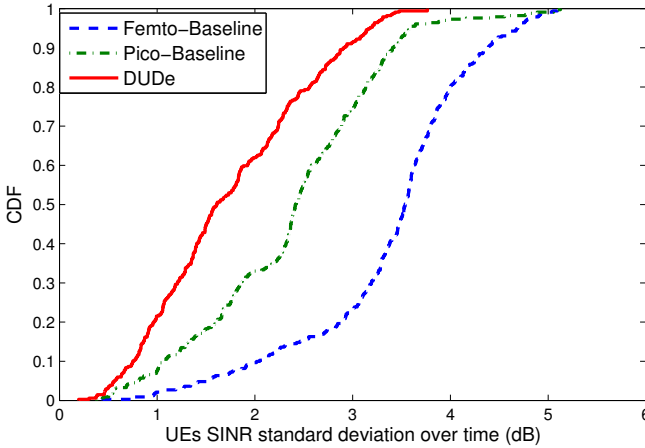


Figure 8.11: Uplink UEs SINR standard deviation over time.

8.5.1 Inter-band Carrier Aggregation

CA provides great advantages when carried out in HetNets. In particular, the inclusion of CA in such context has been recognised as a feasible way of providing multi-site radio resource allocation feature that was first introduced in Release 11 [3]. Also, CA in general (multi or same site aggregation) allows to improve capacity by extending the available bandwidth, and supports mobility and interference management techniques.

Several studies have focused specifically on the implementation of CA in the UL, where power capabilities of the UEs constitute the most limiting constraint [24, 23]. It is observed that the potential gains of CA transmissions are strongly related to the power demanded, which essentially depends on the bandwidth allocations and UE pathloss. Cell association determines which eNB is serving the UE. The eNB being closer to the UE means that less power is needed to transmit data in an allocated piece of bandwidth. To assure a correct performance of the aggregated bandwidth transmissions in the UL, it is crucial to account for the UE maximum transmit power. In MCell-only deployments, cell edge UEs are less likely to transmit in CA, however in heterogeneous deployments, the distance to the eNB is shorter given the higher cell density. In such a context, if UEs are associated based on the DL RSRP, the UL CA transmission is going to be highly restricted. Decoupling strategies are more lax in adopting aggregated transmissions; mainly, due to the improvement in UL power availability brought by decoupled associations. This is important, since CA is intended to be applied in both UL and DL, and with traditional DL received power association rules, the applicability of CA in the UL is seriously conditioned by the lack of power availability.

8.5.2 Cooperative Multi-Point

Base station cooperation, in the form of Coordinated Multi-Point (CoMP) transmission or reception, has gained popularity in the context of HetNets as a means to increase the UE achievable throughput. eNBs within the same cluster communicate via backhaul links (i.e., via the X2 interface) with the objective of minimizing the inter-cell interference and capitalize on the benefits of distributed antenna systems. In fact, interference within a cooperation cluster can be effectively cancelled [1, 18]. This level of coordination and cooperation can be carried out in both UL and DL, and the realization of such coordination relies strongly on the availability of sufficient backhaul capacity, first to serve the UE in the cell cluster, and second to communicate with other cells in the cooperation cluster. This backhaul dependency can be very limiting in situations of high load, and in capacity limited links.

The increased flexibility provided by decoupled UL and DL associations provides advantages when selecting UL and DL coordinated transmissions or receptions. In particular, there is no need to have both UL and DL simultaneous connection to the entire cooperating set of base stations and the UE could have unequal UL and DL active links (as in the case of CA). This flexible association inside the cluster, and the interoperability of DUDe with CoMP goes one step further in the device-centric network, since the UE can select independently the number and position of DL and UL serving cells, according to several input parameters, as backhaul capacity, power limitation, throughput maximization, among others.

8.5.3 Millimeter Wave

The ever increasing network traffic demands have led to several trends in cellular networks including the densification of the network and the emergence of heterogeneous networks where SCells are playing a key role in providing capacity in hotspots and dense urban areas. Another approach was to explore new unused frequency bands that would help satisfy the projected future traffic needs. Popular sub 6 GHz frequencies are becoming scarce and would no longer be able to cope with the increasing network traffic.

A solution to this is to resort to higher frequencies in the millimeter-wave bands (20 - 100 GHz) where a significant amount of spectrum is underutilized or completely unused. The several GHz of available spectrum promise a spectacular increase in capacity which qualifies millimeter-wave technologies as one of the main enablers of future 5G networks.

In reality, millimeter-wave bands will not replace sub 6 GHz bands, at least initially, where sub 6 GHz will still be needed to provide coverage and ubiquitous service since millimeter-wave frequencies have poor propagation properties and are quite sensitive to blockages. The existence of millimeter-wave networks as an overlay to the conventional sub 6 GHz networks would require a change in the conventional cell association techniques. DUDe is expected to play a key role in millimeter-wave networks for several reasons.

Recent studies on electromagnetic field exposure [9] have shown that to be compliant with the mandatory exposure limits at frequencies above 6 GHz, the maximum transmit power of devices would have to be reduced by several dBs below the conventional power levels in current cellular networks. This has significant implications on the UL; since the UL coverage depends on the transmit power of devices and since mmWaves have unfavourable propagation properties, a reduction in the transmit power of devices could result in serious degradations in the link quality in the UL. A possible solution to this problem would be to decouple the UL and DL where for some UEs the UL will be served by UHF MCells with a better link budget, whereas the DL would be served from the mmWave SCells. That is, while previously DUDe was discussed in the context of associating UEs to MCells in the DL and SCells in the UL, for mmWaves the opposite strategy might prove useful.

In addition, mmWave SCells are expected to have a very limited coverage area and consequently the conventional association techniques based on the DL received power will leave the mmWave SCells very much under-utilized considering the vast amounts of spectrum available in the millimeter-wave bands. SCell biasing or range extension is a technique that was introduced in LTE release 10, where effectively a cell selection bias is applied to SCells in order to expand the coverage of SCells and improve their offloading effect. A similar approach could be employed for millimeter-wave SCells to solve the aforementioned problem by using aggressive range extension values to attract as many UEs as possible in order to make use of the large spectrum chunks available at millimeter-wave frequencies.

DUDe would play an important role in this setup as it would allow to set different values of range extension for UL and DL in order to meet the requirements of both links. For instance, if the UL is required to be highly reliable whereas the DL can be less reliable but is more demanding in terms of throughput, a high value of DL SCell range extension can be used while not applying any range extension in the UL. Therefore DUDe would offer the flexibility to cater for the different needs for the UL and DL in millimeter-wave scenarios.

8.5.4 *Different Duplexing Techniques*

DUDe can function with both FDD and TDD, with different implications from a system level and spectrum point of view.

TDD allows much more flexibility in trading DL and UL resources as compared to FDD. With DUDe, it was demonstrated that fewer UL resources are needed to achieve the same UL rate as compared to coupled operation. This could lead the way to DUDe having a positive effect on the DL rate by allowing the DL to use more resources via dynamic TDD.

Another benefit of TDD is the possibility of estimating the DL channel via UL reference signals. This is quite important especially for channels with large dimensionality like in the case of massive MIMO. However, when DUDe is used this reciprocity is broken as DL and UL transmissions originate and terminate at different BSs. Much of the existing spectrum is paired FDD spectrum, therefore for these two reasons massive MIMO may need to be supported without channel reciprocity.

In the medium to long term, DUDe along with different emerging technology trends could lead to a rethinking of the traditional FDD/TDD paradigms. DUDe, hyper-densification, the use of millimeter wave frequencies and highly directional antennas, could allow for duplexing approaches over the spatial domain. For instance, the same band could be used for two different devices located in different locations, one receiving in the DL from a certain BS and the other transmitting in the UL to another BS. A spatial UL/DL coordinated scheduling mechanism, would effectively allow full-duplex like gains without the complicated interference cancellation mechanisms of full-duplex. In addition, once analog/digital interference-cancellation mechanisms are realised to support full temporal duplex, DUDe can be beneficial as it enables a generalized decoupled access that would allow the support of a DL and not necessarily the same UL user in the same frequency band.

8.6 Enabling Radio Access Network Architectures for Decoupling

The access network configuration presents several challenges to enable the use of decoupled connections, while maintaining interference and energy efficiency at reasonable levels. 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 [22, 35]; also, the use of one carrier to exclusively transmit control signals will lead to poor spectral efficiency. In particular, the most challenging part of devising architectures that support DUDe is that an increased amount of control information needs to be signalled back to the corresponding serving cell.

The 3GPP has proposed several architectural alternatives for DL dual connectivity and the architecture needed to support the U-Plane aggregation from different eNBs is expected to be very similar to those proposed for dual connectivity, based on the bearer split concept [4]. Alternatively, 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 layer 1 and layer 2 control signals while maintaining the RAN latency requirements.

8.6.1 Network Procedures

When a UE accesses the system and associates with eNB, a radio resource control connection is set up. The RRC is the protocol that handles all the control plane signalling of Layer 3 between the UE and the EUTRAN. Among the different tasks, the RRC is in charge of the connection configuration, radio bearer establishment and release, mobility procedures and configuration of power control mechanisms. The amount of resources the RRC consumes dynamically changes, depending on the cell traffic type, number of UEs being served and the connection timers. Short timer values generate high RRC signalling overhead whereas high timer values result in poor use of idle mode [2].

Layer 1 and layer 2 control channels are crucial to support the user plane data transmission. PDCCH and PUCCH physical control channels are in charge of delivering information related to the active transmission, as the scheduling decisions or the acknowledgements of the transmitted information, and the channel state information to perform accurate link adaptation. In particular, RAN control signals that are of paramount importance to handle scheduling and other MAC layer procedures are:

- control information carried in the PUCCH: DL CQI, buffer status reports, scheduling requests and power headroom reports;
- downlink control information (DCI) carried in the PDCCH 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 UL;
- downlink HARQ acknowledgment messages carried in the Physical Hybrid-ARQ Indicator Channel (PHICH); and
- RRC messages that configure the UE connection and release, as well as the PUCCH position and resources and SRSs for UL channel state information configuration.

Among all these control information, the one most stringent in terms of latency requirement is the HARQ RTT, which is approximately 8 ms, considering information processing at both sides and frame transmission. The HARQ in the UL follows a synchronous process: with a periodicity of 8 sub-frames the same HARQ id process is used, and no explicit information is exchanged about the process id. This means that there is a strict relation among the sub-frame number and the HARQ process id, which results in strong delays if one of the acknowledgements cannot be sent in the corresponding sub-frame. The architecture that enables the use of DUDe shall be the one that minimizes delay in the acknowledgment synchronization process and assures that the RAN RTT is maintained.

8.6.2 *DUDe with Assisting Connections*

The use of supporting UL and DL connections in order to transmit the control signals is an immediate solution to feasibly implement DUDe. The terminal is connected to both cells and aggregates the data flows; in such a way, all signalling layer 1/layer 2 and RRC can be handled locally. Architecture alternatives along these lines were presented in [14].

A simple way to support this is with the use of CA, where each carrier component is configured separately to carry a shared and a control channel. This configuration is shown in Figure 8.12. This solution allows to keep the RAN latency at desired levels since no further delay is sensibly introduced. However, potential disadvantages of this configuration are: first, having two simultaneous UL connections can seriously jeopardise the UL performance in terms of UE energy efficiency. Second,

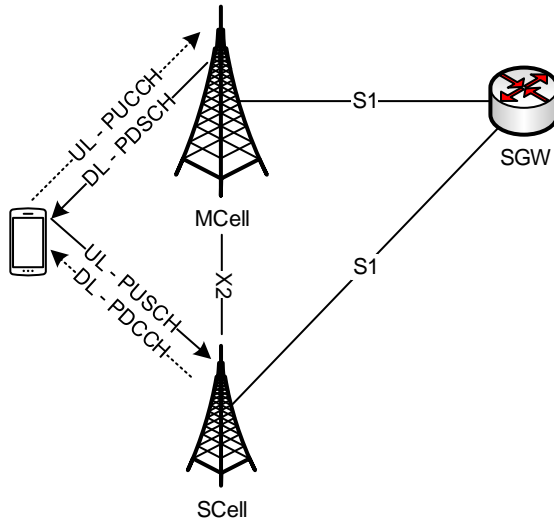


Figure 8.12: Example of DUCe configuration with assisting connections

this configuration does not maximize capacity over the available spectrum, since one component carrier is exclusively used to handle control information.

8.6.3 User Plane Bearer Split for DUCe

Latest releases of LTE-A (Release 12 and 13) consider new architecture alternatives for dual connectivity, with the introduction of the novel bearer split concept, an alternative that eases having two simultaneous transmissions in different eNBs, a MCell and a SCell for instance [4]. On the other hand, the UL feasibility of adopting the bearer split can be argued in terms of power consumption, and UL data should be either transmitted directly to the MCell, or forwarded to the MCell by the SCell [22].

Based on this, in a DUCe context, assisting PUSCH or PUCCH connections may not be carried out in the MCell; and PDSCH and PDCCH connections may not be carried out in the SCell. Having no control information signalled back to the corresponding serving cell through the user interface (i.e., using supporting UL and DL connections), the delivery of layer 1 and layer 2 signalling and RRC relies on the non ideal backhaul connection, the X2 interface, between both serving cells.

Following the LTE architecture of distributed cells and using the bearer split concept two possible solutions arise for the decoupling:

1. The radio bearer is managed at the serving gateway (SGW), and the DL flows through the S1 from the SGW to the MCell and the UL flows through the S1

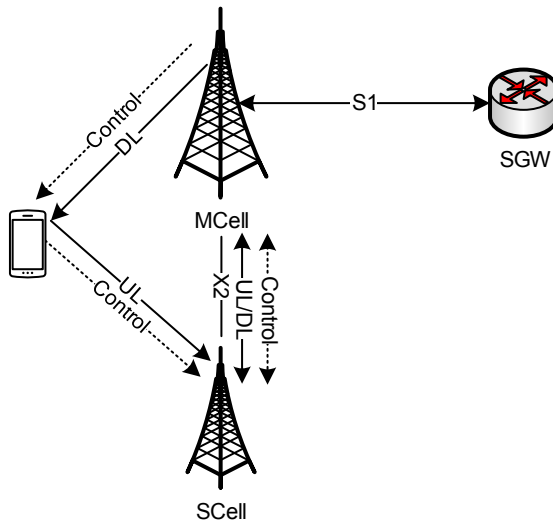


Figure 8.13: Protocol stack information flow for radio bearer managed at MCell

to the SGW from SCell. This option has reduced flow control among both serving eNBs, since only part of the control signals, for example HARQ acknowledgements, needs to be forwarded through the X2. The architecture largely follows the one in Figure 8.12, where information flow for DL and UL goes through the S1 interface.

2. The Radio bearer is managed at the MCell, 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 back-haul interface, while respecting the 8 ms HARQ round trip time requirement. Master/slave configuration is for UL and DL, and the processing of each MAC PDU is done in the corresponding cell. Figure 8.13 shows a diagram of the information flow for DL and UL.

Both configurations allow to maximize the spectrum usage since all carriers can be configured for U-Plane information transmission. Current heterogeneous networks pose a big challenge to meet the RAN RTT fixed by the HARQ process, since delays that range from 5 ms to 30 ms are expected in the X2 interface [4].

Similar to what has been done in the user plane, the control plane that enables the full DUDe is analysed based on the RRC protocol architecture presented in [4] for dual connectivity. Two alternatives have been presented:

- Based on prior negotiation of parameters between SCell and MCell, the MCell generates the final RRC message. This alternative is good, as only one RRC connection is active, but requires an assisting UL connection with the MCell.
- In this case, the RRC message is built by the SCell based on input parameters provided by the MCell. This solution adds complexity in the UE side as it needs to handle parallel RRC procedures [40].

In light of this, to completely decouple UL and DL poses a huge effort in the RRC connection.

8.6.4 *Centralized solution*

Previous architectural solutions present limitations in terms of spectrum usage, power efficiency and meeting latency requirements. If both eNBs, MCell and SCell, are able to share the same BBU, then complete DUDe can be handled, and mitigates most of the drawbacks presented hitherto. The concept under the BBU sharing is the centralized based architecture, also known as C-RAN, which breaks the static relationship between BBU and RRH, and each RRH does not necessarily needs to be mapped with one specific BBU. In particular, the radio signals from/to a RRH can be processed by a virtual eNB, which facilitates enhanced real-time cooperation among them. Given this, virtualization technology (network function virtualization, NFV) will maximize the flexibility in the RAN network allowing different levels of coordinated transmissions, or separated associations. Figure 8.14 shows the C-RAN architecture for DUDe.

The C-RAN architecture needs to use a new fronthaul interface that allows the communication among the RRH and the BBU, which can potentially increase the RAN latency. Different protocol functional split options can be recognised which variate the delays and capacity requirements over the fronthaul interface [30, 11, 25]. There is a benefit in locating part of the 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 (i.e, dark fiber, wireless or other technologies) and at the same time enable the possibility of having enhanced cooperative radio resource management procedures. Realistic fronthaul delays can range from 250 μ s to 4 ms [30].

When the DL serving RRH and the UL serving RRH share the same BBU, one RRC connection is held. To assume this situation, the BBU has one single physical cell identifier (PCI).

8.6.5 *On the road to 5G based architectures*

The next few years will witness a surge in research and development in the different areas of 5G. The ITU and 3GPP are ramping up their activities on requirements and standardization of 5G which is expected to be rolled out commercially by the early 2020's. In addition, there is an increasing consensus about the main requirements of

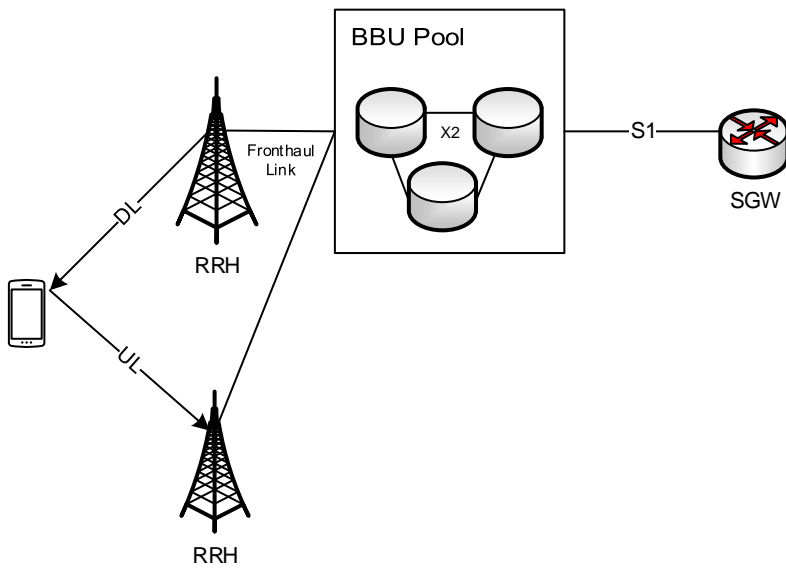


Figure 8.14: C-RAN Architecture with DUDe

5G regarding capacity and latency as well as the key technical features of 5G, including ultra base station densification, massive MIMO, millimeter wave frequencies and possibly a "cell-less" architecture [5, 7].

With this view of 5G, does the new 5G standard need to include extra features to natively support DUDe?

An important question to answer is whether a simple evolution of current 3GPP architecture design would be able to efficiently support DUDe in future heterogeneous 5G deployments. In the previous subsections, a discussion was undertaken on how the LTE-A architecture already supports an implementation of DUDe when different BSs are linked via fiber to the same baseband unit. It was also discussed how the support of DUDe in 4G could be limited to different frequencies in the case of carrier aggregation or dual connectivity. Intuitively, any future 5G releases in 3GPP should allow for same-frequency dual connectivity, which, despite having implications on resource and interference management, is not considered to be a major upgrade.

Another issue to address is to ensure proper encryption of all data and control channels, specially when communication via the X2 interface is used between BSs. In LTE, each BS can support tens of IP Security (IPSec) tunnels. However, the management of security via IPsec is so problematic that operators tend to deploy only a few IPsec gateways per country, where most of the gateways are deployed near the SGW. This means that traffic that logically goes via the X2 is actually routed via the SGW, which adds a lot of delay that renders the DUDe operation rather inefficient. Whilst LTE-A has more IPSec gateways deployed closer to the mobile edge, 5G architecture designs need to consider efficient encryption of the X2 traffic to reduce the extra latency in 4G.

In addition, some integration work is needed for some emerging techniques that have proved efficient for coupled systems. The integration of DUDe with decoupled Control/Data planes and License Assisted Access (LAA) will require some architecture changes. Self-organizing networking (SON) will also be instrumental in managing and coordinating the increased degrees of freedom introduced by DUDe.

Given, the above discussion, it can be concluded that a native support of DUDe does not require major design changes in 5G from an architectural perspective.

8.7 DL/UL Challenges and Opportunities for Higher Layers

In this section, the main focus is on the transport layer and in particular on the TCP, given that this is the prominently used transport layer protocol to achieve reliable end-to-end data transfer in IP based networks. Despite the fact that TCP has been initially designed for elastic applications it is currently commonly used in various popular streaming applications. It is worth noting that Real Media and Windows Media, the two dominant streaming media applications, both are based on TCP streaming. Here, we discuss challenges in the higher layer and the TCP in two folds: First, the effect of UL channel on the TCP performance and second, is the support available at the TCP to enable DL/UL decoupling. In this section, both sides of the argument are widely elaborated and the following questions are mainly answered:

1. How DL/UL asymmetry can impact performance of TCP?
2. What are the effects of DL/UL joint resources allocation on the TCP?
3. What are the effects of DUDe on the TCP?
4. How DUDe can be enabled by the the available mechanisms in TCP?

8.7.1 Effect of UL on the performance of higher layers

TCP connections, as being inherently bi-directional, require ACKs from the receiver in order to achieve reliable communication. On one hand, the common assumption in data centric networks is that the DL carries heavier data load, thus less bandwidth is specified for the UL path. On the other hand, channel measurements are carried out based on the DL channel. Moreover, in many of today's applications such as MTC, the sender can be the mobile terminal, which is different from traditional applications, and therefore UL capacity plays an important role.

The effect of link asymmetry on the performance of TCP is widely studied in wired networks. Limited available bandwidth and congestion on the reverse path breaks down the principle of ACK clocking, and may cause an increase in the RTT, which can degrade the TCP throughput on the forward path [6]. Several research works explore these issues and a range of solutions have been proposed. Some of these proposals require explicit support from routers or middle boxes, whereas others are end-to-end schemes. For example, ACK congestion control [16] attempts to reduce the sending rate of ACK traffic, with the assumption that a reduction in ACKs rate may help reduce the congestion itself.

Within mobile and wireless communications, a wealth of work has gone into optimizing the performance on the DL channel of wireless networks such as for example, honing its congestion awareness mechanism so that it is minimally affected by random wireless losses, and optimizing achieved fairness of the end-to-end TCP rate. Limited volume of research in fact addresses the effect of such imbalance on the TCP performance. The work in [26], however, brings the requirements of TCP on the UL path, into the actual radio resource allocation mechanisms for DL. Therefore, the limited capacity or lower channel quality on path will not adversely affect the end-to-end performance. This solution can be implemented by adding one bit in the optional part of TCP header to explain the state of congestion control TCP is in, i.e. *Congestion Avoidance* or *Slow Start*.

Modifications to the TCP Congestion Control

The core of TCP congestion control algorithm are the slow start and the congestion avoidance phases. In the slow start, the achievable throughput by the TCP flow simply depends on the actual value of CW and RTT. On the other hand, throughput in the congestion avoidance phase can be expressed by the TCP steady state throughput [31]. Introducing a new flag in the TCP header, SS/CA flag, can illustrates the actual

state of TCP, i.e. SS/CA equal to zero represents being in *Slow Start*, and one being in *Congestion Avoidance*¹.

TCP Congestion Control and DL/UL resource allocation

While resource allocation mechanisms commonly consider DL only, it is often useful to consider the joint DL/UL allocation problem. Such allocation techniques can provide a dynamic border between DL capacity and UL capacity, in order to allocate the total amount of resources more efficiently. This proportion can be defined dynamically based on various system constraints to guarantee the requested QoS. To fully illustrate the joint resource allocation, here an OFDMA resource allocation problem is detailed that is also constrained with the requirements of the bi-directional connection of the TCP so that the delivery of TCP ACK packets for the specific allocated bandwidth to the DL is guaranteed.

The DL resource allocation problem per se, could aim to maximize sum rate [39] subject to power consumption, minimize the overall power consumption while satisfying the minimum rate requirements [21], consider certain degree of QoS or fairness [27, 28]. Here, a joint allocation of DL/UL resources, to provide the maximum throughput on the DL, and also to guarantee the delivery of the DL packets with the appropriate data rate on the UL, is discussed. The objective of DL allocation scheme is presented in two variants. The first defined problem, i.e. (P1), aims to maximize sum rate, constrained to the proportional data rate on each individual TCP flow. The proportional rate is weighted with the theoretical TCP throughput, which is the throughput that a TCP flow can achieve, depending on the end-to-end RTT and the PER of the corresponding flow-this throughput can be considered as the actual capacity of the end-to-end path. In the second problem, i.e. (P2), the difference between the allocated wireless link rate and the theoretical achievable TCP throughput is studied. The (P2) aims to maximize the sum rate while minimizing the gap between the allocated data rate to each flow and the theoretical throughput that can be achieved by that TCP flow.

Formulation of the OFDMA resource allocation - (P1)

The formulated optimization problem (P1) aims to maximize DL sum rate, while TCP fairness is assured by imposing a set of nonlinear constraints into the problem. In other words, the proportional DL rate among users with respect to the TCP theoretical throughput, are subject to constraints. In the UL, assuming TCP receiver acknowledges every single packet, the minimum required data rate would be a proportion of the DL data rate for each specific TCP flow depending on the size of the ACK packet, which can be increased for example using the SACK option, $R_{u_i} \geq \rho R_{d_i}$, i.e. $0 < \rho < 1$.

$$(P1) : \text{Maximize } \sum_{i=1}^n \sum_{j=1}^m c_j a_{ij} w \log_2 \left(1 + \frac{p_{ij} G_{ij}}{\sigma^2 c_3} \right),$$

¹The SS/CA flag can be implemented in the TCP header, option section with Kind = 30, and Length = 1 [36].

$$\text{subject to: } \sum_{i=1}^n a_{ij} \leq 1, \quad \forall j \in \{1, \dots, m\} \quad (8.1)$$

$$\sum_{i=1}^n \sum_{j=1}^m c_j a_{ij} p_{ij} \leq P_T, \quad (8.2)$$

$$\sum_{j=1}^m (1 - c_j) a_{ij} p_{ij} \leq P_i, \quad \forall i \in \{1, \dots, n\} \quad (8.3)$$

$$\frac{R_{d_i}}{B_i} = \frac{R_{d_1}}{B_1}, \quad \forall i \in \{2, \dots, n\} \quad (8.4)$$

$$R_{u_i} \geq \rho R_{d_i}, \quad \forall i \in \{1, \dots, n\} \quad (8.5)$$

$$\sum_{j=1}^m c_j \leq m_d, \quad m_d \in \{1, \dots, m\}, \quad (8.6)$$

$$p_{ij} \geq 0, \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\} \quad (8.7)$$

$$a_{ij} \in \{0, 1\}, \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\} \quad (8.8)$$

$$c_j \in \{0, 1\}. \quad \forall j \in \{1, \dots, m\} \quad (8.9)$$

In (P1), despite the classic approaches in solely maximizing throughput, subcarriers (a_{ij}) and transmission power over each subcarrier (p_{ij}) are allocated so that certain performance metrics of TCP are guaranteed. These performance metrics are provided via the constraints that are detailed below. Constraints (8.1) ensure that every subcarrier is assigned to only one user. We assume c_j represents the allocation of subcarrier j to DL ($c_j=1$) or UL ($c_j=0$). Thereby, constraints (8.2) and (8.3) restrict the total available power at the base station, P_T , and at each mobile user, P_i . Moreover, constraint (8.4) provide fairness among TCP flows with maintaining proportional rate with respect to the TCP throughput for each user. Constraint (8.5) provide the required data rate for UL, in order to guarantee delivery of the DL allocated resources, and finally constraint (8.6) bounds the number of DL subcarriers to m_d . Note that the optimal value of m_d can be found solving the problem (P1) iteratively for different values of m_d .

Formulation of OFDMA resource allocation - (P2)

Similar to (P1), (P2) aims to maximize DL sum rate, but the instantaneous rate allocation is constrained by the TCP theoretical throughput. The difference between the allocated DL rate to the i th user and the theoretical TCP throughput of flow i is denoted by D_i .

$$D_i = | \alpha B_i - R_{d_i} |. \quad (8.10)$$

In Equation (8.10), α represents the overhead of the TCP/IP header. The resource allocation problem in this case attempts to minimize D_i while maximizing the sum rate. Therefore, (P2) can be defined as a multi objective optimization problem. There are various approaches to formulate such a multi-objective problem; we use a

well-studied approach that combines the multiple objectives into a single objective function whose solution is Pareto optimal.

$$(P2) : \text{Maximize } \sum_{i=1}^n \sum_{j=1}^m c_j a_{ij} w \log_2 \left(1 + \frac{p_{ij} G_{ij}}{\sigma^2 c_3} \right) - \mu \sum_{i=1}^n D_i .$$

$$\text{subject to: } \sum_{i=1}^n a_{ij} \leq 1, \quad \forall j \in \{1, \dots, m\} \quad (8.11)$$

$$\sum_{i=1}^n \sum_{j=1}^m c_j a_{ij} p_{ij} \leq P_T, \quad (8.12)$$

$$\sum_{j=1}^m (1 - c_j) a_{ij} p_{ij} \leq P_i, \quad \forall i \in \{1, \dots, n\} \quad (8.13)$$

$$R_{u_i} \geq \rho R d_i, \quad \forall i \in \{1, \dots, n\} \quad (8.14)$$

$$\sum_{j=1}^m c_j \leq m_d, \quad m_d \in \{1, \dots, m\}, \quad (8.15)$$

$$p_{ij} \geq 0, \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\} \quad (8.16)$$

$$a_{ij} \in \{0, 1\}, \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\} \quad (8.17)$$

$$c_j \in \{0, 1\}. \quad \forall j \in \{1, \dots, m\} \quad (8.18)$$

Constraints (8.11)-(8.18) are the same as (8.1)-(8.3) and (8.5)-(8.9). As mentioned above, problem (P2) has a Pareto optimal solution; thus the solution is not unique and it depends on the value of μ that balances the two objectives. In the above problem, increasing the value of μ shift the allocation balance towards TCP throughput, while decreasing the value of μ shift the balance towards a data rate maximization problem.

Finding Solutions for (P1) and (P2)

The well-used method for solving the multiple variable optimization problems is to decouple the problem [33], i.e. to allocate a_{ij} and p_{ij} separately. In the subcarrier allocation, it is assumed that power is equally distributed in all the subcarriers, therefore the solution is suboptimal. Afterwards, to a certain subcarrier allocation, an optimization problem can be reformulated over the continues variable p_{ij} . Thus, using the water filling approach, power will be distributed optimally.

The principle of the DL algorithm is to allocate the subcarrier with the highest channel gain available for each user. In addition to that, in the first round of allocation, users with the highest theoretical TCP throughput (largest value of B_i) are first to select subcarrier. Thereafter, in each round of the allocation, in solving problem (P1) users with the lowest proportional rate have priority to select the best available subcarrier. This step performs differently solving problem (P2), i.e. user with the smallest objective function selects the next subcarrier.

Subcarrier allocation in the UL in both problems is in order to satisfy the UL minimum rate requirements. The initial value of m_d maintains the proportion of ρ

Table 8.1: Parameters used in the numerical study

Bandwidth	5MHz
Total number of OFDMA subcarriers	52
maximum power at the base station	43 dBm
maximum power at the mobile device	23 dBm
thermal noise power (Johnson-Nyquist noise)	107 dBm
target BER	10E4
average SNR of wireless channel	15 dB
wireless channel model	ITU pedestrian model
TCP Maximum segment size	1460 bytes
TCP flow type	long-lived with SS/CA = 1
number of mobile users	1 to 15 users
end-to-end RTT	uniform in [10 ms, 200 ms]
Initial number of DL subcarriers (m_d)	32

between number of DL and UL subcarriers – $m_d = m \cdot \frac{1}{1+\rho}$. Afterwards, in few iterations, the largest value of m_d which satisfies constraint (8.4) will be found; clearly this value maximizes the objective function. The above procedure for problem (P1) is detailed in Algorithm 1, in which Ω_i is the set of allocated subcarriers to the user i in the DL and Ψ_i is the set of allocated subcarriers to this user in the UL. Algorithm 2 details the subcarrier allocation solving problem (P2) in which steps (a), (b), (d), and (e) are similar to Algorithm 1, and only step (c) is restated.

The problem of power allocation with pre-defined subcarrier allocation, is based on the reformulation of (P1) into a maximization problem over continuous variable p_{ij} (similarly for (P2)).

In the Algorithms 1 and 2, Ω_{i_1} and Ω_{i_2} are mutually exclusive, if $i_1 \neq i_2$; the same assumption is also valid for Ψ_i . Problem (P2) also can be rewritten as (P2') over the continuous variable p_{ij} .

Reducing the UL effect on TCP performance

Numerical evaluations are carried out using parameters listed in Table 8.1, to demonstrate reduced effect of UL imbalance on the overall system KPIs (based on [26]). The benchmarks are the sum rate maximization problem, denoted by (BM1), and also the sum rate maximization with an equal rate constraint, denoted by (BM2). The m_d is initialized with 32 subcarriers for the DL, while $\rho = 0.2$. Afterwards and in step (f) of Algorithm 1, the optimal value of m_d is calculated in few iterations. The two benchmark problems (BM1) and (BM2), operate at $m_d = 32$; thus the results comparison here mainly show how the end-to-end performance is benefitted from setting the border between DL and UL adaptive to the current load of the system.

The two examined KPIs here are DL TCP throughput that also reflects the UL TCP throughput, and fairness among TCP flows. Fairness is quantified based on Jain's fairness index, which is a well-used index in computer and communication networks. The presented numbers in Figures 8.15-8.17 are average values taken

Algorithm 1 Subcarrier Allocation Algorithm for the optimization problem (P1)

a) Initialization

1. $m_{d_1} = \frac{1}{1+\rho} \cdot m$.
2. Set $R_{d_i} = 0$ and $\Omega_i = \emptyset$ for $i=1$ to n and $C_d = \{1, \dots, m_d\}$.
3. Set $R_{u_i} = 0$ and $\Psi_i = \emptyset$ for $i=1$ to n and $C_u = \{m_d + 1, \dots, m\}$.
4. Sort the users' index in the descending order of B_i .

b) for $i=1$ to n

1. Find the subcarrier k satisfying $|G_{ik}| > |G_{ij}|$ for all $j \in C_d$.
2. Let $\Omega_i = \Omega_i \cup \{k\}$ and $C_d = C_d - \{k\}$.
3. Update R_{d_i}

c) while $C_d \neq \emptyset$

1. Find user l satisfying $R_{d_l}/B_l < R_{d_i}/B_i$ for all $i \in \{1, \dots, n\}$.
2. For user l , find the subcarrier k satisfying $|G_{lk}| > |G_{lj}|$ for all $j \in C_d$.
3. Let $\Omega_l = \Omega_l \cup \{k\}$ and $C_d = C_d - \{k\}$.
4. Update R_{d_i}

d) for $i=1$ to n

1. Find the subcarrier k satisfying $|G_{ik}| > |G_{ij}|$ for all $j \in C_u$.
2. Let $\Psi_i = \Psi_i \cup \{k\}$ and $C_u = C_u - \{k\}$.
3. Update R_{u_i}

e) while $C_u \neq \emptyset$

1. Find user l satisfying $(R_{u_l} - \rho R_{d_l}) < (R_{u_i} - \rho R_{d_i})$ for all $i \in \{1, \dots, n\}$.
2. For user l , find the subcarrier k satisfying $|G_{lk}| > |G_{lj}|$ for all $j \in C_u$.
3. Let $\Psi_l = \Psi_l \cup \{k\}$ and $C_u = C_u - \{k\}$.
4. Update R_{u_i}

f) If $\sum_{i=1}^n R_{u_i} \geq \rho \cdot \sum_{i=1}^n R_{d_i}$.

1. Find the largest $m_d \in \{m_{d_1}, \dots, m\}$ such that constraint (8.5) are satisfied.
 2. Else, find the largest $m_d \in \{1, \dots, m_{d_1}\}$ such that constraints (8.5) are satisfied.
-

Algorithm 2 Subcarrier Allocation Algorithm for the optimization problem (P2)

c) while $C_d \neq \phi$

1. Find user l satisfying $R_{d_l} - \mu D_l < R_{d_i} - \mu D_i$ for all $i \in \{1, \dots, n\}$.
2. For user l , find the subcarrier k satisfying $|G_{ik}| > |G_{ij}|$ for all $j \in C_d$.
3. Let $\Omega_l = \Omega_l \cup \{k\}$ and $C_d = C_d - \{k\}$.
4. Update R_{d_i}

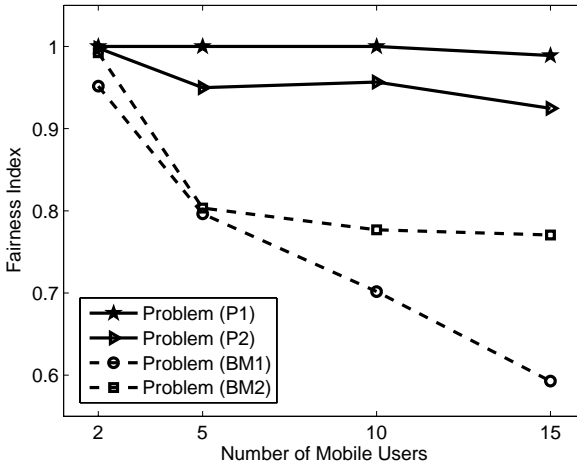


Figure 8.15: Jain's Fairness Index as achieved by solving resource allocation problems (P1), (P2), (BM1), and (BM2) Vs. the number of mobile users

from 150 Monte Carlo simulations. Observed from Figure 8.15, it can be seen that the achieved fairness index by resource allocation (P1) is increased approximately 30% in average compared with the results of allocation scheme (BM1). In addition, Figure 8.15 shows that, as the number of mobile users competing over the wireless link is increased, distribution of the resources in a fair manner is more challenging. The highest increase in fairness index is in the fifteen-user case, where (P1) performs 70% better than the benchmarks. Moreover, Figures 8.16 and 8.17 show that joint allocation of the DL and UL improve the total aggregated throughput as well as the aggregated DL throughput by approximately 15%.

DL and UL, Joint and Decoupled

Enabling DUDe greatly contributes to the success of DL/UL TCP-ware resources allocation, while the joint resource allocation improves the efficiency of both UL and DL communications. First, the joint resource allocation will allow UL to have

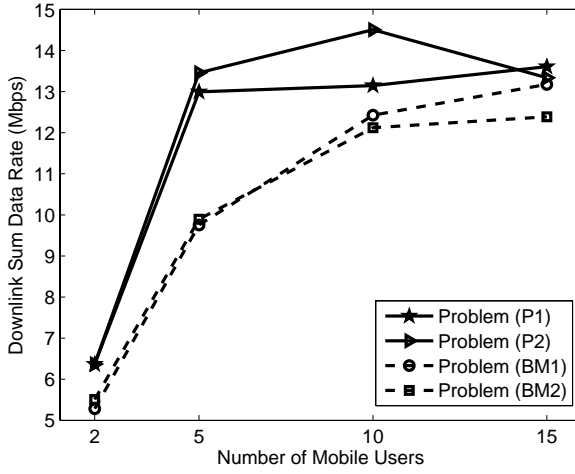


Figure 8.16: DL sum data rate (Mbps) as achieved by solving resource allocation problems (P1), (P2), (BM1), and (BM2) Vs. the number of mobile users.

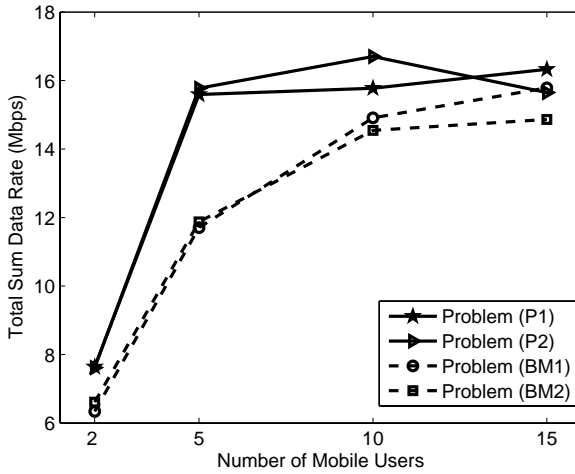


Figure 8.17: Total (UL+DL) sum data rate (Mbps) as achieved by solving resource allocation problems (P1), (P2), (BM1), and (BM2) Vs. the number of mobile users

sufficient capacity for carrying TCP ACKs for the DL and vice versa. Therefore, allocated resource at the RAN will be efficiently utilized by the end-to-end connection. Second, the decoupled DL/UL will provide additional flexibility to the joint resource allocation. Hence, decoupled DL/UL while resources to the two are allocated in coordination with each other and in awareness from the higher layers can provide significant extra efficiency to the end-to-end communications both in terms of latency and data rate.

8.7.2 *Enabling DUDe through Multiple TCP flows*

One of the challenges facing DUDe architecture is the possibility of having multiple TCP flows, that can handle the separate DL and UL connections. One of the major existing efforts in the Internet is the MultiPath TCP (MPTCP) [17]. MPTCP enables TCP to present a regular TCP interface to applications, while several IP interfaces are used simultaneously. In other words, data communication of a flow is spread over multiple *subflows*. The MPTCP connections begin regularly, and if there are extra paths available, additional TCP sessions, termed *subflows*, are created on these paths, and are combined with the existing session, which continues to appear as a single connection to the applications at both ends. Therefore MPTCP will allow for switching between coupled and decoupled DL/UL for different traffic flows. While MPTCP could be a great enabler for DUDe architecture, the challenge here is the availability of two different IP addresses. In other words, MPTCP can only be used to handle the DL and UL, if the UE is multihomed. Example of other efforts in running multiple and parallel TCP flows include MulTCP [29], which behaves as if it was a collection of multiple virtual TCP connections. Such initiative in the Internet domain is often for increasing data rate and better utilization of the end-to-end bandwidth.

8.8 **Conclusions**

Since the inception of cellular networks, the UL connection was always associated with the same base station that has been chosen for the DL reception. In this chapter, this postulate has been revisited and the feature of Downlink/Uplink Decoupling (DUDe) was introduced. DUDe is a new architectural paradigm where UL and DL are not constrained to associate to the same base station and are effectively treated as separate networks. This concept is becoming very relevant with the expected densification in future cellular networks where each terminal will have multiple access point in proximity and allowing the terminal to choose the best UL and DL connection(s) will result in a much more efficient operation of cellular networks. In this chapter, the main challenges faced by heterogeneous networks have been identified and how DUDe can be a solution to some of these challenges. The performance gains from DUDe were also demonstrated in terms of UL capacity, coverage, load balancing and reliability. The interoperability of DUDe with the different new trends in cellular networks has been highlighted. Finally, the realization of DUDe in terms

of architecture has been studied thoroughly, where the different options of implementing DUDe in LTE have been illustrated, with a glimpse of what is needed in 5G networks for DUDe to be natively supported as well as a study on the effects of DUDe on higher layers.

Bibliography

- [1] 3GPP. Coordinated Multi-Point Operation for LTE Physical Layer Aspects. TS 36.819, 3rd Generation Partnership Project (3GPP), September 2012.
- [2] 3GPP. LTE Radio Access Network (RAN) Enhancements for Diverse Data Applications. TR 36.822, 3rd Generation Partnership Project (3GPP), September 2012.
- [3] 3GPP. Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 11). TS 36.300, 3rd Generation Partnership Project (3GPP), September 2012.
- [4] 3GPP. Study on Small Cell Enhancements for E-UTRA and E-UTRAN; Higher Layer Aspects. TR 36.842, 3rd Generation Partnership Project (3GPP), September 2014.
- [5] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and C. Zhang. What will 5G Be? *IEEE Journal on Selected Areas in Communications*, June 2014.
- [6] H. Balakrishnan, V. N. Padmanabhan, G. Fairhurst, and M. Sooriyabandara. TCP Performance Implications of Network Path Asymmetry. *IETF RFC 3449*, December 2002.
- [7] F. Boccardi, R.W. Heath, A. Lozano, T.L. Marzetta, and P. Popovski. Five Disruptive Technology Directions for 5G. *52(2):74–80*, Feb. 2014.
- [8] Federico Boccardi, Jeffrey Andrews, Hisham Elshaer, Mischa Dohler, Stefan Parkvall, Petar Popovski, and Sarabjot Singh. Why to Decouple the Uplink and Downlink in Cellular Networks and How To Do It. 2015. To appear. Available at: <http://arxiv.org/abs/1503.06746>.
- [9] D. Colombi, B. Thors, and C. Tornevik. Implications of EMF Exposure Limits on Output Power Levels for 5G Devices Above 6 GHz. *Antennas and Wireless Propagation Letters, IEEE*, 14:1247–1249, 2015.
- [10] A. Daeinabi, K. Sandrasegaran, and X. Zhu. Performance Evaluation of Cell Selection Techniques for Picocells in LTE-Advanced Networks. In *2013*

- 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*, pages 1–6, May 2013.
- [11] Uwe Dtsch, Mark Doll, Hans-Peter Mayer, Frank Schaich, Jonathan Segel, and Philippe Sehier. Quantitative Analysis of Split Base Station Processing and Determination of Advantageous Architectures for LTE. *Bell Labs Technical Journal*, 18(1):105–128, 2013.
- [12] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer. Downlink and Uplink Decoupling: A Disruptive Architectural Design for 5G Networks. In *2014 IEEE Global Communications Conference (GLOBECOM)*, pages 1798–1803, Dec 2014.
- [13] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer. Load & Backhaul Aware Decoupled Downlink/Uplink Access in 5G Aystems. In *2015 IEEE International Conference on Communications (ICC)*, pages 5380–5385, June 2015.
- [14] ST. Ericsson. Further Discussions on UL/DL Split. Technical Report R2-131678, 3GPP TSG-RAN, May 2013.
- [15] ST. Ericsson. Physical Layer Aspects of Dual Connectivity. Technical Report R1-130566, 3GPP TSG-RAN, February 2013.
- [16] S. Floys, A. Arcia, and J. Iyengar. Adding Acknowledgement Congestion Control to TCP. *IETF Internet Draft*, January 2009.
- [17] A. Ford, M. Handley, and O. Bonaventure. TCP Extensions for Multipath Operation with Multiple Addresses. *IETF RFC 6824*, January 2013.
- [18] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and Wei Yu. Multi-Cell MIMO Cooperative Networks: A New Look at Interference. *IEEE Journal on Selected Areas in Communications*, 28(9):1380–1408, December 2010.
- [19] X. Gu, X Deng, Q. Li, L. Zhang, and W. Li. Capacity Analysis and Optimization in Heterogeneous Network with Adaptive Cell Range Control. *International Journal of Antennas and Propagation*, 2014:10, April 2014.
- [20] I. Guvenc, Moo-Ryong Jeong, I. Demirdogen, B. Kecioglu, and F. Watanabe. Range Expansion and Inter-Cell Interference Coordination (ICIC) for Picocell Networks. In *2011 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–6, Sept 2011.
- [21] Zho Han, Zhu Ji, and K. J. Ray Liu. Power Minimization for Multi-Cell OFDM Networks Using Distributed Non-cooperative Game Approach. In *Global Communications Conference (GLOBECOM)*. IEEE, November 2004.
- [22] Huawei. Handling of UL Traffic of a DL Split Bearer. Technical Report R2-140054, 3GPP TSG-RAN, February 2014.

- [23] Maria Lema, Mario Garcia-Lozano, and Silvia Ruiz. MPR-Aware Scheduler for Carrier Aggregation Transmissions in LTE Uplink. *Wireless Personal Communications*, 84(2):1417–1438, 2015.
- [24] Maria A. Lema, Mario Garcia-Lozano, Silvia Ruiz, and David 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)*, pages 2191–2196, Sept 2013.
- [25] 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*, pages 1–5, June 2014.
- [26] T. Mahmoodi, V. Friderikos, and H. Aghvami. Using Traffic Asymmetry to Enhance TCP Performance. *Computer Networks*, 57(1):317–329, January 2013.
- [27] T. Mahmoodi, V. Friderikos, O. Holland, and H. Aghvami. Cross-layer Optimization to Maximize Fairness among TCP Flows of different TCP Flavors. In *Global Communications Conference (GLOBECOM)*. IEEE, November 2008.
- [28] T. Mahmoodi, V. Friderikos, O. Holland, and H. Aghvami. TCP-aware Resource Allocation Problem in OFDMA based Wireless Networks. In *International Workshop on Cross layer Design (IWCLD)*. IEEE, June 2009.
- [29] P. Mulroy, S. Appleby, M. Nilsson, and B. Crabtree. The use of MulTCP for the delivery of equitable quality video. In *International Packet Video Workshop*. IEEE, May 2009.
- [30] NGMN. Further Study on Critical C-RAN Technologies. Deliverable, NGMN, 2015.
- [31] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP Throughput: A Simple Model and its Empirical Validation. *SIGCOMM Computer Communications Review*, 28(4):303–314, 1998.
- [32] K.I. Pedersen, Yuanye Wang, B. Soret, and F. Frederiksen. eICIC Functionality and Performance for LTE HetNet Co-Channel Deployments. In *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5, Sept 2012.
- [33] Zukang Shen, Jeffrey G. Andrews, and Brian L. Evans. Adaptive Resource Allocation in Multiuser OFDM Systems With Proportional Rate Constraints. *Transactions on Wireless Communications*, 4(6):2726–2737, 2005.
- [34] S. Singh, Xinchun Zhang, and J.G. Andrews. Joint Rate and SINR Coverage Analysis for Decoupled Uplink-Downlink Biased Cell Associations in HetNets. *IEEE Transactions on Wireless Communications*, 14(10):5360–5373, Oct 2015.

- [35] K. Smiljkovikj, P. Popovski, and L. Gavrilovska. Analysis of the Decoupled Access for Downlink and Uplink in Wireless Heterogeneous Networks. *IEEE Wireless Communications Letters*, 4(2):173–176, April 2015.
- [36] W. Richard Stevens. *TCP/IP illustrated, Volume I The protocols*. Addison Wesley, February 2000.
- [37] Hua Wang, C. Rosa, and K.I. Pedersen. Dedicated Carrier Ceployment in Heterogeneous Networks with Inter-Site Carrier Aggregation. In *2013 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 756–760, April 2013.
- [38] Hua Wang, C. Rosa, and K.I. Pedersen. Uplink Inter-Site Carrier Aggregation between Macro and Small Cells in Heterogeneous Networks. In *2014 IEEE 80th Vehicular Technology Conference (VTC Fall)*, pages 1–5, Sept 2014.
- [39] Ian C. Wong and Brian L. Evans. Optimal Downlink OFDMA Resource Allocation with Linear Complexity to Maximize Ergodic Rates. *Transactions on Wireless Communications*, 7(2):962–971, 2008.
- [40] ZTE. Comparison between CP solution C1 and C2. Technical Report R2-132383, 3GPP TSG-RAN, August 2013.