Proceedings of EUCNC 2014, Bologna, IT, June 23-26, 2014 - full paper here

Effective dynamic coordinated scheduling in LTE-Advanced networks

Giovanni Nardini(1), Giovanni Stea(1), Antonio Virdis(1), Dario Sabella(2), Marco Caretti(2)

(1) Dipartimento di Ingegneria dell'Informazione, University of Pisa, Italy
g.nardini@ing.unipi.it, {g.stea, a.virdis}@iet.unipi.it

(2) Telecom Italia, Turin, Italy – {dario.sabella, marco.caretti}@telecomitalia.it

***Abstract*— This paper proposes a dynamic coordinated scheduling strategy among LTE-Advanced cells. We argue that in current networks dynamic coordinated scheduling is practically feasible only at small scales (e.g., three cells), and we discuss how to optimally select which group of cells should be coordinated to reap the maximum coordination benefit, weighing the above choice with architectural considerations, such as cost-effectiveness and latency. Our results show that, if we take a single group of coordinated cells and assume random, unpredictable interference, coordinating *overlapping* cells is the best choice. However, if we assume that neighboring group of cells do coordinate autonomously, the performance is less sensitive to the way we group them, hence architectural feasibility and cost make the difference.**

***Index Terms*—LTE-A, CoMP, Optimization**

I. Introduction

THE Long Term Evolution (LTE) of the Universal Mobile Telecommunication System (UMTS) [1] promises near-ubiquitous coverage and ultra-high bandwidth. Its new developments are referred to by the collective name of LTE-Advanced (LTE-A). LTE-A downlink transmissions are point-to-multipoint, with a base station (or enhanced Node-B, eNB) scheduling several Resource Blocks (RBs) to User Equipments (UEs), at Time Transmission Intervals (TTIs, or subframes) of 1 ms. Neighboring cells use the same frequency band, hence interfere with each other's UEs unless inter-cell *coordination* is assumed. More specifically, Coordinated Scheduling (CS) is a CoMP (Coordinated Multi-Point Transmission and Reception) technique that allows several cells to decide who addresses whom and using what RBs, so that pairs of cell-UEs transmissions can be scheduled concurrently with a tolerable interference. It is well known that CS-CoMP may increase the cell throughput and – by controlling interference – improve the energy efficiency of LTE-A networks. CoMP techniques for LTE-A have received a considerable attention in the last few years. A survey is reported in [9]. CS-CoMP techniques can be either *static* or *dynamic*. In the first case, resources are partitioned into *shared* and *exclusive* subbands, and each cell schedules its mobiles in either subband based on their interference conditions. Accordingly, static techniques boil down to planning the width and position of exclusive subbands in a multi-cell environment, which may result in an efficient (but not flexible) usage of frequency resources. On the other hand, in the dynamic case there is no prior frequency planning: coordination is performed periodically and may take into account traffic load and channel conditions. This can result in a more flexible and adaptive usage of resources, although the drawback is the higher communication overhead required among coordinated entities.

Previous work has been focused mainly on static approaches, such as Partial Frequency Reuse [3] and Soft Frequency Reuse [4] schemes. These techniques are easy to implement but they cannot cope with changes in data traffic. A dynamic coordination scheme was presented in [5], where each cell autonomously pre-assigns RBs to its users, assuming that a certain restriction pattern (i.e., muting of some interferers) is applied to dominant interferers. A central controller gathers per-RB restriction requests from cells, arbitrates them and allocates each RB independently and without any global outlook (e.g., inter-cell fairness). Moreover, each cell has to send a per-RB utility measure to the central controller, which causes high communication overhead and makes it hard to run coordination at fast time scales.

In our previous work on the subject [8] we show that it is practically infeasible to coordinate more than few cells at a time (e.g., three), because of two reasons:

- We cannot expect UEs to be able to report more than one or two highest interferers with the high time granularity needed by the centralized scheduler.
- The size and complexity of the coordination problem grow exponentially with the number of coordinated cells, and become prohibitive already with few of these (e.g., 5-6).

In this paper we present a practical heuristic to solve the coordination problem at a small scale (e.g., three cells). The heuristic is fast, and it can be used at timescales comparable to a single TTI. We show that such small-scale coordination is beneficial, even in a larger-scale environment. We do this by simulating a relatively large-scale network (consisting of 57 cells) where triples of cells coordinate autonomously, and showing that a significant benefit in cell throughput, fairness and energy consumption is attainable. Moreover, we argue that the above coordination benefits do depend on *how* we select the set of coordinated cells, and we compare two such selection schemes: *intra-site* coordination, whereby co-located cells at the same site coordinate themselves, and *inter-site* coordination, where cells not co-located but radiating in the same area coordinate themselves (in the latter case the cells are supposed to be connected by means of a fiber ring). We discuss that coordination benefits have to be weighed against cost and technology considerations, such as the need for additional cabling and data transfer latency.

In the rest of the paper, Section II reports the system model. We describe our coordination algorithm in Section III, and evaluate it in Section IV. We report conclusions in Section V.

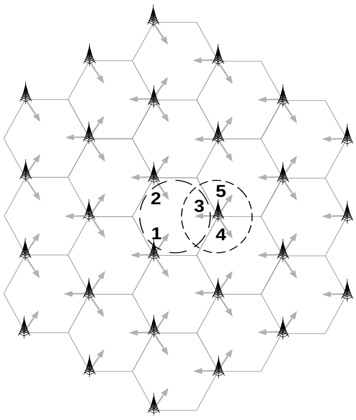

Figure 1 – Network model

II. System Model

We assume that the network is deployed as in Figure 1. Every second vertex of a hexagon hosts a *site*, which groups three cells, each one with a sectorial antenna radiating at an angle of 120°. Therefore, each hexagon has three inward-facing antennas (hence hosts three heavily overlapping cells). UEs are statically associated to one of the cells, but they can measure the level of interference perceived by other two interferers, and they can report to the serving cell the Channel Quality Indicator (CQI) when either or both the two interferers are muted. Coordination may take place:

a) *Intra-site*, i.e. among co-located cells at the same site (e.g., antennas 3,4,5 in Figure 1).
b) *Inter-site*, i.e. among non-co-located cells radiating inward in the same hexagon (e.g., antennas 1,2,3 in Figure 1).

In a distributed-RAN setting, the first approach requires little or no additional expenditure, since co-located cells already share the same site. Moreover, a negligible coordination latency can be expected. The second approach, instead, requires remote antennas to be connected, which is normally done through additional wiring, i.e. using fiber-optics, which adds to the CAPEX and incurs higher latency. In a cloud-based RAN (C-RAN, [2]) architecture, we can assume that all sites are wired through high-speed, low-latency fiber to a central hub, where resource allocation is performed. Hence these solutions are equivalent as for CAPEX and latency.

We denote with A, B, C the coordinated cells (using either coordination approach), and if the cell index is a value taken by variable *x*, then $x = A \Rightarrow (x+1) = B,\ (x-1) = C$. Let $N(x)$ be the number of UEs associated to cell *x*. UEs can be identified by couple *x,j*, where $1 \le j \le N(x)$. Second, given a cell *x*, we use two superscripts to denote the interference from the other two coordinated cells. The first symbol identifies cell *x*-1, whereas the second is for cell *x*+1. Symbol "+" means "active", and "-" is for "inactive". This way, $CQI^t_{x,j}$, where $t \in T = \{++, -+, +-, --\}$, denotes the four possible CQIs for a UE *j* associated to cell *x*: $CQI^{++}_{x,j}$ is the one reported when both *x*-1 and *x*+1 are active, etc. Set *T* represents the four *Interference Subbands (ISs)* for a UE. Our goal is to allocate resources effectively in the above settings.

III. Coordination Algorithm

The coordination problem among three cells can be formulated as an optimization problem [8], which we recall hereafter for the sake of completeness. Afterwards, we discuss its complexity and propose a heuristic solution strategy. Note that both the formulation as an optimization problem and the heuristic are the same irrespective of whether intra-site or inter-site coordination is assumed.

Call $s^t_{x,j}$ the number of RBs given to UE *x,j* within IS *t*. Call $Q_{x,j}$ that UE's buffer level and $r_{x,j}$ the (unique) CQI that will be used for UE *x,j*. Let $b^t_{x,j}$ be a binary variable that is 1 if UE *x,j* has a RB within IS *t*, and zero otherwise. Call M the number of RBs at each cell (i.e., the frame size). Finally, let *R* be a large positive constant (e.g., the largest CQI). The maximum overall throughput is the optimum of the following problem:

$$\max \sum_{x \in \{A,B,C\}} \sum_{j=1}^{N(x)} r_{x,j} \cdot \left(\sum_{t \in T} s^t_{x,j} - p_{x,j} \right)$$

$$s.t.$$

$$r_{x,j} \cdot \sum_{t \in T} s^t_{x,j} \le Q_{x,j} + p_{x,j} \qquad \forall x, j \qquad (i)$$

$$r_{x,j} \le CQI^t_{x,j} + R \cdot \left(1 - b^t_{x,j}\right) \qquad \forall x, j, t \qquad (ii)$$

$$b^t_{x,j} \le s^t_{x,j} \le M \cdot b^t_{x,j} \qquad \forall x, j, t \qquad (iii)$$

$$p_{x,j} \le CQI^t_{x,j} - 1 + R \cdot \left(1 - b^t_{x,j}\right) \qquad \forall x, j, t \qquad (iv)$$

$$\sum_{t \in T} \sum_{j=1}^{N(x)} s^t_{x,j} + \sum_{j=1}^{N(x-1)} s^{--}_{x-1,j} + \sum_{j=1}^{N(x+1)} s^{--}_{x+1,j} + \max\left\{ \sum_{j=1}^{N(x-1)} s^{-+}_{x-1,j},\ \sum_{j=1}^{N(x+1)} s^{+-}_{x+1,j} \right\} \le M \qquad \forall x \qquad (v)$$

$$\sum_{x} \left(\sum_{j=1}^{N(x)} s^{--}_{x,j} + \max\left\{ \sum_{j=1}^{N(x-1)} s^{-+}_{x-1,j},\ \sum_{j=1}^{N(x+1)} s^{+-}_{x+1,j} \right\} \right) + \max\left\{ \sum_{j=1}^{N(x-1)} s^{++}_{x-1,j},\ \sum_{j=1}^{N(x)} s^{++}_{x,j},\ \sum_{j=1}^{N(x+1)} s^{++}_{x+1,j} \right\} \le M \qquad (vi)$$

$$b^t_{x,j} \in \{0,1\}, \quad s^t_{x,j} \in \square^+ \qquad \forall x, j, t \qquad (vii)$$

$$r_{x,j}, p_{x,j} \in \square^+ \qquad \forall x, j \qquad (viii)$$

In the objective, the rate of each UE *x,j* is multiplied by the RBs allocated to that UE, whatever their IS. Moreover, variable $p_{x,j}$ denotes the padding, which is to be discounted from the objective. Constraint (*i*) states that UEs can transmit up to the length of their buffer, plus some padding. The latter must be counted in as well, otherwise buffers would never be emptied. Constraint (*ii*) limits the rate to the minimum CQI among all the ISs where a UE is allocated RBs. For example, if a UE is given both RBs with zero interference (i.e., $b^{--}_{x,j} = 1$) *and* with interference from both neighbors (i.e., $b^{++}_{x,j} = 1$), it will use the *smallest* CQI, i.e., $r_{x,j} = CQI^{++}_{x,j}$. Since *R* is a large positive constant, (*ii*) holds if $b^t_{x,j} = 0$, hence the CQIs of the ISs where the UE is not allocated RBs do not count as a limit. Constraint (*iii*) ensures that $b^t_{x,j} = 0 \Rightarrow s^t_{x,j} = 0$ and $b^t_{x,j} = 1 \Rightarrow s^t_{x,j} \ge 1$ for consistency. Constraint (*iv*) limits the padding to strictly less than one RB at the chosen rate, *R* being used in the same role as in constraint (*ii*). Constraint (*v*) states that cell *j*'s frame should be large enough to include all the RBs allocated to its UEs *x,j*, whatever their ISs *t*, and still have enough room for the RBs allocated to UEs of *other* cells, which require cell x to be muted in those RBs. This last term can be further split into: the RBs where the other cells require exclusive transmission (i.e., those two with a $--$ superscript), and RBs where other cells require only x to be muted (i.e., those in the *max* bracket), which can overlap. For example, the RBs allocated by cell A should be those where:

- A's UEs transmit;
- B requests *both* A *and* C not to transmit;
- C requests *both* A *and* B not to transmit;
- B requests A not to transmit (whereas C may transmit);
- C requests A not to transmit (whereas B may transmit).

The maximum of the last two terms is taken, instead of their sum, to allow overlapping. An example of frame structure for three cells A, B, C is shown in Figure 2. Constraint (*vi*) forces RBs where muting of one or two cells apply to occupy the *same positions* in the three frames.

This is a mixed integer-nonlinear problem (MINLP), with $O\left(K \cdot N \cdot |T|\right) = O\left(K \cdot N \cdot 2^K\right)$ variables and constraints. Nonlinearity is due to the products in the objective function and constraint (*i*). MINLPs are NP-hard in general, hence it is hardly feasible to attempt to solve this problem in a TTI's time, even when the number of UEs is small (i.e., 10-20).

For this reason, we devise a simpler and faster heuristic, which relies on eNBs to do part of the work, so as to make the coordination task easier. That task is actually performed by a *master node* (MN): the latter may be co-located at one of the coordinated nodes (e.g., in a distributed-RAN architecture), or it may be running in the cloud (in a C-RAN architecture).

The heuristic is described as follows (see also Figure 3):

Step 1: each of the three cells makes a *provisional schedule* of its UEs in isolation (i.e., without considering coordination constraints). In doing so, the cell selects which UEs are to be scheduled, and in which IS. A UE is matched to a protected IS only if this warrants a *significant* throughput increase: more specifically, single muting is only allowed in return for a *doubling*, and double muting for a *tripling* of the UE throughput. This ensures that no illicit overbooking of the protected ISs happens. Then, cells make their schedule assuming that no constraint exists on the width of the ISs (except that the frame has a finite size M). The provisional schedule at cell x is a *list of IS requests*, i.e. integers $S_x^t = \sum_{j=1}^{N(x)} s_{x,j}^t$. These are communicated to the MN.

Step 2: the MN receives the *IS requests* from its three cells. It checks whether they are mutually compatible, by testing constraints (*v*-*vi*). If they are not, then it reduces proportionally the requests until they are. It then splits the common frame into ISs, and sends back to the cells the list of *granted* ISs, specifying which cells transmits when. Note that, in doing so, it may actually *increase* the width of an IS beyond the request of a cell. If, for instance, $S_B^{-+} > S_C^{+-}$, and both the requests can be met, a number of $S_B^{-+} - S_C^{+-}$ RBs will be in *double muting*, since neither A (which is muted) nor C (which only needs S_C^{+-}) are using them (see also Figure 2). Hence, the actual width of S_B^{--} will increase accordingly.

Step 3: cells can now perform the *actual* scheduling, since they have a mutually consistent view of who transmits where. Note that, since ISs are in general larger/smaller than initially requested, the actual scheduling will differ from the provisional one. Some UEs may find themselves promoted to a more protected IS, others may find demoted to a less protected one. However, if the provisional and actual scheduling algorithms are the same, this is not a problem.

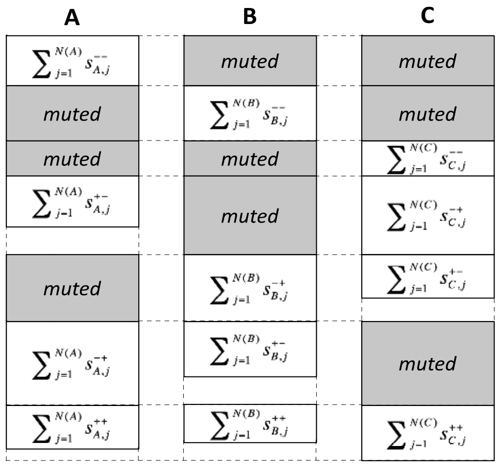

Figure 2 – Frame structure for three coordinated cells

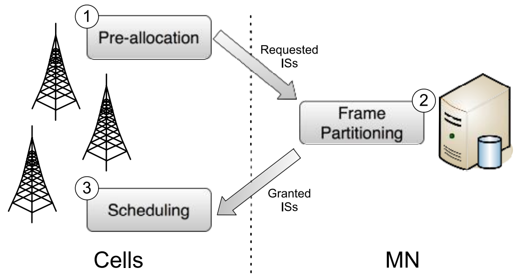

Figure 3 – Heuristic algorithm

First of all, we observe that the complexity of the heuristic is affordable: cells are requested to do no more than their scheduling job, and the MN only has to check a small number of straightforward inequalities. Moreover, the information being exchanged between the cells and the MN is very limited and does not depend on the number of UEs or the traffic load. For this reason, it can be run *dynamically, at a TTI timescale*.

Note that this coordination scheme does not require cells to use a particular scheduling algorithm: as long as they can partition UEs among ISs, scheduling can be done according to any criterion (e.g., Proportional Fair, Max C/I, time-based priority, etc.). It is also worth noting that this scheme allows a last word (i.e., step 3) to the cells, which is done on purpose to achieve scheduling consistency: suppose, in fact, that the PF criterion would select UE j at node x to be scheduled in the double muting IS, since it is just above the tripling threshold. If the MN reduces the double muting IS for x, we do want to allow cell x to choose whether to schedule j in some other IS (e.g., either of the single muting ones) rather than having to drop it altogether. Similarly, if the double muting IS is larger than expected, we still want x to be in control of which additional UEs will be promoted to it (e.g., starting from those nearer to the tripling threshold).

Comparing the described solution with the dynamic algorithm in [5], we observe that the latter requires more information to be exchanged between cells and the central controller, since M triples {muting pattern, utility value, RB} have to be sent, whereas our algorithm only transmits IS widths. Furthermore, the MN of our scheme assembles the frame so as to accommodate conflicting requests, reducing them proportionally when necessary. Instead, when cells request mutual restriction on the same RB, the central controller in [5] selects the highest-utility cell *for that RB*, and never modifies the initial allocations made by cells. This has several unwanted consequences: for instance, if three cells all request *the same* RB where the other two are muted, our algorithm will allocate three double-muting RBs, whereas the one in [5] will allocate only *one* RB, to the cell with the highest utility. This may underutilize the bandwidth or cause long-term unfairness.

IV. PERFORMANCE EVALUATION

Our simulations are run using SimuLTE [6], a system-level simulator based on the OMNeT++ framework [7]. The intra- and the inter-site scenarios under examination are depicted in Figure 4. UEs are randomly deployed only for the central triple, whereas surrounding triples only produce interference on UEs served by the three central cells. When coordinating the three central cells, we consider two types of interference from the surrounding triples. In Type-1, cells generate power on randomly chosen RBs, which vary from one TTI to the next. In Type-2, the frame partitioning generated by the coordination algorithm in the central triple is replicated on each triple of the surrounding tiers, applying a random perturbation to the width of the ISs and shifting their position by a random quantity. This way, we model a scenario where instances of our algorithm are run independently in all triples. In both cases, the bandwidth occupation is kept similar to that in the central triple. UEs are static and equally distributed among cells. The attenuation pattern of sectorial antennas is: $A(\theta)=\min\{12\cdot(\theta/70^{\circ})^{2}, 25\}$, θ being the angle between cell and receiver. For the inter-site case, we assume ideal communication through the X2 interface (null latency and infinite bandwidth). Simulation parameters are reported in Table 1. VoIP traffic is considered, with the parameters of Table 2. Statistics are gathered at the three central cells.

In Figure 5, we show the cumulative distribution function (CDF) of the MAC-level user throughput, with 50 UEs per cell. Since CDFs with Type-1 and Type-2 interference are overlapping, we show only one. As expected, fairness among UEs increases dramatically when coordination is employed, for both intra- and inter-site scenario. We also report an ideal baseline, obtained by sending the same VoIP traffic on a lightly loaded wired Gb link. The CDF obtained with coordination almost matches that one. Figure 6 reports the CDF of user throughput with 75 UEs per cell. When Type-1 interference is assumed, inter-site coordination performs better than the intra-site one, and results are similar with Type-2 interference. Note that in the latter case cell-edge UEs have higher throughput than in the former case, where the random interference is not predictable. Since intra-site coordination leaves out the major interferers (which are, in fact, cells radiating inward in the hexagons), such unpredictability degrades performance more than in the inter-site case. However, when the coordination algorithm is run in every triple (albeit independently), the position of each IS is less subject to oscillation in the frame, making it simpler to estimate the interference from surrounding triples. This can be observed in Figure 7, where the average error rate with Type-1 interference is higher than with Type-2 at high loads. Figure 8 shows the average number of RBs allocated by one cell. We observe that fewer RBs are used when inter-site coordination is employed. Benefits over the uncoordinated case are clear in both scenarios.

Table 1 – System parameters

Parameter	Value
Cellular layout	Hexagonal grid
Inter-site distance	500 m
Carrier frequency	2 GHz
Bandwidth	10 MHz
Number of RBs	50
Path loss model	Urban Macro [10]
Fading model	Jakes (6 tap channels)
eNB Tx Power	46 dBm
Scheduler	Max C/I
Simulated time	100s

Table 2 – VoIP model parameters

Talkspurt duration (Weibull dist.)	Shape 1.423, Scale 0.824
Silence duration (Weibull dist.)	Shape 0.899, Scale 1.089
Codec Type	GSM AMR Narrow Band (12.2 kbps) w. VAD
VAD Model	One-to-one conversation
Header Compression	Active (RTP+UDP+IP headers = 6 bytes)
Packet length	32 bytes/frame + 6 bytes Hdr + 1 byte RLC

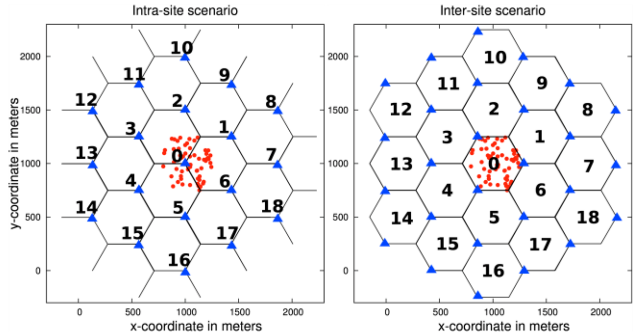

Figure 4 – Simulation scenarios

Figure 9 presents the average of the frame partitioning, i.e. the output of the second step of the coordination algorithm. These results refer to the scenario with Type-2 interference. Denoting with A, B, C the cells of the central triple, each IS is identified by the name of the cells that are enabled in that IS. For example, AB represents the IS in which C is muted, whereas A represents the IS in which both B and C are muted. It can be observed that the coordination algorithm enforces mutual exclusion among cells as much as possible. In the intra-site scenario, single-muting ISs are wider than in the inter-site case. In the former case a UE hardly ever perceives high interference from *both* the other two coordinated cells, thus double muting is less effective than in the inter-site scenario, where double muting ISs protect a UE from its major interferers. This is shown in Figure 10, where we represent the Signal to Interference and Noise Ratio (SINR) that a UE perceives from its serving cell when double muting is enforced. SINR increases going from red to green. The area with good SINR is wider when muting is applied in the inter-site case (Figure 10, right) than in the intra-site case (Figure 10, left).

V. CONCLUSIONS AND FUTURE WORK

We presented a low-complexity heuristic to solve the coordinated scheduling problem among triples of cells. It relies on cells doing part of the work, which reduces the amount of information to be exchanged between computation entities. We showed that the choice of the triple of cells makes a difference in an *uncoordinated* environment: in that case, selecting overlapping cells (*inter-site* CoMP) is considerably better than selecting cells at the same site (intra-site CoMP), although more expensive in terms of CAPEX and latency. Quite surprisingly, in a *coordinated* environment, the two choices perform the same (although inter-site CoMP still requires fewer RBs to accomplish the same task). Future work will include leveraging small-scale coordination to achieve larger-scale coordination, in a layered approach.

VI. ACKNOWLEDGEMENTS

The subject matter of this paper includes description of results of a joint research project carried out by Telecom Italia and the University of Pisa. Telecom Italia reserves all proprietary rights in any process, procedure, algorithm, article of manufacture, or other result of said project herein described.

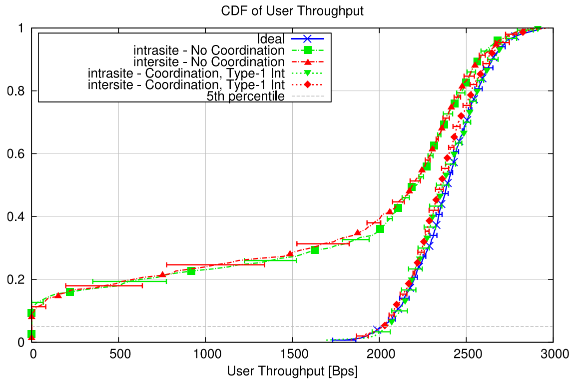

Figure 5 – CDF of UE throughput, 50 UEs per cell

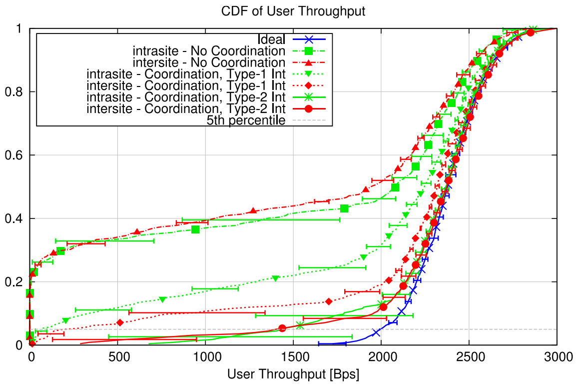

Figure 6 – CDF of UE throughput, 75 UEs per cell

Figure 7 – Average H-ARQ Error Rate

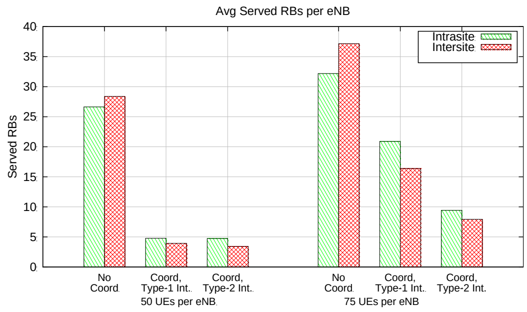

Figure 8 – Average number of allocated RBs per cell

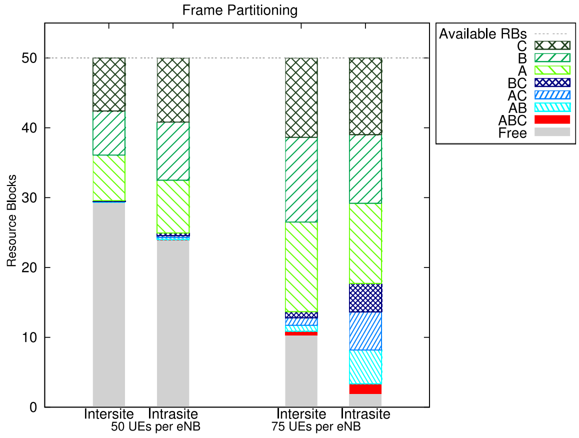

Figure 9 – Frame Partitioning

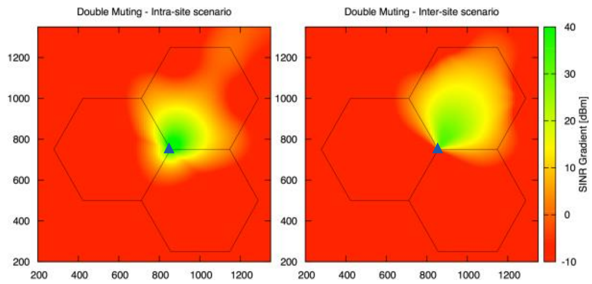

Figure 10 – SINR distribution

REFERENCES

[1] 3GPP - TS 36.300 Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2.

[2] C-RAN – The Road Towards Green RAN, v. 2.5 white paper, China Mobile Research Institute, Oct. 2011.

[3] M. Sternad, T. Ottosson, A. Ahlen, A. Svensson, "Attaining both coverage and high spectral efficiency with adaptive OFDM downlinks," proc. VTC 2003-Fall, pp. 2486-2490, 6-9 Oct. 2003.

[4] 3GPP, "Soft Frequecy Reuse Scheme for UTRAN LTE," 3rd Generation Partnership Project (3GPP), Project Document R1-050507, May 2005.

[5] M. Rahman and H. Yanikomeroglu, "Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination," IEEE Trans. on Wireless Communications, vol. 9, pp. 1414-1425, April 2010.

[6] SimuLTE webpage. http://www.simulte.com.

[7] OMNeT++, http://www.omnetpp.org

[8] G. Nardini, G. Stea, A. Virdis, M. Caretti, D. Sabella, "Improving network performance via optimization-based centralized coordination of LTE-A cells", proc. CLEEN 2014, Istanbul, TK, April 6, 2014.

[9] E. Pateromichelakis, M. Shariat, A. ul Quddus, and R. Tafazolli, "On the Evolution of Multi-Cell Scheduling in 3GPP LTE / LTE-A", IEEE Communications Surveys and Tutorials, 2012.

[10] 3GPP TR 36.814 v9.0.0, "Further advancements for E-UTRA physical layer aspects (Release 9)," March 2010.