SHARING

SELF-ORGANIZED HETEROGENEOUS ADVANCED RADIO NETWORKS GENERATION

Deliverable D4.1

New opportunities, challenges and innovative concept candidates for SON/Heterogeneous Networks

Date of delivery	11/07/2014
Contractual date of delivery	30/06/2014
Project number	C2012/1-8
Editor(s)	Kimmo Hiltunen (ERICSSON)
Author(s)	Name Surname (company acronym) Sofia Martinez-Lopez (FT), Yasir Khan (FT), Yolanda Fernandez (TTI), Adrian Sanchez (TTI), Grégory Gougeon (SIR), Kimmo Hiltunen (ERICSSON), Edgar Ramos (ERICSSON), Engin Zeydan (AVEA), Mourad Khanfouci (MERCE), Mehdi Bennis (UOULU), Mohamad Assaad (SUPELEC), Yves Louet (SUPELEC)
Dissemination level	PU
Workpackage	4
Version	V0.1
Total number of pages	130

Abstract:

This deliverable provides an initial view on new opportunities, challenges and innovative concepts for SON and/or heterogeneous network deployments. In addition to the concept descriptions, this deliverable will present some initial evaluation results, where available.

New concepts are presented within the areas of: intra-system radio access offloading, inter-system radio access offloading, energy saving mechanisms and spectrum resource allocation. In addition to presenting new concepts, it will also discuss the performance and deployment strategies of heterogeneous network deployments within various scenarios.

Keywords:

Offloading, Balancing, Performance, Heterogeneous Networks, RRM, SON, WiFi, Backhaul, Energy-Efficiency, Spectrum, Management, Algorithm

Document Revision History

Version	Date	Author	Summary of main changes
0.1	04/04/2014	Kimmo Hiltunen	Initial draft
1.0		Kimmo Hiltunen	Final version

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This deliverable provides an initial view from SHARING Work Package 4 "Ressource optimization for Heterogeneous Networks" on new opportunities, challenges and innovative concepts for SON and/or heterogeneous network deployments. In addition to the concept descriptions, this deliverable will present some initial evaluation results, where available.

New concepts are presented within the areas of: intra-system radio access offloading, inter-system radio access offloading, energy saving mechanisms and spectrum resource allocation. In addition to presenting new concepts, it will also discuss the performance and deployment strategies of heterogeneous network deployments within various scenarios.

In order to address the rapidly increasing data traffic volumes, one efficient way to enhance the system performance is to densify the network deployment, either by deploying new macro cells, or with the help of heterogeneous network deployments. However, together with the network densification, several new challenges appear, related to for example the inter-cell interference orthe overall cost of deployment. Therefore, several techniques and deployment strategies have been identified, allowing the network to cope with localized peak traffic loads and optimizing the utilization of network resources. In case of intra-system radio access offloading (SHARING WP4 Task 4.1) the presented innovations and concepts are related to (i) SON-based load balancing and inter-cell interference management (via adjustments of eNodeB transmission powers, antenna tilt angles, handover margins, eICIC muting ratios and biased cell selection), (ii) load balancing with the help of large scale antenna systems or middleware deployment, (iii) mobility management between macro and low-power nodes, and (iv) backhaul offloading via proactive data caching. As indicated by the initial results, the presented innovations are indeed able to enhance the system performance both in terms of coverage and capacity, which can be expected to contribute to reduce CAPEX and OPEX for the operators.

An efficient and cost-effective integration of cellular and WiFi technologies, referred to as inter-system offloading has recently attracted significant interest from academia, industry, and standardization bodies alike. On one hand, the inherent constraints of small cell networks, particularly due to cross-tier and co-tier interference, motivate offloading some of the traffic to the WiFi band to alleviate interference and ease congestion. On the other hand, due to the uncontrolled and unlicensed nature of WiFi, the competition for resources among a large number of hotspot users can yield dramatically poor throughput. In such a scenario, offloading some of this traffic to a well-managed small cell network operating over the licensed spectrum can improve the performance.

In SHARING Work Package 4 Task 4.2, smart traffic offloading and steering strategies taking into account users' QoS requirements, indoor/outdoor locations, terminal speed, users' contexts, interference levels, and network loads will be investigated. Both UE-assisted and network-assisted offloading scenarios taking into account network parameters such as latency, reliability, user traffic and mobility patterns, network capacity, congestion, content type, packet loss, channel characteristics, will be explored. A decentralized and dynamic traffic offloading framework, in which small cells seamlessly steer their traffic between cellular and WiFi RATs, will be proposed. Of interest is quantifying the improvements in terms of cell throughputs and cell-edge throughput, for the "macro-only", "HetNet," "WiFi-only" and "HetNet+WiFi" offloading strategies as described in D2.2. Task 4.2 will also study efficient WiFi-LTE interconnections for fast handover and propose traffic steering strategies between LTE and WiFi, leveraging the prediction of QoS and power consumption on WiFi and LTE enabling QoS-aware and energy-efficient load balancing. Task 4.2 will also look into aspects of load balancing, procedures for interference control for multi-RAT, inter-LTE offloading to WiFi by middleware deployment, mobility management, backhaul-aware load balancing, and UE-specific load balancing policies, for providing seamless and transparent QoS. The initial evaluation results presented in this deliverable indicate that the proposed inter-system traffic offloading mechanisms are able to improve both the user quality and the overall system capacity.

Energy saving strategies have become one of the key issues in network deployments. Currently deployed base stations need to permanently signal their presence and availability, and continuously listen to the radio environment to detect incoming users even during the low-traffic hours (e.g. nights). At the same time, the massive large scale deployment of small low-power nodes can be ecologically worthwhile only if armed with smart energy efficiency and power saving mechanisms. Motivated by this, energy-efficient sleep mode strategies for small cell base stations are required in order to reduce the power consumption of cellular networks. These strategies allow the hardware components in the base station to be smartly switched off in idle conditions, such that the energy consumption is modulated over the variations in traffic load and space and is carried out across all available tiers. Furthermore, in a deployment scenario with a large number of overlaid cells, serving cell selection can be one key mechanism to reduce the number of activated cells. Indeed, in this case, the classical approach where the mobile station is connected to the strongest access point is no longer an optimal strategy in terms of interference and energy efficiency. On the other hand, due to the large number of overlaid cells, optimizing the serving cell selection becomes an issue. One solution lies in enabling a dynamic activation of local access points limiting the load of the buiser ones. A possible improvement is to leverage on predictive RRM aspects, based on the estimation of traffic evolution offering the required amount of capacity in both time and space. As a result, SHARING Work Package 4 Task 4.3 intends to investigate algorithms that focus on self-organizing mechanisms to minimize the energy consumption of the radio access network. For example, Task 4.3 aims to develop and investigate energy saving sleep mode strategies and to investigate the network performance when self-organizing energy-saving mechanisms are combined with enhanced RRM mechanisms, such as novel cell association techniques for small cells, cell range expansion, and enhanced-ICIC. Furthermore, Task 4.3 studies mechanisms to save energy in the transmitter analog front end, especially regarding the base station power amplifier. The presented simulation results indicate clearly that the proposed innovations result in reduced energy consumption. However, further work is still needed to find out the real potential of the proposed energy saving mechanisms.

SHARING Work Package 4 Task 4.4 studies various self-organizing techniques for radio resource allocation, taking into account carrier aggregation features and capabilities of the future networks in the context of heterogeneous network deployments. For example, the task will develop distributed and centralized radio resource management strategies that ensure the stability of the network in the context of bursty traffic. Furthermore, the task will develop a learning framework in order to improve the performance of RRM strategies.

Task 4.4 will also investigate an interference management mechanism based on dynamic carrier selection in LTE-A systems with carrier aggregation. This interference management is provided by SON-flavoured algorithms for the selection of primary and secondary component carriers for a group of base stations. The component carrier selection will be based on UE measurements and statistical information about the usage of the component carriers. Furthermore, the topic of dynamic component carrier selection will be discussed in the context of multi-flow carrier aggregation, where UEs aggregate separate flows coming from different nodes. Finally, tight time synchronization is necessary for performing advanced RRM. Therefore, the task will propose a cooperative distributed synchronization mechanism, which will improve the basic time reference provided by the macro network.

1 INTRODUCTION

Traffic in the mobile networks is expected to grow very rapidly in the coming years [Cis14][Eri13]. This traffic growth will be caused both by the evolution of mobile terminals (an increasing penetration of smartphones, tablets and mobile computers) and the increased use of more traffic-heavy services, especially video. It is also expected that the wider introduction of various cloud-based services and machine-to-machine communication will accelerate the traffic growth even further.

In general, there are three possible ways to increase the capacity of a mobile network: increased spectrum, improved spectral efficiency and network densification. From the spectrum point of view, the capacity can be enhanced by deploying additional carriers, or by increasing the carrier bandwidth. The spectral efficiency can be improved both by improving the signal-to-interference-ratio for the link between the transmitter and the receiver and by introducing new techniques to enhance the utilization of the high signal-to-interference-ratio conditions. These techniques include for example advanced multi-antenna techniques (for example MIMO and beamforming), higher order modulations and advanced interference management (for example interference cancellation, inter-cell interference coordination and coordinated multipoint transmission and reception). However, although the achievable capacity gains via additional spectrum, and improvements in spectral efficiency are considerable, the substantial growth that is predicted for the mobile broadband revolution will require also actions to densify the mobile networks, i.e. to increase the spatial reuse of the radio resources.

The traditional way to densify mobile networks has been to deploy new macro cells, either by adding new sectors to existing sites, or by deploying new macro sites. The benefit of a densified macro deployment is that the network performance can be improved with a relatively small amount of required new hardware, or new sites. However, as new macro sites are becoming increasingly difficult and often expensive to deploy, at least within urban environments, focus is put on the efforts to find more cost-efficient ways to densify the current networks.

An alternative to deploying new macro sites is to deploy low-power sites within traffic hotspots, i.e. the introduction of heterogeneous network deployments. In case of the heterogeneous network deployment, macro cells will provide wide area coverage, while the small low-power cells deployed within traffic hotspots will take care of the majority of the traffic volume. The downside of heterogeneous network deployments compared to the densified macro deployments is that a considerably larger number of new cells is required to be able to offer the same system performance. Even though the cost of a low-power site will typically be lower than the cost of a macro site, the overall situation may turn out to be quite challenging from the total network cost point of view.

The total cost of the network deployment consists of a large number of different components, related to both the capital expenditure (CAPEX) and the operational expenditure (OPEX). Examples of the CAPEX-related components include the base station equipment, site equipment and network roll-out. The OPEX-related components include for example site rental, energy, and operation and maintenance. The total cost of the network deployment can be reduced for example by improving the cell capacity and coverage, by introducing various types of SON functions, and by introducing different kinds of energy saving mechanisms.

SHARING Work Package 4 will look into many of the topics listed above. Task 4.1 will investigate ways to perform intra-RAT traffic offloading as efficiently as possible, including topics such as load balancing and interference management, mobility management, and backhaul offloading. Task 4.2 will evaluate the impacts of inter-RAT traffic offloading, such as traffic offloading between LTE and WiFi. Task 4.3 will study different ways to reduce the energy consumption, both via RRM reconfiguration and hardware optimization. Finally, Task 4.4 will consider different ways to enhance the system performance with the help of spectrum resource allocation mechanisms.

This deliverable presents an initial view on new opportunities, challenges and innovative concept candidates for SON and heterogeneous network deployments. Chapters 2-5 present concepts from all four tasks of SHARING WP4. Furthermore, Chapter 6 discusses the performance and deployment strategies of heterogeneous network deployments within different scenarios.

2 INTRA-SYSTEM RADIO ACCESS OFFLOADING

In order to address the fast data traffic boost experienced in today's mobile cellular networks, several techniques and deployment strategies have been identified, allowing the network to cope with localized peak of traffic load and optimizing the utilization of network resources. These techniques are known as traffic offloading, and in the context of SHARING Task 4.1 they are carried out within the same radio access technology.

There are several areas that can be targeted with these kind of innovations related to traffic offloading. The first group would belong to optimization of mobility management parameters in order to adjust the cell sizes and the intended coverage areas. This is useful in particular for the heterogeneous network deployments, where the cells might have overlapping coverage areas aimed at providing enhanced coverage and capacity within specific areas. The network would then decide how to move traffic between the overlapping cells in order to maximize performance, capacity, resource utilization as well as energy efficiency, which are becoming more important nowadays. These factors are applicable also for the introduction of new and less traditional deployments and their interactions and coexistence with legacy systems, where traditional mobility procedures, channel estimation and link adaptation might not be applicable. The second group of innovations aims to handle the capacity and performance requirements by managing the available spectrum in order to reduce the interference between neighbouring transmission points and at the same time maximize the system performance. Finally, the third group of innovations targets the optimization of the use of the backhaul connection.

The heterogeneous nature of the deployments targeted by the innovations has inherent challenges that have to be addressed. The first of those challenges is the power imbalance between the overlapping cell layers. The difference in transmission power between the small cells and the macro cells has an impact on the perceived interference between the cells that affects the power control algorithms, the cell selection procedures and indirectly the control channel quality and channel estimation. Additionally, some of the regularly used algorithms to optimize the mobility parameters in macro environments are not possible to be scaled down for small cell environments, such as calculation of the UE speed and estimation of the UE location. Another challenge related to the power imbalance is the actual detection of the small cells, especially when the cell layers are operating on different carrier frequencies.

Yet another aspect of the heterogeneous network deployments is their flexibility, meaning that some deployments might be used "when required" to assist the network to provide the desired capacity or performance. Due to this flexibility the environment and the performance have to be monitored in different parts of the network in order to provide the decision bases to network controllers to activate or deactivate features and deployment options. The collection of the required KPIs, measurements and activity in the network is itself an additional challenge together with the actual organization and processing of the data that in many cases is distributed and might accumulate large volumes of information and the final execution of the configuration decision, especially challenged in distributed environments. It is important to keep in mind that the decisions to modify the network setup will have a direct impact on the balance between network performance, capacity and energy consumption.

2.1 SON-based Load Balancing and Interference Management

All the innovations presented in this section are based on the use of an iterative self-optimization process. The process consists of a self-optimization block and a network simulation block. Depending on the innovation in question, the self-optimization process is utilized to optimize the eNodeB transmission powers, antenna tilt angles, cell selection offsets and related eICIC muting ratios, or the handover parameter settings.

2.1.1 Load balancing by deploying power optimization in LTE macro cell networks

2.1.1.1 Introduction

A typical challenge for an operator is the improvement of QoS in a congested region of the network. While increasing capacity by deploying small cells is a long term solution to alleviate congestion, however, short term QoS objectives such as the amelioration of services during an open air concert, or rush hour radio traffic congestion may not call for a more permanent solution such as a deployment of small cells. Use of Load Balancing (LB) with neighbouring cells can in such cases be an effective solution to improve the QoS in congested cell by optimizing base station parameters such as pilot power and antenna tilts to change the effective cell boundary/coverage thereby forcing the excess mobiles to connect to the less loaded neighbours. Load balancing aims at offloading traffic from one or several congested cells to less loaded neighbouring cell(s), thus increasing the overall capacity and enhancing end-user QoS. LB can be carried out by adjusting either the cell coverage or the mobility parameters [RAM11]. The former involves changes to antenna tilt or variations in pilot power to increase/decrease the coverage area of the cell, whereas the latter involves tuning the handover or cell-reselection parameters to trigger handovers closer or further away from the initial handover point.

A pilot power based load balancing is an automated self-optimization concept, which not only improves the QoS based on the KPI requirements but also improves the OPEX and CAPEX savings by taking manual operations of parameter adjustments out of the optimization loop.

Load balancing through cell coverage adjustments on the other hand, has been thoroughly studied for legacy systems [JLI05], [ISOM04] and lately for enterprise femtocell networks [MOH12]. Ensuring a "well-covered" network without coverage issues, pilot power and thus total transmit power adjustments can serve as a powerful tool to enhance the overall capacity and user QoS by balancing the traffic between the cells [RAM11]. Note that cell coverage adjustments through pilot power adjustments are accompanied by corresponding proportional modifications in traffic channel transmit power in order to maintain the coherence between coverage and capacity.

2.1.1.2 Solution description

Let us assume a cluster of M cells in a network. A typical load balancing scenario consists of identification of one or several congested cells (with load beyond a specific threshold) within a central high traffic (hotspot) zone G_1 and their neighbouring surrounding cells within a normal traffic zone G_2 which are less loaded. Load balancing based on pilot power optimization aims at reducing the loads of the congested cells by transferring/offloading traffic to (less loaded) neighbouring cells, thus increasing the loads of the neighbouring cells. In accordance with this goal, the self-optimization algorithm balances/equalizes the loads of the congested cells and the neighbouring cells. It performs this optimization by adjusting each of the individual pilots and thus total transmit powers, and proposes a vector of total transmit power combination.

The iterative self-optimization process proposed in this work involves a simulation block, which accepts the individual pilot power settings as an input and computes the corresponding observations (KPIs) as shown in Figure 1. The objective function constitutes a data point and acts as an input to the self-optimization block. The self-optimization block consists of:

- Model sub-block, which takes in the data point, appends it to the existing data set and updates the Kriging model.
- Optimization sub-block, which uses the real relationships (surrogates) and uses a sequential optimization technique called search and poll algorithm to obtain the optimum pilot power settings.

The optimum pilot power settings are then fed into the simulator block to obtain the next data point in an iterative manner.

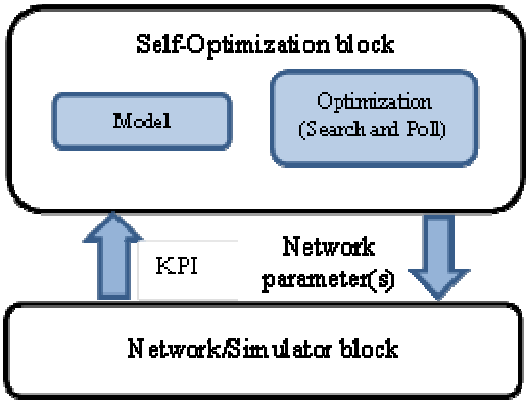


Figure 1. A centralized self-optimization proposal

Optimization objective

Denoting the total transmit power of cell s as P_s (dBm) and the vector of total transmit powers by $\mathrm{x} = [\mathrm{P}_1\ \mathrm{P}_2\ \mathrm{P}_3 \dots \mathrm{P}_\mathrm{M}]^\mathrm{T}$, respectively, the objective function of load balancing can be written as:

$$\begin{aligned} x^* &= \arg\min_x f(x) \\ s.t.\ \ P_{min} &\le P_s \le P_{max}\ \ \forall s \in G_1 \cup G_2 \end{aligned} \tag{1}$$

where $\mathrm{f(x)} = \max[\mathrm{load}_1(\mathrm{x}), \dots, \mathrm{load}_\mathrm{s}(\mathrm{x}), \dots \mathrm{load}_\mathrm{M}(\mathrm{x})]$, $\mathrm{load}_\mathrm{s}(\mathrm{x})$ is the load of cell s $\forall s \in G_1 \cup G_2$ and P_{min}, P_{max} are the minimum and the maximum allowable total transmit power values to prevent capacity and coverage issues such as coverage holes and pilot pollution.

Kriging model

Let Y denote the variable (KPI and/or a function of KPIs) that we would like to model, i.e. $f(\mathbf{x})$. The aim of Kriging is to find a stochastic model for $Y(x)$, which can be expressed as follows [PICH13],

$$Y(x) = \mu + Z(x) + \epsilon(x) \tag{2}$$

where $\mu \in \mathbb{R}$ is the unknown but deterministic constant trend, $\mathrm{Z(x)}$ is the underlying stochastic model which is a Gaussian process with zero mean and anisotropic covariance kernel $\phi(\mathrm{x}, \mathrm{x}')$, $\forall \mathrm{x} \in \mathbb{R}$. $\epsilon(\mathrm{x})$ is the additive noise term with zero mean and covariance kernel $\tau^2\delta(\mathrm{x}, \mathrm{x}')$, $\forall \mathrm{x} \in \mathbb{R}$, $\delta(\mathrm{x}, \mathrm{x}')$ is the Kronecker delta function and τ^2 is the noise variance. $\epsilon(\mathrm{x})$ is assumed to be independent from $\mathrm{Z(x)}$.

Given the noisy observations of $Y(x)$, $\mathrm{Y^n} = [\mathrm{y_1 y_2 y_3 \dots y_n}]^\mathrm{T}$ at x values $\mathrm{X^n} = [\mathrm{x_1 x_2 x_3 \dots x_n}]^\mathrm{T}$, where $\mathrm{y_i} = \mathrm{Y(x_i)}$, $\forall \mathrm{i} = 1, \dots, \mathrm{n}$, Kriging finds the *Best Linear Unbiased Prediction* (BLUP) $\hat{\mathrm{Y}}_\mathrm{n}(\mathrm{x})$ at any unobserved point x, "best" being in terms of conditional quadratic error criterion.

Essentially using n points Kriging gives us closed for expressions for mean $m_{Y_n}(x)$ and variance $s^2_{Y_n}(x)$ at any given point x as [PICH13],

$$m_{Y_n}(x) = \hat{\mu}_n + \phi_n(x)^T \Phi_n^{-1}(Y^n - \hat{\mu} 1_n) \tag{3}$$

$$s^2_{Y_n}(x) = \tau^2 \left[A_n + \frac{A_n^2}{1_n^T \phi_n^{-1} 1_n} \right] \tag{4}$$

where $A_n = 1 - \phi_n(x)^T \Phi_n^{-1} \phi_n(x)$. Furthermore, Φ_n is the $n \times n$ correlation matrix whose $(i,j)^{th}$ entry is equal to

$$\Phi_n(i,j) = \begin{cases} \phi(x_i, x_j) + \tau^2 & i = j \\ \phi(x_i, x_j) & i \neq j \end{cases} \text{ for } i,j = 1 \dots n \quad (5)$$

where $\hat{\mu}_n = \frac{1_n^T \Phi_n^{-1} Y^n}{1_n^T \Phi_n^{-1} 1_n^T}$, is the best linear unbiased estimate of μ, $\Phi_n(x) = [\phi(x, x_1), \phi(x, x_2), \dots, \phi(x, x_n)]^T$ is the correlation of the unobserved point with the existing (observed) points, 1_n is the $n \times 1$ vector of 1's and $\hat{\tau}^2$ is the estimated noise variance. A Gaussian anisotropic covariance kernel is assumed, given by $\phi(x, x') = \sigma^2 \exp[-\theta(x - x')^p], \forall x, x' \in \mathbb{R}$. The estimation of the covariance kernel parameters p, σ^2 and θ are based on the Maximum Likelihood Estimation (MLE) which maximizes the likelihood function of Y^n conditioned on p, σ^2 and θ. This is equivalent to minimizing the negative log-likelihood with respect to p, σ^2 and θ given by [PICH13]

$$\log(\det[\Phi_n]) + (Y^n - \hat{\mu} 1_n)^T \Phi_n^{-1} (Y^n - \hat{\mu} 1_n) \quad (6)$$

Search and poll algorithm

Once we are capable of obtaining the predictions at any given point using the initial models $\hat{F}_n(\mathbf{x})$ for $f(\mathbf{x})$ using the n initial data points, we start the iterative optimization procedure by using *search and poll algorithm*. The search and poll algorithm consists of two steps:

1. An exploratory SEARCH step which, at each iteration: a) finds the global minimum of the current surrogate model $\hat{F}_n(\mathbf{x})$, b) performs a function evaluation at the found minimum, gets the corresponding value of the real function and obtains a new data point, c) includes the new data point in the Kriging data set then updates the surrogate $\hat{F}_n(\mathbf{x})$, and d) looks for the global minimum of the updated surrogate $\hat{F}_n(\mathbf{x})$. The SEARCH continues as long as distinct global optima are found successively and the surrogate function improves (successful SEARCH). When two successive SEARCH steps give the same global optimum $[x]'$ (unsuccessful SEARCH), the POLL step is initiated around $[x]'$, called as the poll point.
2. The POLL step carries out real function evaluations at the immediate neighbourhood points of $[x]'$ in order to find a better point than $[x]'$. If such a point is found (successful POLL), that point is included in the Kriging data set and the SEARCH step is initiated again. Else, if no immediate neighbourhood point of $[x]'$ yields a better function evaluation (unsuccessful POLL), the algorithm stops with $[x]'$ as the output.

2.1.1.3 Scenario

This innovation is related to scenario 2.1.2 defined in [D.2.2]. A daily traffic pattern which has been extracted from an operational network is assumed, see Figure 2. On this figure, the daily traffic variation of the two traffic zones (G_1 and G_2) as well as the two peak-traffic hours occurring during the day (around noon and in the evening) can be observed.

The self-optimization is carried out for the second busiest hour of the day, i.e. optimum total transmit power settings are found for the traffic values of the second busiest hour. The optimized network settings are then applied to all the hours of the day. This is a common exercise in network optimization, traditionally adopted since the early stages of cellular networks, due to its proven robustness.

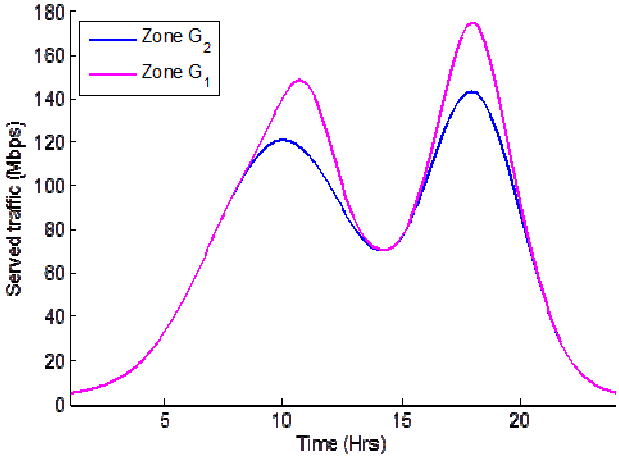


Figure 2. Variation of traffic demand over a 24 hour period.

A regular LTE sub-network consisting of $M = 6$ macro eNBs is considered. Two traffic zones are considered: a central G_1 with a higher traffic (hotspot) and surrounding G_2 with a regular traffic intensity, such that the 5th and 6th eNBs are in G_1 and the other 4 eNBs are in G_2. An analytical simulator which simulates Poisson call arrivals and computes KPIs such as cell loads, Blocked Call Rates (BCR) and File Transfer Times (FTT) based on flow level analytical models is used [COM13]. The simulator computes KPIs based on the traffic demand and the total transmit powers $P_s \forall s \in G_1 \cup G_2$. During the self-optimization process, changes to the transmit powers are allowed with a granularity of $1\,dBm$. The minimum and maximum allowable pilot powers are chosen as $6.2\,dBm$ and $18.2\,dBm$ respectively to ensure full coverage without pilot pollution while the minimum and maximum allowable total transmit power are chosen as $P_{min} = 37\,dBm$ to $P_{min} = 49\,dBm$ respectively to ensure sufficient capacity. Simulations are carried out for the downlink and no mobility is assumed.

Simulation parameters are provided as part of the description of scenario 2.1.2 [D.2.2].

Each simulation run with the injected transmit powers provides the corresponding observations (KPIs), which constitutes a data point to the self-optimization block for model building and optimization. The new transmit power setting calculated by the self-optimization block is fed into the simulator to obtain the next data point in a recursive manner.

2.1.1.4 Future work

An envisaged centralized recursive self-optimization algorithm for a LTE deployment, which uses a surrogate of the network model and a search and poll algorithm for optimization, is expected to improve the QoS by balancing the load between a highly loaded cell and the neighbouring less loaded cells by changing coverage via pilot power optimizations. Furthermore, it will be interesting to see if the optimized parameters for the second busiest hour will equally improve the QoS if applied to all the other hours of the day. The proposed framework requires KPIs averaged over a longer time period (typically an hour to a few hours) and thus are not suitable as a distributed solution. The pilot power based optimization approach does not require any inter-site coordination and information exchange since the algorithm is deployed high up in the OMC where the load KPIs are readily available for optimization. Non-ideal backhaul between macro-sites would not be a problem.

Results using the above described scenario will be provided the future.

2.1.2 Capacity optimization through Active Antenna Systems in LTE macro cell networks

2.1.2.1 Introduction

Traditionally, antenna tilt based optimization has been used as an "offline" process during the network planning/deployment phase, basically for coverage optimization. Planning tools which are based on propagation models have been used to find the optimum antenna tilt settings which are set in the beginning of the operational phase and then modified manually when/if necessary. With the arrival of Active Antenna Systems (AAS), remote modification of antenna tilts during the operational lifecycle of a network has become possible. This possibility, together with the consideration of the significant impact of the antenna tilts on almost every aspect/metric of the radio access network, has recently given rise to self-optimization solutions of Capacity and Coverage Optimization (CCO) based on AAS.

AAS based tilt optimization within a centralized optimization framework at the OMC plane is envisaged to provide improved interference mitigation and thus enhancements in network capacity apart from CAPEX and OPEX gains for the operator.

Antenna tilt is defined by 3rd Generation Partnership Project (3GPP) as an important network optimization parameter that can be used to carry out capacity optimization by reducing inter-cell interference through optimal coverage boundary adjustment of neighbouring cells' AAS for network capacity optimization [RAM01]. AAS involves electronically orienting the antenna beam in elevation and/or azimuth or even varying the shape of the transmission beam. Besides capacity optimization, several studies for antenna tilt based load balancing has shown to provide performance gains in inhomogeneous user traffic distribution [VLA13, GAR04].

2.1.2.2 Solution description

This innovation is based on the use of a similar iterative self-optimization process as described in Section 2.1.1. Here, however, the process aims to optimize the eNodeB antenna tilt settings.

Optimization objective

A typical AAS based optimization scenario includes adjustments to each of the antenna tilts in M. Here the self-optimization algorithm performs optimization by adjusting each of the antenna tilts and proposes an optimum vector of antenna tilt combination.

Let us denote the antenna tilt of cell s as θ_s (°) and the vector of total transmit powers by $\mathrm{x} = \left[\theta_1\ \theta_2\ \theta_3 \ldots\ \theta_{|\mathrm{T}|}\right]^{\mathrm{T}}, \forall \mathrm{s} \in \mathrm{T}$, $|\mathrm{T}|$ being the cardinality of T. The AAS based self-optimization objective function can be given by:

$$\begin{aligned} x^* &= \arg\min_x f(x) \\ s.t.\ \ & c(x) \le Th_c, \theta_{min} \le \theta_s \le \theta_{max}\ \ \forall s \in T \end{aligned} \tag{7}$$

where $f(x)$ and $c(x)$ are the objective function to be optimized and the constraint function, respectively, which are both functions of one or more predetermined KPIs defined over T. $\mathrm{c(x)} = [\mathrm{BCR}_1(\mathrm{x}), \mathrm{BCR}_2(\mathrm{x}), \ldots . \mathrm{BCR}_\mathrm{s}(\mathrm{x})], \forall \mathrm{s} \in \mathrm{T}$, where $\mathrm{BCR}_\mathrm{s}(\mathrm{x})$ is the Block Call Rate (BCR) of cell s $\forall \mathrm{s} \in \mathrm{T}$. Th_c is the constraint threshold on BCR. Finally, θ_{min} and θ_{max} are the minimum and maximum allowable antenna tilt values, respectively.

We can now define several different $\mathrm{f}(\mathbf{x})$ through aggregation of different KPIs, each of which can be used for network capacity optimization as follows:

1. $\mathrm{f(x)} = \mathrm{f}_{5\%}(\mathrm{x}) = \mathrm{R}_{\mathrm{max}5\%} - \sum_{\forall \mathrm{s} \in \mathrm{T}} \rho_\mathrm{s}^{5\%}(\mathrm{x})$, a minimization problem for $\mathrm{f}_{5\%}(\mathrm{x})$ for the maximization of cell edge capacity, where $\rho_\mathrm{s}^{5\%}$ is a 5%-ile throughput of cell s, and $\mathrm{R}_{\mathrm{max}5\%}$ is a predetermined value which is greater than the maximum expected sum of

the 5%-ile throughput over T, i.e. $\mathrm{R}_{\mathrm{max5\%}} > E\left[\max[\sum_{\forall s\in \mathrm{T}} \rho_s^{5\%}(\mathrm{x})]\right]$ where $\mathrm{E}[.]$ denotes expectation,

2. $\mathrm{f(x)} = \mathrm{f}_{50\%}(\mathrm{x}) = \mathrm{R}_{\mathrm{max50\%}} - \sum_{\forall s\in \mathrm{T}} \rho_s^{50\%}(\mathrm{x})$, a minimization problem for $\mathrm{f}_{50\%}(\mathrm{x})$ for the maximization of cell edge capacity, where $\rho_s^{50\%}$ is a 50%-ile throughput of cell s, and $\mathrm{R}_{\mathrm{max50\%}}$ is a predetermined value which is greater than the maximum expected sum of the 50%-ile throughput over T, i.e. $\mathrm{R}_{\mathrm{max50\%}} > E\left[\max[\sum_{\forall s\in \mathrm{T}} \rho_s^{50\%}(\mathrm{x})]\right]$ where $\mathrm{E}[.]$ denotes expectation,
3. $\mathrm{f(x)} = \mathrm{f}_{\mathrm{FTT}}(\mathrm{x}) = \sum_{\forall s\in \mathrm{T}} \mathrm{FTT}_s(\mathrm{x})$, for the minimization of the total download time (for FTP traffic), where FTT_s is the File Transfer Time (FTT) of cell $\mathrm{s}, \forall \mathrm{s} \in \mathrm{T}$.

Search and poll algorithm

Once we are capable of obtaining the predictions at any given point using the initial models $\hat{\mathrm{F}}_\mathrm{n}(\mathbf{x})$ and $\hat{\mathrm{C}}_\mathrm{n}(\mathbf{x})$ for $f(\mathbf{x})$ and $c(\mathbf{x})$ respectively using the n initial data points, we start the iterative optimization procedure by using the search and poll algorithm. The search and poll algorithm consists of two steps:

1. An exploratory SEARCH step which, at each iteration: a) finds the constrained minimum of the surrogate model $\hat{\mathrm{F}}_\mathrm{n}(\mathbf{x})$ such that $\hat{\mathrm{C}}_\mathrm{n}(\mathbf{x}) \leq Th_c$, b) performs a function evaluation at the found constrained minimum, gets the corresponding value of the real function and obtains a new data point, c) includes the new data point in the Kriging data set updates the surrogates $\hat{\mathrm{F}}_\mathrm{n}(\mathbf{x}), \hat{\mathrm{C}}_\mathrm{n}(\mathbf{x})$, and d) looks for the constrained minimum of the updated surrogate $\hat{\mathrm{F}}_\mathrm{n}(\mathbf{x})$ such that $\hat{\mathrm{C}}_\mathrm{n}(\mathbf{x}) \leq Th_c$. The SEARCH continues as long as distinct global optima are found successively and the surrogate function improves (successful SEARCH). When two successive SEARCH steps give the same global optimum $[\mathrm{x}]'$ (unsuccessful SEARCH), the POLL step is initiated around $[\mathrm{x}]'$, called as the poll point.
2. The POLL step carries out real function evaluations at the immediate neighbourhood points of $[\mathrm{x}]'$ in order to find a better point than $[\mathrm{x}]'$. If such a point is found (successful POLL), that point is included in the Kriging data set and the SEARCH step is initiated again. Else, if no immediate neighbourhood point of $[\mathrm{x}]'$ yields a better function evaluation (unsuccessful POLL), the algorithm stops with $[\mathrm{x}]'$ as the output.

2.1.2.3 Scenario

A regular LTE sub-network consisting of $|\mathrm{T}| = 7$ target cells and $|\mathrm{O}| = 12$ observation cells are considered as described in scenario 2.1.3 in [D.2.2]. Furthermore, it is assumed that $\mathrm{T} \equiv \{1,2,3,5,6,9,13\}$ and $O \equiv \{4,7,8,10,11,12,14,15,16,17,18,19\}$. Two traffic zones are considered: G_1 with regular traffic intensity λ_1 and G_2 with higher traffic intensity λ_2, such that the 1st, 5th and 6th target cells are in G_2, the other 4 target cells and all observation cells are in G_1. A flow level simulator which computes cell level KPIs such as loads, BCRs, 5%-ile throughputs, 50%-ile throughputs and FTTs is used [COM13]. The simulator computes KPIs based on the traffic demand and the antenna tilts of the cells in T.

AAS boosts the capacity of eNB site efficiency by using an electronically steering beam which can be oriented in the desired azimuth and elevation thereby improving the gain received at a user location in the network. In the current work, we deal with only changes in antenna tilt θ. The tilt determines the coverage extent of the cell and can be used to reduce intercell interference in LTE thereby increasing capacity of the network. The radiation pattern of an antenna determines the gain at a fixed location (ψ, φ) where ψ is the elevation angle computed using the antenna height and the ground distance from the antenna to that location; φ is the azimuth angle computed using the antenna and the location coordinates.

For a trisectorial site, 3GPP defines azimuth, elevation and the total radiation patterns at location (ψ, φ) given respectively by

$$A_H(\varphi) = -\min\left[12\left(\frac{\varphi}{\varphi_{3dB}}\right)^2, A_m\right] \quad (8)$$

$$A_V(\psi) = -\min\left[12\left(\frac{\psi-\theta}{\psi_{3dB}}\right)^2, SLA_v\right] \quad (9)$$

$$A(\varphi,\psi) = -\min[-[A_H(\varphi) + A_V(\psi)], A_m] \quad (10)$$

where the backward attenuation factors A_m and SLA_v are assumed to be equal to 25 and 20 dB, respectively. Furthermore, the half power beam widths of the azimuth (ψ_{3dB}) and elevation beam patterns (φ_{3dB}) are assumed to be equal to 10 dB and 20 dB, respectively. The total radiation pattern $A(\varphi,\psi)$ is used to compute the pathloss at any given location.

During the self-optimization process, changes to the antenna tilts are allowed with a granularity of 2°. The minimum and maximum allowable antenna tilts are chosen as θ_{min} = 2° and θ_{max} = 12° to ensure full coverage without pilot pollution. Simulations are carried out for the downlink and no mobility is assumed. The other simulation parameters are defined as part of scenario 2.1.3 [D.2.2].

2.1.2.4 Future work and results

A centralized self-optimization framework has been proposed for AAS-based constrained capacity optimization in LTE-A. As shown, statistical learning can be used to model the NP-KPI functional relationships and a pattern search algorithm can be applied in an iterative manner to optimize the network for capacity purpose using different capacity based objective functions. It will be worth comparing if a single objective function optimization is sufficient to achieve the best improvements for all the other objective function values. This is crucial for an operator as it can reduce the optimization problem to a single objective rather than struggling to achieve a balance amongst the competing objectives. Future results will provide a proof-of-concept for the utilization of the novel technique for a wide range of network objective optimizations using AAS tilts based centralized self-optimization framework.

2.1.3 Antenna tilt optimization for interference management in LTE-A heterogeneous network deployments

2.1.3.1 Introduction

With LTE networks reaching the critical Shannon limit, traditional capacity boosters such as macro site densification using more macros and improving the network architecture are seen to be less than permanent solutions. Therefore, network densification with the help of small cells (picos) within the context of Heterogeneous Networks (HetNets) is seen as one of the viable solution for meeting the demands of growing mobile data traffic [RAM11]. HetNets consist of several layers (macro, pico, femto) comprising of base stations operating at fixed transmission power levels. Spectral efficiency can be increased by operating these cell layers on the same carrier frequency, thereby addressing the demands of growing traffic. However, efficient inter-layer and interference management techniques need to be deployed in order to achieve the full spectral potential [RAM11].

Antenna tilt based optimization has been specified in 3GPP as an important network parameter which affects inter-cell interference by adjusting adjacent cell boundaries.

Traditionally, the adjustment of antenna tilts has been carried out by mechanical changes to the antenna assembly involving manual site visits and thus reduced autonomy. Later, Remote Electrical Tilts (RETs) provided the operators with the flexibility of remotely tilting the antenna thereby reducing OPEX. Recently, Active Antenna Systems (AAS) have improved antenna tilt performance by providing the operator with more flexibility to adapt the beam shape, vertical and horizontal orientation at a much faster rate.

Given this flexibility, a question remains to assess the performance of AAS based optimization within the context of centralized self-optimization mechanism.

2.1.3.2 Solution description

This innovation is based on the use of a similar self-optimization process as described in Section 2.1.1. In this section, the process aims to find the optimum antenna tilt settings.

Optimization objective

A typical AAS based optimization scenario includes adjustments to each of the antenna tilts in M. Here the self-optimization algorithm performs optimization by adjusting each of the antenna tilts and proposes an optimum vector of antenna tilt combination.

Let us denote the antenna tilt of cell s as θ_s (°) and the vector of total transmit powers by $\mathrm{x} = \left[\theta_1\ \theta_2\ \theta_3 \ldots\ \theta_{|\mathrm{M}|}\right]^{\mathrm{T}}, \forall \mathrm{s} \in \mathrm{M}$, $|\mathrm{M}|$ being the cardinality of M. The AAS based self-optimization objective function can be given by:

$$\begin{gathered} x^* = \arg\min_x f(x) \\ s.t.\ \ c(x) \le Th_c, \theta_{min} \le \theta_s \le \theta_{max}\ \ \forall s \in M \end{gathered} \tag{11}$$

where $f(x)$ and $c(x)$ are the objective function to be optimized and the constraint function, respectively, which are both functions of one or more predetermined KPIs defined over M. $\mathrm{c(x)} = [\mathrm{BCR}_1(\mathrm{x}), \mathrm{BCR}_2(\mathrm{x}), \ldots . \mathrm{BCR}_\mathrm{s}(\mathrm{x})], \forall \mathrm{s} \in \mathrm{M}$, where $\mathrm{BCR}_\mathrm{s}(\mathrm{x})$ is the Block Call Rate (BCR) of cell s $\forall \mathrm{s} \in \mathrm{M}$. Th_c is the constraint threshold on BCR. Finally, θ_{min} and θ_{max} are the minimum and maximum allowable antenna tilt values.

We can now define several different $\mathrm{f}(\mathbf{x})$ through aggregation of different KPIs, each of which can be used for network capacity optimization as follows:

1. $\mathrm{f(x)} = \mathrm{f}_{5\%}(\mathrm{x}) = \mathrm{R}_{\mathrm{max}5\%} - \sum_{\forall \mathrm{s} \in \mathrm{M}} \rho_\mathrm{s}^{5\%}(\mathrm{x})$, a minimization problem for $\mathrm{f}_{5\%}(\mathrm{x})$ for the maximization of cell edge capacity, where $\rho_\mathrm{s}^{5\%}$ is a 5%-ile SINR of cell s, and $\mathrm{R}_{\mathrm{max}5\%}$ is a predetermined value which is greater than the maximum expected sum of the 5%-ile SINR over M, i.e. $\mathrm{R}_{\mathrm{max}5\%} > E\left[\max[\sum_{\forall \mathrm{s} \in \mathrm{M}} \rho_\mathrm{s}^{5\%}(\mathrm{x})]\right]$ where $\mathrm{E}[.]$ denotes expectation,
2. $\mathrm{f(x)} = \mathrm{f}_{50\%}(\mathrm{x}) = \mathrm{R}_{\mathrm{max}50\%} - \sum_{\forall \mathrm{s} \in \mathrm{M}} \rho_\mathrm{s}^{50\%}(\mathrm{x})$, a minimization problem for $\mathrm{f}_{50\%}(\mathrm{x})$ for the maximization of cell edge capacity, where $\rho_\mathrm{s}^{50\%}$ is a 50%-ile SINR of cell s, and $\mathrm{R}_{\mathrm{max}50\%}$ is a predetermined value which is greater than the maximum expected sum of the 50%-ile SINR over M, i.e. $\mathrm{R}_{\mathrm{max}50\%} > E\left[\max[\sum_{\forall \mathrm{s} \in \mathrm{M}} \rho_\mathrm{s}^{50\%}(\mathrm{x})]\right]$ where $\mathrm{E}[.]$ denotes expectation.

Search and poll algorithm

Once we are capable of obtaining the predictions at any given point using the initial models $\hat{\mathrm{F}}_\mathrm{n}(\mathbf{x})$ and $\hat{\mathrm{C}}_\mathrm{n}(\mathbf{x})$ for $f(\mathbf{x})$ and $c(\mathbf{x})$ respectively using the n initial data points, we start the iterative optimization procedure by using search and poll algorithm as follows. The search and poll algorithm consists of two steps:

1. An exploratory SEARCH step which, at each iteration: a) finds the constrained minimum of the surrogate model $\hat{\mathrm{F}}_{\mathrm{n}}(\mathbf{x})$ such that $\hat{\mathrm{C}}_{\mathrm{n}}(\mathbf{x}) \leq Th_c$, b) performs a function evaluation at the found constrained minimum, gets the corresponding value of the real function and obtains a new data point, c) includes the new data point in the Kriging data set updates the surrogates $\hat{\mathrm{F}}_{\mathrm{n}}(\mathbf{x}), \hat{\mathrm{C}}_{\mathrm{n}}(\mathbf{x})$, and d) looks for the constrained minimum of the updated surrogate $\hat{\mathrm{F}}_{\mathrm{n}}(\mathbf{x})$ such that $\hat{\mathrm{C}}_{\mathrm{n}}(\mathbf{x}) \leq Th_c$. The SEARCH continues as long as distinct global optima are found successively and the surrogate function improves (successful SEARCH). When two successive SEARCH steps give the same global optimum $[\mathrm{x}]'$ (unsuccessful SEARCH), the POLL step is initiated around $[\mathrm{x}]'$, called as the poll point.
2. The POLL step carries out real function evaluations at the immediate neighbourhood points of $[\mathrm{x}]'$ in order to find a better point than $[\mathrm{x}]'$. If such a point is found (successful POLL), that point is included in the Kriging data set and the SEARCH step is initiated again. Else, if no immediate neighbourhood point of $[\mathrm{x}]'$ yields a better function evaluation (unsuccessful POLL), the algorithm stops with $[\mathrm{x}]'$ as the output.

2.1.3.3 Scenario

An LTE-A heterogeneous network deployment consisting of 27 macro cells, where $\mathrm{M} = [\mathrm{eNB}_1, \mathrm{eNB}_2, \mathrm{eNB}_3, \mathrm{eNB}_{11}, \mathrm{eNB}_{17}, \mathrm{eNB}_{18}, \mathrm{eNB}_{24}]$ and $\mathrm{P} = [\mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3]$ pico cells are considered as shown in Figure 3. Traffic arrives with a regular intensity in all of the 27 macro cells, however KPIs are collected from M and P only. Simulations are carried out for the downlink only. A semi-dynamic LTE-A system simulator which performs correlated Monte Carlo snapshots with a time step of $t = 1$ sec is used.

For a trisectorial site, 3GPP defines azimuth, elevation and the total radiation patterns at location (ψ, φ) given respectively by equations (8), (9) and (10). In the equations, the backward attenuation factors A_m and SLA_v are assumed to be equal to 20 dB and 25 dB respectively. Parameters ψ_{3dB} and φ_{3dB} are the half power beamwidths of the azimuth and elevation beam patterns, and they are assumed to be equal to 10 dB and 20 dB, respectively. The total radiation pattern $A(\varphi, \psi)$ is used to compute the pathloss at any given location.

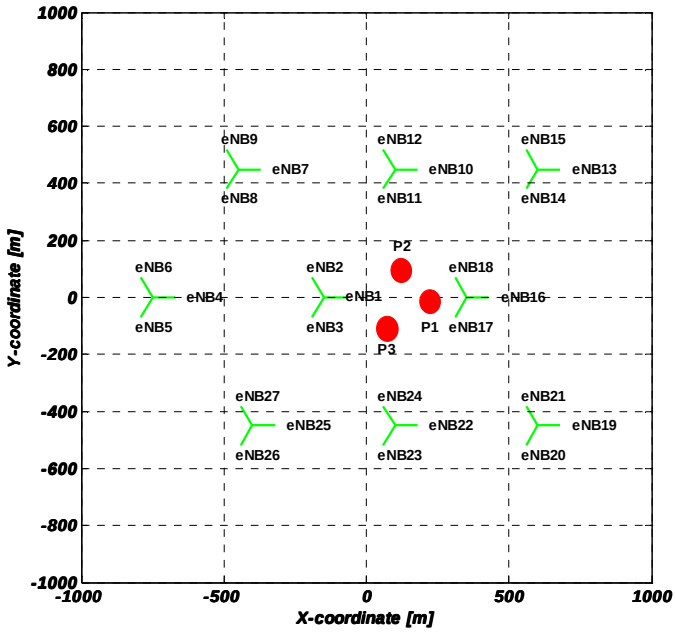


Figure 3. LTE-A network layout with 27 macro and 3 pico eNBs.

At each time step, new UEs arrive, are checked for call admission, and are either admitted or blocked based on capacity or coverage criteria. Connected UEs are allowed to download a file of fixed size. The UEs that finish communication leave the system, while the UEs that continue their communication are moved according to a pre-determined mobility model, after which the mobile-base station attachments are re-calculated. Each simulation lasts for 3000 seconds and at the end the performance metrics (KPIs) are collected. The parameters used in the system simulation are as listed as part of scenario 2.3.2 defined in [D.2.2].

The following KPIs are observed for evaluating the performance of the centralized self-optimization algorithm:

1. $\mathrm{SINR}_{5\%}$ and $\mathrm{SINR}_{50\%}$ for the 5%-ile and 50%-ile SINR of the entire network
2. $\mathrm{Th}_{\mathrm{P5\%}}$ and $\mathrm{Th}_{\mathrm{P50\%}}$ for the 5%-ile and 50%-ile Throughput for the pico cells in P
3. $\mathrm{Th}_{\mathrm{M5\%}}$ and $\mathrm{Th}_{\mathrm{M50\%}}$ for the 5%-ile and 50%-ile Throughput for the macro cells in M
4. $\mathrm{BCR}_{\mathrm{M}}$ and $\mathrm{BCR}_{\mathrm{P}}$ for BCRs in M and P respectively

2.1.3.4 Future work and results

Future results will provide an insight into the performance of the centralized recursive self-optimization algorithm and challenges for a LTE-A HetNet deployment with AAS based antenna tilt optimization. One of the critical results will be to observe the impact on the pico cell KPIs and how they are influenced by a macro-only optimization scenario. For the considered dense urban environment, it will be seen if either or both of the cell-center capacity optimization and cell-edge capacity optimization are needed to obtain the best improvement of all other macro, pico and overall network KPIs. The proposed approach does not require any inter-site coordination and information exchange as it is employed as a slow closed-loop control system at the OMC plane where the KPIs are readily available.

2.1.4 Enhanced Inter-Cell Interference Coordination (eICIC) for interference management in LTE-A networks

2.1.4.1 Introduction

In a co-channel HetNet deployment of macro and pico cells, the operator is always faced with a critical challenge of reducing the inter-cell interference of pico cells. While pico densification increases the capacity of a network, the high macro cell transmission power increases the interference towards the pico cells thereby degrading the QoS of the pico cell-edge users. Downlink time domain (TDM) enhanced Inter-Cell Interference Coordination (eICIC) defined for HetNet SON by 3GPP Release 10 [RAM11] helps to manage interference by providing synchronized interference coordination in the time domain in which the interfering eNB partially mutes a pre-defined percentage (called as the muting ratio) of the sub-frames in order to allow the interfered eNB to serve its users with a higher data rate. This is achieved through X2 signalling between the interfering and the interfered eNBs. Furthermore, at the instant of connection initiation, due to the difference in Received Signal Reference Power (RSRP), a user closer to the pico cell edge may be biased towards the macro while sufficient capacity/resources are still available with the pico. In such cases, a pico cell can operate with an increased cell range, termed "Cell Range Extension" (CRE) [RAM11] through a bias term added to the RSRP from the pico cell in the cell selection process.

A combination of CRE and TDM eICIC has been shown to improve network performance for the macro-pico scenario through simulations with different bias and muting ratio values [PED12,WAN12].

The envisaged difficulty is for the operator to choose the right muting ratio and CRE combination for optimum network performance. Thus, the proposed solution involves incorporation of eICIC within a centralized self-optimization framework which makes use of Kriging regression to build a functional relationship between the muting ratio, CRE pair and the network KPIs.

TDM eICIC

TDM eICIC provides the necessary interference mitigation against the macro interference suffered by the pico users in the CRE scheme. Co-channel interference is minimized by restricting macro layer transmissions in a fraction of the sub-frames. During the sub-frames in which macro layer does not transmit useful data to the UE, only the signals that are mandatory are transmitted and thus these sub-frames are also called "Almost Blank Sub-frames (ABS)". During these sub-frames when the macro layer uses ABS, there is less interference towards the pico users at the cell-edge. It is of course of prime importance that perfect sub-frame synchronization in time and phase exists between the macro and the pico layer and that the pico layer is aware of the ABS pattern thanks to X2 interface signalling, so as to serve the cell-edge users with priority during ABS.

Pico cell range extension

In a co-channel HetNet deployment of macro and pico cells, the serving cell selection is performed through comparison of the UE RSRP measurements from the nearby macro and pico cells. Due to the difference between macro and pico transmit powers, a simple comparison would yield an unbalanced load partition between macro and pico cells, resulting in smaller capacity gains than is typically expected from HetNet deployments. Therefore, to increase the pico coverage area, an approach called as Cell Range Extension (CRE) is adopted. In case of CRE, a cell-specific positive bias is applied to the RSRP measurements performed on pico cells. As a result, the selection of the pico cells will be favoured [RAM11]:

$$ID_{serving} = \arg\max_{i}\{RSRP_i + RE_i\} \tag{12}$$

where $\mathrm{RE_i} = 0\mathrm{dB}$ for macro cells and $\mathrm{RE_i} \geq 0$ for pico cells. CRE increases the pico cell coverage and thus enhances the offloading effect from the macro towards the pico.

2.1.4.2 Solution description

This innovation is based on the use of an iterative self-optimization process as described in Section 2.1.1. Here, the process is utilized to find the optimum values for both the cell range extension (cell selection offset, RE) and the eICIC muting ratio (mR).

Optimization objective

The aim within the eICIC based self-optimization context is to find the optimum $[\mathrm{mR\ RE}]^{\mathrm{T}}$ value by optimizing a pre-determined objective function subject to certain constraints. For the considered TDM eICIC + CRE problem, we define a Joint Performance Metric (JPM) as the weighted sum of the 5%-ile and the 50%-ile of UE SINR:

$$JPM(mR, RE) = (1-\lambda)SINR_{50\%}(mR, RE) + \lambda SINR_{5\%}(mR, RE) \tag{13}$$

where λ $(0 \leq \lambda \leq 1)$ is the fairness parameter. $\lambda = 0.5$ can be assumed to have a compromise between cell center and cell edge users. The constrained optimization problem is to maximize the above JPM with an upper limit on network BCR. For this, we use the following minimization formulation:

$$\min_{RE,mR}[f(mR,RE)] \tag{14}$$

where $\mathrm{f(mR,RE)} = (1-\mathrm{JPM_{norm}}) + \mathrm{g(mR,RE)}$, $\mathrm{JPM_{norm}}$ is the normalized JPM:

$$JPM_{norm} = \frac{JPM - \min\limits_{RE,mR}(JPM)}{\max\limits_{RE,mR}(JPM) - \min\limits_{RE,mR}(JPM)} \tag{15}$$

and $\mathrm{g = g(mR,RE) = max_{RE,mR}(0, BCR - BCR_{limit})}$ is the constraint limit on network BCR which is added as a penalty function ($\mathrm{BCR_{limit}}$).

The surrogate models for $\mathrm{SINR_{50\%}(mR,RE)}$, $\mathrm{SINR_{5\%}(mR,RE)}$ and $\mathrm{BCR(mR,RE)}$ are constructed separately by Kriging.

Search and poll algorithm

Once we are capable of obtaining the predictions at any given point using the initial models $\widehat{\mathrm{JPM}}_{\mathrm{n}}(\mathrm{mR,RE})$ and $\widehat{\mathrm{G}}_{\mathrm{n}}(\mathrm{mR,RE})$ for $\mathrm{JPM_{norm}}$ and $\mathrm{g(mR,RE)}$ respectively using the n initial data points, we start the iterative optimization procedure by using search and poll algorithm as follows. The search and poll algorithm consists of two steps:

1. An exploratory SEARCH step which, at each iteration: a) finds the global minimum of the function $\mathrm{f(mR,RE)}$ which is constructed using the updated surrogate models $\widehat{\mathrm{JPM}}_{\mathrm{n}}(\mathrm{mR,RE})$ and $\widehat{\mathrm{G}}_{\mathrm{n}}(\mathrm{mR,RE})$, b) performs a function evaluation at the found minimum, gets the corresponding value of the real function and obtains a new data point, c) includes the new data point in the Kriging data set updates the surrogates $\widehat{\mathrm{JPM}}_{\mathrm{n}}(\mathrm{mR,RE})$ and $\widehat{\mathrm{G}}_{\mathrm{n}}(\mathrm{mR,RE})$, and d) looks for the global minimum of the updated function $\mathrm{f(mR,RE)}$. The SEARCH continues as long as distinct global optima are found successively and the surrogate function improves (successful SEARCH). When two successive SEARCH steps give the same global optimum $[\mathrm{mR,RE}]'$ (unsuccessful SEARCH), the POLL step is initiated around $[\mathrm{mR,RE}]'$, called as the poll point.
2. The POLL step carries out real function evaluations at the immediate neighbourhood points of $[\mathrm{mR,RE}]'$ in order to find a better point than $[\mathrm{mR,RE}]'$. If such a point is found (successful POLL), that point is included in the Kriging data set and the SEARCH step is initiated again. Else, if no immediate neighbourhood point of $[\mathrm{mR,RE}]'$ yields a better function evaluation (unsuccessful POLL), the algorithm stops with $[\mathrm{mR,RE}]'$ as the output.

2.1.4.3 Scenario

A heterogeneous network consisting of 21 macro eNBs (7 sites with 3 cells per site) and 1 pico per macro in the central site is considered as shown in scenario 2.3.1 [D.2.2]. The traffic is distributed around the whole simulation area of 21 macros, but KPIs are collected only for the central 3 macros and picos. Simulations are carried out for downlink. The applied simulation parameters are listed in scenario 2.3.1 [D.2.2]. A semi-dynamic LTE-A system simulator which performs correlated Monte Carlo snapshots with a time step of t seconds is used ($t = 1$ sec). At each time step, new users are admitted or blocked based on channel conditions and resource availability, user buffers are updated, users that have finished their communication leave the system, users that continue their communication are moved according to a pre-determined mobility, mobile-base station attachments are re-calculated and performance metrics are collected. Each simulation lasts for 3000 seconds and the resulting KPIs provide a data point to the self-optimization block for model-building and optimization. The new network parameter setting calculated by the self-optimization block is fed into the simulator for the next 3000 seconds of simulation.

We consider $\mathrm{eNB_M}$ $(\mathrm{M} \in \mathrm{G1} = \{1,2,3\})$ as the central macro group, and $\mathrm{eNB_P}$ $(\mathrm{P} \in \mathrm{G2} = \{4,5,6\})$ as the pico group. Each element of $\mathrm{G1}$ with the respective one of $\mathrm{G2}$ forms a macro-pico pair of one sector. Let mR denote the macro eNB muting ratio and RE the cell selection offset for the pico eNB cell range extension. Since the UE traffic distribution is spatially stationary for all the macro-pico pairs (see UE placement in the scenario 2.3.1 [D.2.2] which is in accordance with 3GPP recommendations [3GPP10a]) we consider that mR and RE values are the same $\forall \mathrm{M} \in \mathrm{G1}$ and $\forall \mathrm{P} \in \mathrm{G2}$, and modified simultaneously by the centralized self-optimization.

2.1.4.4 Future results

A centralized recursive self-optimization algorithm for a LTE-A HetNet deployment with RE + TDM eICIC, which uses a surrogate of the network model and a search and poll algorithm for optimization to improve the cell-edge SINR and throughput will be tested using a system-level simulator. The search and poll algorithm is a 'brute force' method and while it has no restrictions, Kriging imposes a complexity issue when building surrogate models (on a desktop computer) with very high dimensions (greater than 45 base stations) and very fine mesh (number of possible values on each dimension). Although scalability can be an issue, it can be solved by using a meta-surrogate approach where an initial surrogate can be used to build a rough estimation followed by a secondary surrogate build by refining either the number of dimensions or the mesh size. Note that this approach requires an inter-site coordination and information exchange over the X2 interface for setting the ABS patterns. A future work could be to investigate the performance of the proposed optimization framework for non-ideal backhaul between macro-sites.

2.1.5 Mobility Load Balancing in LTE macro cell networks

2.1.5.1 Introduction

Mobility Load Balancing is one of the solutions to balance load which involves changes to mobility parameters in order to postpone or prepone the time instant of handovers between neighbouring cells. These are useful to increase capacity in dense traffic areas consisting of a loaded cell surrounded by less loaded neighbours. MLB implies a deep insight into the traffic conditions which are very complex and costly, in particular when involving manual changing of parameters [RAM11].

An automated MLB within the framework of SON provides a closed loop optimization for the operator thereby reducing OPEX in the short term and CAPEX in the long term. The use of MLB within the context of SON can also improve user QoS.

Intra-LTE MLB consists of tuning the Handover Margin (HM) parameter of the handover procedure defined by the A3 event for LTE [RAM11]. The aim of MLB is to balance the loads between neighbouring eNBs by offloading loaded eNBs towards less loaded eNBs. The HM from cell *s* to cell *t* is denoted by $HM_{s,t}$ in dB where $s \in G_1, t \in G_2 \cap \aleph_s$ and $\aleph_s$ is the neighbor set index of cell *s*.

The decision to move a user equipment (UE) from serving cell *s* to target cell *t* is taken if the following conditions are satisfied [RAM11]:

1. $RSRP_t \geq RSRP_s + HM_{s,t} + Hyst$
2. $RSRP_t \geq RXmin_t$
3. There exist sufficient resources in the target cell.

where both the $RSRP$ and the $HM_{s,t}$ are expressed in dBm. Furthermore, $RXmin$ is the minimum received signal power level for establishing a successful communication in dBm. Finally, $Hyst$ is a constant independent of the eNBs and of the mobile station and is set here to 0.

If condition 1 is satisfied, conditions 2 and 3 are checked. In case of non-validity of conditions 2 or 3, the UE is rejected by the target cell, the handover procedure is terminated and the UE is bounced back to the serving cell.

For better sensitivity of the eNB loads to HM changes, we assume a bi-directional change of HM, i.e. $HM_{t,s} = 10\ dB + HM_{s,t}$.

2.1.5.2 Solution description

This innovation is based on the use of a similar iterative self-optimization process as described in Section 2.1.1. Here, the aim of the process is to find an optimum value for the handover margin.

Optimization objective

For the scenario 2.1.1 considered in [D.2.2], let us assume a uniform setting of the HM parameter, i.e. only one $HM_{s,t}$ value, $\forall s \in G_1, t \in G_2 \cap \aleph_s$ (shown by red arrows in the figure). Let us denote this value of $HM_{s,t}$ by x. Then, the objective function of the MLB can be written as:

$$\begin{gathered} x^* = \arg\min_x f(x) \\ s.t.\ \ BCR^t(x) \le Th_{BCR},\ \forall t \in G_2 \cap \aleph_s \\ DCR^t(x) \le Th_{DCR},\ \forall t \in G_2 \cap \aleph_s \end{gathered} \tag{16}$$

where $f(x) = \sum_{\forall s \in G_1, t \in G_2 \cap \aleph_s} [load_s(x) - load_t(x)]^2$, $load_s(x)$ and $load_t(x)$ are the cell loads, $BCR^t(x)$ and $DCR^t(x)$ are the Block Call rate (BCR) and Drop Call Rate (DCR) of cell *t* respectively. Th_{BCR} and Th_{DCR} are the upper limit thresholds on the BCR and DCR respectively $\forall s \in G_1, t \in G_2 \cap \aleph_s$.

Constrained expectation of improvement algorithm

In order to choose the next infill point within the surrogate based optimization framework discussed above and for locating the optimum of the surrogate approximated so far by the Kriging model, we use the maximization of expectation of improvement criterion that balances surrogate exploration and exploitation as described in [FOR01]. Suppose that we have built the Kriging models for $f(x)$, $BCR^t(x)$, $DCR^t(x)$, $\forall t \in G_2 \cap \aleph_s$ based on the n observation data points $(\mathbf{X}^n, \mathbf{Y}^n)$ as:

1. $\hat{F}_n(x)$ is the Kriging model for $f(x)$ which is Gaussian with conditional mean m_{F_n} and variance $s^2_{F_n}(x)$,
2. $\widehat{BCR}^t_n(x)$ is the Kriging model for $BCR^t(x)$ which is Gaussian with conditional mean $m_{BCR^t_n}(x)$ and variance $s^2_{BCR^t_n}(x)$,
3. $\widehat{DCR}^t_n(x)$ is the Kriging model for $DCR^t(x)$ which is Gaussian with conditional mean $m_{DCR^t_n}(x)$ and variance $s^2_{DCR^t_n}(x)$,

To minimize function $f(x)$ let us define an "improvement" achieved by Kriging prediction $\hat{F}_n(x)$ over the n observed values as $I_n(x) = f_n^{min} - \hat{F}_n(x)$ where $f_n^{min} = min_{x \in \{x_1, x_2, \ldots, x_n\}} F(x)$ [PICH13, FOR01]. This implies that $I_n(x)$ is also Gaussian with mean $f_n^{min} - m_{F_n}$ and variance $s^2_{F_n}(x)$. The optimum x value, x^*, which minimizes $\hat{F}_n(x)$ is that value which maximizes our expectation of having an improvement over all the n observed values. The solution can be given by [FOR01]

$$x^* = \arg\max_x E[I_n(x)] \tag{17}$$

where $\mathrm{E}[I_n(x)] = s_{F_n}(x)\left\{\frac{u}{2}\left[1 + erfc\left(\frac{u}{2}\right) + \frac{1}{\sqrt{2}}e^{-\frac{u^2}{2}}\right]\right\}$, $u = \frac{f_n^{min} - m_{F_n}(x)}{s_{F_n}(x)}$ and $erfc(.)$ is the complementary error function.

The above equation represents the minimization of the objective function $f(x)$. To include constraints in the optimization process we need to restrict the solution space to those values of x which satisfy the constraints on BCR and DCR. We do this by first defining the measure of "feasibility":

$$F_n(x) = \prod_{t \in \mathrm{G}_2 \cap \aleph_s} \Pr\left[\widehat{BCR}_n^t \leq Th_{BCR}\right] \Pr\left[\widehat{DCR}_n^t \leq Th_{DCR}\right] \tag{18}$$

It can be shown that this equation can be expressed in a closed form solution as [FOR10]

$$F_n(x) = \prod_{t \in \mathrm{G}_2 \cap \aleph_s} \left\{1 - \frac{1}{2}\left[\frac{Th_{BCR} - m_{BCR_n^t}(x)}{\sqrt{2s^2_{BCR_n^t}(x)}}\right]\right\}\left\{1 - \frac{1}{2}\left[\frac{Th_{DCR} - m_{DCR_n^t}(x)}{\sqrt{2s^2_{DCR_n^t}(x)}}\right]\right\} \tag{19}$$

This measure of feasibility can be incorporated as a constrained optimization problem by defining a joint optimization function as a multiplication of expectation of objective improvement and feasibility of constraint as

$$x^* = \arg\max_x \mathrm{E}[I_n(x)]\, F_n(x) \tag{20}$$

2.1.5.3 Scenario and results

The performance evaluation presented in this section is based on scenario 2.1.1 described in [D.2.2]. A typical regular network layout comprising 15 sites with 3 sectors per site is used. The environment considered is dense urban. The assumption is that eNB_1 is surrounded by its first tier geographical neighbours eNB_{11}, eNB_{14}, eNB_{15} and eNB_{18}.

An LTE simulator developed in MATLAB has been used for simulations. The simulator performs correlated snapshots, and at each end of time step of *t* seconds (s), new mobiles enter the system, those that end their communications leave and positions of the active mobiles are updated. Variable bit rate (VBR) traffic is considered. A non-uniform traffic distribution with hotspots at cell edges is considered. All mobiles are moving with a specific speed in a random direction.

Each simulation is run for a time period of 3000 seconds. The first 500 seconds is the stabilization time and the observation time is from 501 s to 3000 s during which the KPIs are window-averaged with a period of 2500 seconds. The averaged KPIs are used by the optimization algorithm to determine the new parameter settings which are then fed to the simulator for the next 3000 seconds of simulation. A default value of $HM = 6$ dB for all eNBs is used in the network. The algorithm is free to specify any value of HM as input within the available range of 0 to 12 dB.

Figure 4 shows the optimization objective, the expectation of improvement and the constrained expectation of improvement curves for the optimization steps. At $k = 1;$ which corresponds to the first update, expectation of improvement is high in the region closer to the minimum of $f(HM)$, but the constrained expectation of improvement forces the algorithm to move in the direction of constraint feasibility towards a solution at $HM_{s,t} = 8.242\, dB$. At this iteration, the constraint is satisfied, however there is still scope for further improvement in the objective minimization as indicated by the constrained expectation of improvement curves proposing a global optimized setting at $HM_{s,t} = 7.5\, dB$. Convergence of the algorithm is indicated by a very low $\mathrm{E}[I(HM)]F(HM)$ value of 10^{-134}.

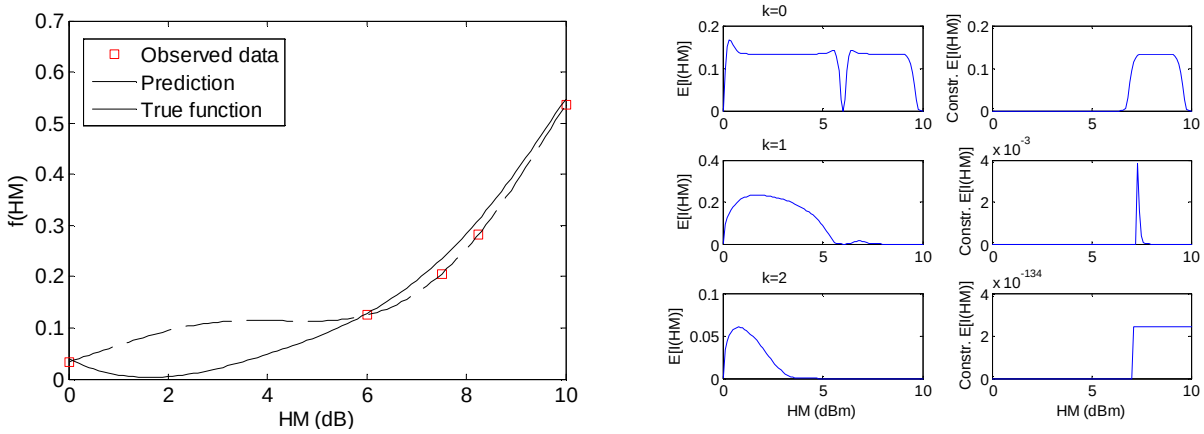


Figure 4. Objective function $f(HM)$ data points and Kriging regression and updates using maximum $\mathrm{E}[I(HM)]\,F(HM)$ for $Th_{BCR} = 0.05$.

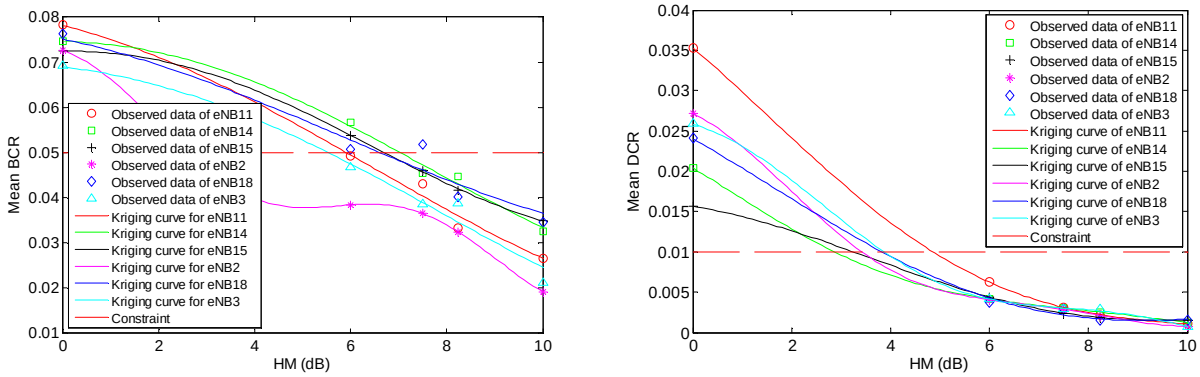


Figure 5. Mean BCR and DCR data points and Kriging regression using 3 initial designs (0 dB, 6dB and 10 dB) and the two optimization updates (7.5dB and 8.1dB).

Figure 5 indicates the regression plots for the initial designs and optimization updates. As can be observed, the update stops at or below the respective BCR and DCR constraint values set by the operator.

Looking at the bar plots in Figure 6, it is clear that optimization unloads the users from the most loaded eNB (which here is eNB1) and increases the load of the neighbouring cell. While doing so, it inadvertently impacts the BCR and DCR of these cells, by reducing BCR, DCR (QoS) in eNB while degrading them in the neighbours within their respective constraint limits. As the BCR constraint is loosened further from 5% to 6.5%, more traffic is offloaded to neighbouring eNBs improving the QoS of eNB further.

The optimization results in a 55.83% and 16% reduction in DCR and BCR respectively for eNB1 and at $Th_{BCR} = 0.05$.

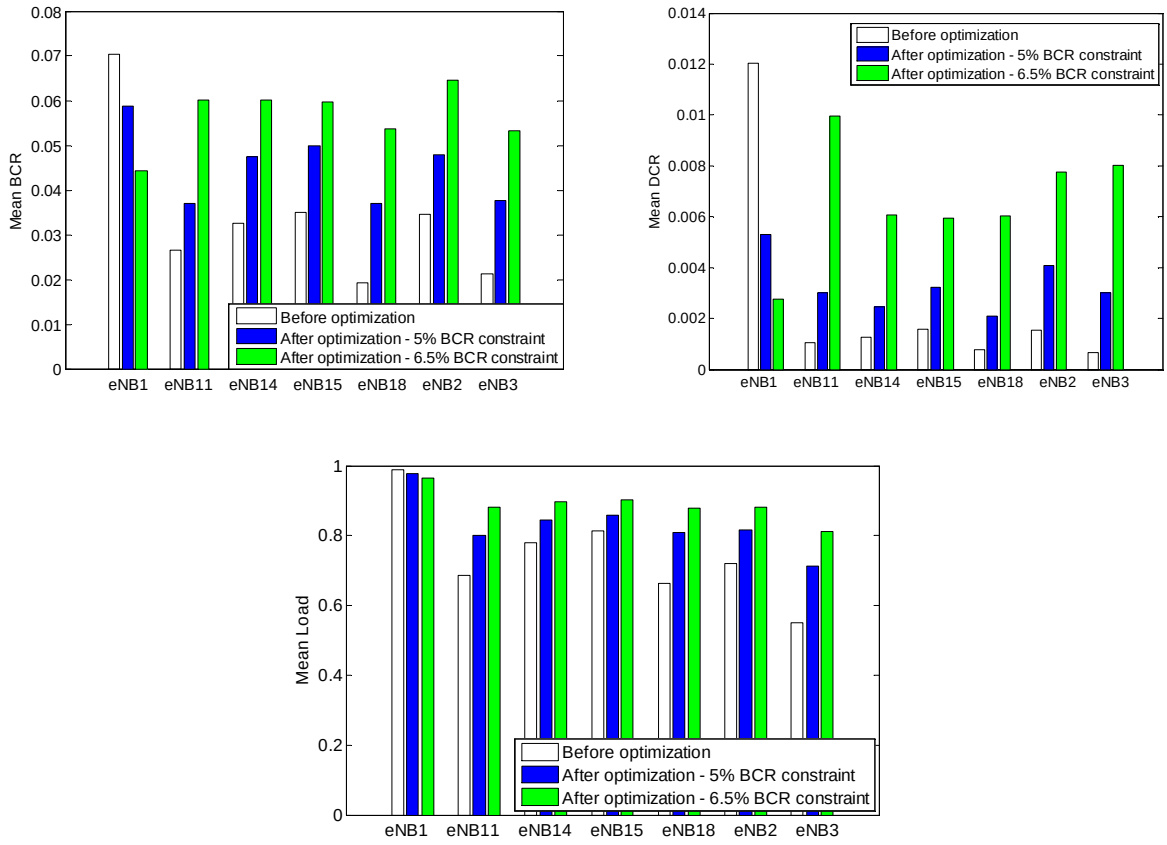


Figure 6. Mean BCR, DCR and Load bar plots for the optimized design for $Th_{BCR} = 0.05$ and $Th_{BCR} = 0.06$.

2.1.5.4 Conclusions

As presented in the results, the use of surrogate based centralized optimization can significantly improve the user QoS when compared with non-optimized network parameter setting. However, when comparing the optimized setting, centralized MLB based optimization performance improves QoS in the most loaded cell, but at an expense of degrading QoS in the neighbouring cells. This tradeoff in QoS degradation and amelioration between the eNBs is expected as the users which were earlier served by the congested cell are now being served by its neighbours while the call arrival rate remains constant. Nevertheless, this solution brings significant gains for the operators in terms of CAPEX and OPEX and thus is an important tool for use in centralized and long term optimization objectives. Finally, it should be noted that the proposed surrogate based centralized SON is not subject to any architecture constraints since the KPIs are readily available at the Operations and Maintenance Center (OMC) and the centralized SON solution can be deployed at the OMC plane. Thus, no coordination is required and the solution is not subject to any timing constraints as they typically operate over a step size of an hour.

2.2 Load Balancing

2.2.1 Cell virtualization based on Large Scale Antenna System

2.2.1.1 Introduction

Small cells can be one of the solutions to the requirement of high spectral efficiency of the explosively growing mobile traffic. Also, they are useful to handle high-dense traffic areas. Typically, the deployment of small cells implies a non-negligible cost for operators in terms of new equipment deployment: location of new sites, leasing, maintenance and increment of the global energy consumption of the network [MAR10].

Large Scale Antenna System (LSAS) at eNB could open the door to more efficient strategies in network deployment than those including small cells. The use of antenna systems having a two-dimensional array structure provides control over both the elevation and the azimuth dimensions [SPEN04][YOO06]. This would allow accurately matching the transmitted energy in the downlink to the user distribution in the coverage area. For example, different beams could point to different hotspots (high density of UEs), thereby increasing the SINR in this area. Those beams could be considered as new cells and replace the typical HetNet deployment (Figure 7), therefore reducing the cost for operators. In this section, those high gain beams are called as Virtual Small Cells (VSCs).

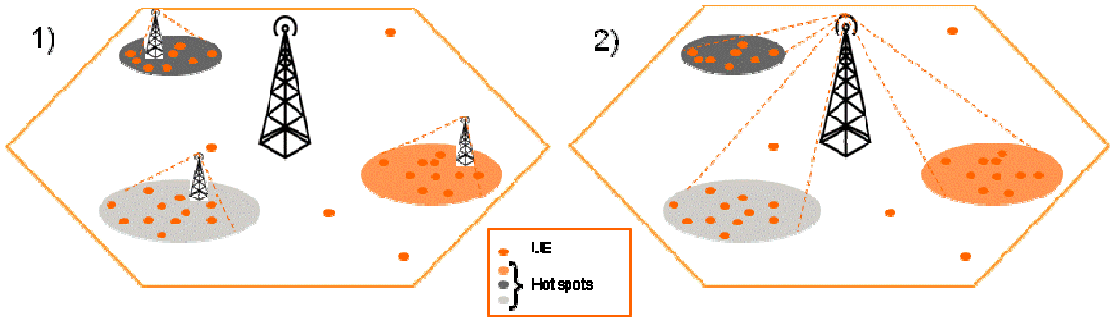


Figure 7. Typical HetNet deployment (1) versus (2).

In addition, VSCs could be reconfigured in time, in order to be adaptable to the changing traffic conditions. This flexibility is an important advantage compared to the typical small cell deployment since it will give the network the capability of being intelligently responsive in terms of energy focalization following the network traffic variations.

The aim of this work is to evaluate the gains in terms of throughput achieved with such VSCs compared to the macro-only network and with a typical HetNet deployment, where small nodes (pico cells) are installed in the vicinity of the high-dense traffic area.

2.2.1.2 Solution description

VSCs are defined by a highly directional beam. This beam is formed in such a way that it will cover a given area, i.e., a hotspot. The required beam to cover a hotspot is defined by its direction in the horizontal and vertical planes, defined as θ_{tilt} and φ_{tilt}, respectively, and its beam width in both planes given by θ_{3dB} and φ_{3dB}, respectively, as shown in Figure 8 and defined here below:

$$\theta_{tilt} = -\tan^{-1}\left(\frac{h}{dist}\right)(rad) \quad (21)$$

$$\varphi_{tilt} = \tan^{-1}\left(\frac{y-y'}{x-x'}\right)(rad) \tag{22}$$

$$\theta^{\circ}_{3dB} = 2\left[\tan^{-1}\left(\frac{h}{dist-r}\right) - \tan^{-1}\left(\frac{h}{dist+r}\right)\right]\cdot\frac{180}{\pi} \tag{23}$$

$$\varphi^{\circ}_{3dB} = 2\left[\tan^{-1}\left(\frac{r}{dist}\right)\right]\cdot\frac{180}{\pi} \tag{24}$$

where h represents the macrocell antenna height, $dist$ is the distance between the macrocell base station and the center of the hot spot, which coordinates are represented by *y'*, *x'*, r stands for the radius of the hot spot and *y*, *x* are the macrocell BS coordinates.

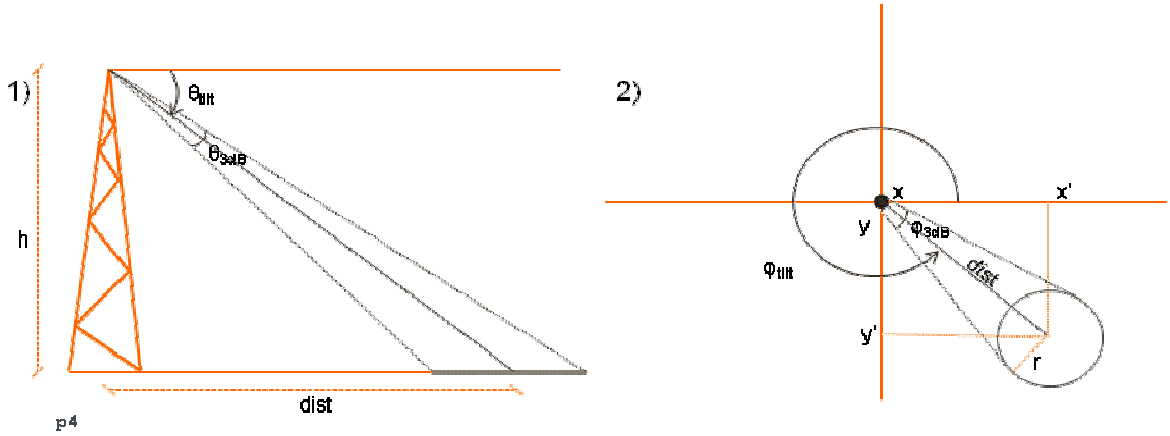


Figure 8. Beam representations in the vertical (1) and horizontal planes (2).

Then, having the beam-width we can easily determine the maximum gain of the group of antennas used to form the beam. The maximum gain, G_{max}, can be computed as the maximum directivity, D_{max}, multiplied by the antenna efficiency e [ORF08][BAL05]

$$G_{max} = eD_{max} \tag{25}$$

where D_{max} is given by solid angle of a sphere in steradians (sr) and the beam solid angle or beam area Ω_A

$$D_{max} = \frac{4\pi}{\Omega_A} \tag{26}$$

Ω_A can be approximated in terms of the half-power angles of the main lobe in vertical and horizontal planes.

$$\Omega_A \cong \theta^{\circ}_{3dB}\cdot\varphi^{\circ}_{3dB} \tag{27}$$

Following Ballanis reference [BAL05] we arrive to following approximation for planar arrays

$$D_{max} = \frac{32400}{\theta^{\circ}_{3dB}\cdot\varphi^{\circ}_{3dB}} \tag{28}$$

Also, with the beam-width we can estimate the antenna length in the vertical and horizontal dimensions as follows:

$$L_v^{\circ} = \frac{50}{\theta_{3dB}^{\lambda}} \tag{29}$$

$$L_h^{\circ} = \frac{50}{\varphi_{3dB}^{\lambda}} \tag{30}$$

where the angles are expressed in terms of λ, the array inter-element distance. Finally, the number of elements required to form the needed beam depend on the direction of the beam and henceforth they are estimated as follows:

$$N_{elem}^{v} = \frac{L_v^{\circ}}{\lambda \cdot \cos(\theta_{tilt})} \tag{31}$$

$$N_{elem}^{h} = \frac{L_h^{\circ}}{\lambda \cdot \cos(\varphi_{tilt})} \tag{32}$$

Methodology

Figure 9 summarizes the methodology for the VSCs implementation at the site level. Based on the cell traffic analysis, the first step is to identify if there are hot spots in the coverage area and where they are. To do so, Radio Environment Maps (REM) could be applied using the received Radio Signal Strenght Indicator (RSSI). The maps should be constructed periodically in order to follow the appearance and disappearance of the hot spots. Virtual small cells are therefore created or eliminated as a function of the network traffic distribution.

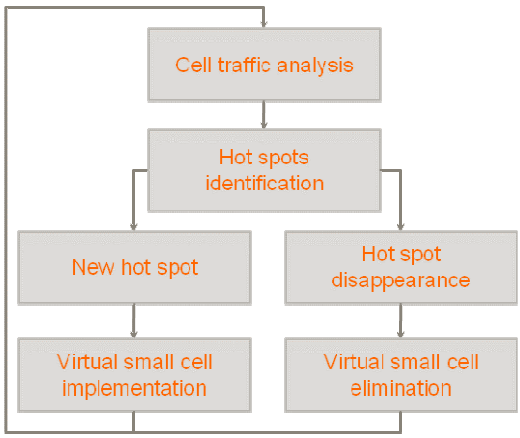


Figure 9. Methodology for the virtual small cell implementation.

2.2.1.3 Scenario and results

In this section we present some results obtained in a system-level 3GPP LTE (Release 8) quasi-compliant simulator. First we present the scenarios considered and then we show the obtained results in terms of UE attachment to the existing serving cells and some plots comparing the SINR and throughput obtained in the network. The presented performance evaluation is based on scenario 2.3.13 described in [D.2.2].

A dense urban hexagonal macrocell layout with Inter Site Distance (ISD) equal to 500 m is considered. Furthermore, a deployment with 19 three-sector sites is assumed, resulting in a total of 57 macro cells with wrap around. The macro cell antenna height is assumed to be equal to 32 m and the maximum antenna gain is assumed to be equal to 14 dBi. The antenna pattern is based on the 3D model given by 3GPP [3GPP10a]

$$A_V(\theta) = -\min\left[12\left(\frac{\theta-\theta_{tilt}}{\theta_{3dB}}\right), SLA_v\right] \quad (33)$$

$$A_H(\varphi) = -\min\left[12\left(\frac{\varphi}{\varphi_{3dB}}\right)^2, A_m\right] \quad (34)$$

$$A(\varphi,\theta) = -\min\{-A_H(\varphi) + A_V(\theta), A_m\} \quad (35)$$

where it is assumed that: half-power beam width in the vertical plane $\theta_{3dB} = 10°$; backward attenuation $SLA_v = 20$ dB, half power beam width in the horizontal plane $\varphi_{3dB} = 70°$ and the backward attenuation $A_m = 25$ dB.

Each macro sector is assumed to contain one hotspot with a radius of 40 m. Two options are considered to cover the hotspots: 1) deployment of small nodes, i.e. one picocell installed in the center of the hotspot; 2) to cover each hotspot with a VSC formed at the macrocell base station using LSAS. The UEs are randomly deployed with a probability of 66% of being in a hotspot, which makes that 2/3 of the traffic will be handled by the small cells, either by the VSCs or by the picocells. UEs are modeled with omnidirectional antenna with gain equal to 0 dBi and an antenna height of 1.5 m.

A full buffer traffic model is assumed during the simulations; the number of users in the cell is constant and the buffers of the users' data flows always have unlimited amount of data to transmit.

The carrier frequency is assumed to be on the 2 GHz band and the bandwidth is assumed to be equal to 10 MHz. The used Path Loss (PL) models for macrocell to UE for Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS), respectively [3GPP10a].

$$PL = 103.4 + 24.2 \cdot \log_{10}(R) \quad (36)$$

$$PL = 131.1 + 42.8 \cdot \log_{10}(R) \quad (37)$$

And for the picocell to the UE

$$PL = 103.8 + 20.9 \cdot \log_{10}(R) \quad (38)$$

$$PL = 145.4 + 37.5 \cdot \log_{10}(R) \quad (39)$$

where R (km) is the 2D distance between the position of the UE and of the transmitting antenna, whether macrocell or picocell (i.e. R is not considering the antenna height).

Finally, it is assumed that when VSCs are implemented in the macrocell, the total transmission power at a macrocell has to be kept below the $P_{max} = 46$ dBm, i.e. the total transmitted power by the macrocell and its hosted VSCs is below the P_{max}. We assume a simplified solution where P_{max} is proportionally divided between the macro and its hosted VSCs.

Figure 10 shows how the UEs are attached to the existing serving cells. Macrocells and small cells are represented by triangles, orange and grey, respectively. UEs are represented by dots, red dots show the UEs attached to the macrocells and the blue ones are the UEs attached to the small cells. The left-hand side of the figure shows one shot of the macrocell-VSC scenario and the one on right-hand side of the figure presents the macrocell-picocell scenario. Table 1 summarizes the UE attachment percentages obtained for both scenarios. In the case of the macro-VSC scenario, the amount of UEs attached to the VSCs is 8% higher than the distribution probability that has been set. This is given by 1) the high beam gain of the VSC and 2) the equal power distribution between the macrocell and the VSC.

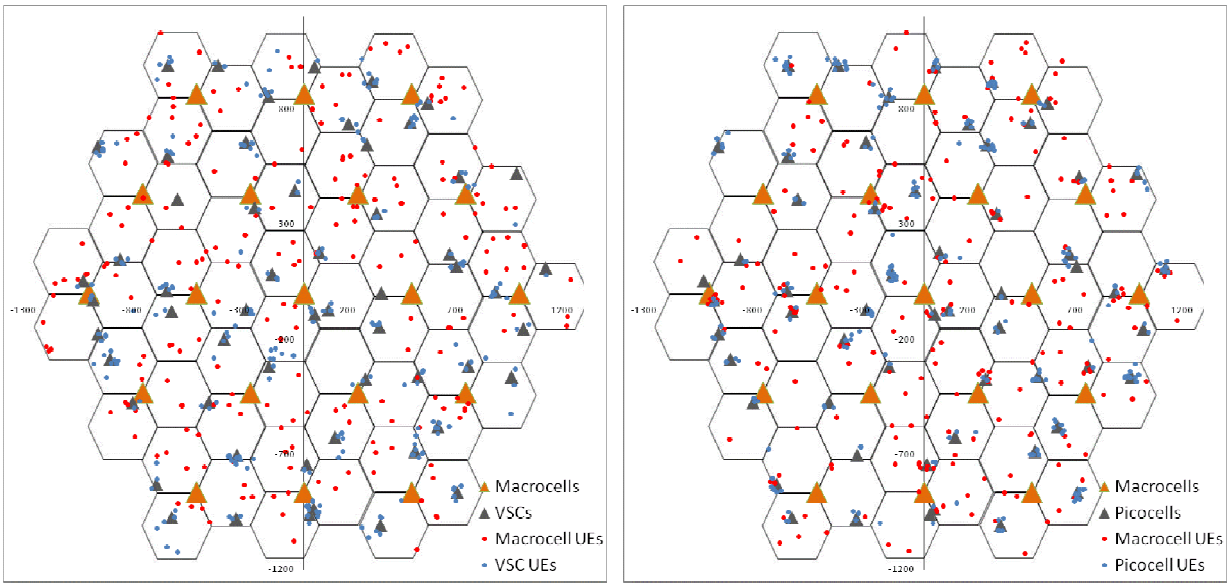


Figure 10. One realization of UE attachment for Macro-VSC and Macro-Picocell scenarios.

Table 1. UE attachment percentage for Macro-VSC scenario and Macro-Picocell scenarios.

Macrocell-VSC scenario		**Macrocell-Picocell scenario**	
Macrocell	VSC	Macrocell	Picocell
26%	74%	47%	53%

Figure 11 presents the Cumulative Distribution Function (CDF) of the average received Signal to Interference Ratio (SIR) for the evaluated scenarios. Onaverage, the SIR of the macrocell-VSC case is 0.5 dB higher than the SIR for the macrocell only case due to the interference generated by the VSCs at the macrocell. As expected, the macrocell-picocell case provides much better results than the macrocell-only scenario; onaverage the SIR is 3.3 dB higher. When the VSCs are introduced, the cell-edge SIR experiences a degradation of 1.4 dB. Table 2 summarizes the results for the average received SIR percentile.

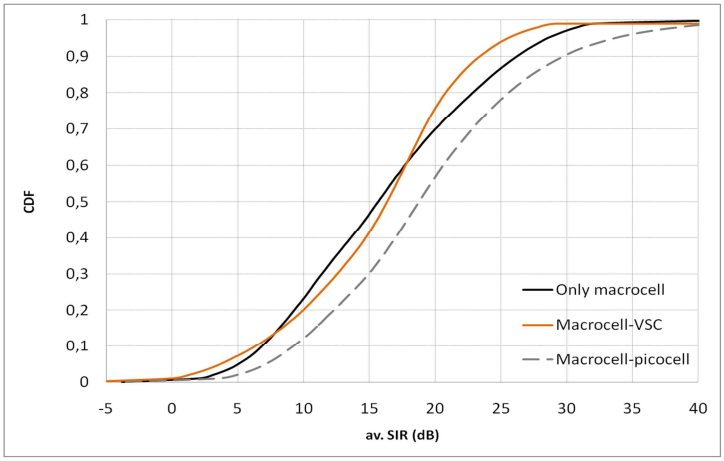


Figure 11. CDF of the average received SIR

Table 2. Summary of the average received SIR.

Metric	**Only macro**	**Macro-VSC**	**Macro-Picocell**
Avg	16	15.5	19.3
5%	5.1	3.7	7.1
50%	15.7	16.3	18.8
95%	28.5	25.5	33.6

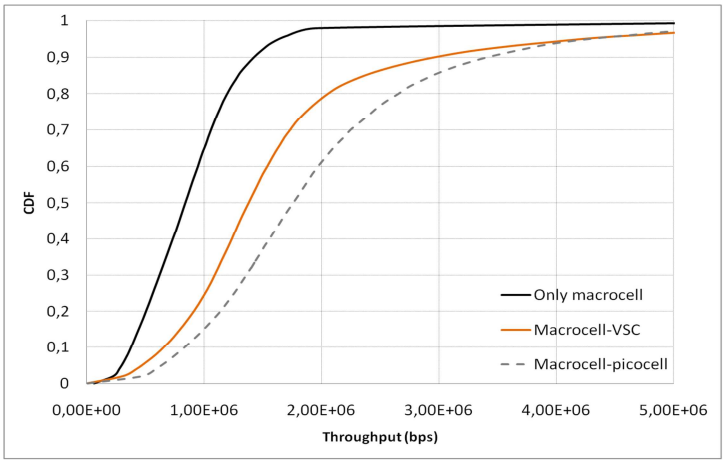


Figure 12. CDF of the user throughput.

Figure 12 shows the CDF of the user throughputs for the evaluated scenarios. Onaverage, the introduction of the VSCs in the network increases the throughput by 0.8 Mbps compared to the only-macrocell case, but reduces the throughput by 0.3 Mbps compared to the typical HetNet deployment (macrocell-picocell case). Comparing the user throughput at the cell-edge we can see that the introduction of the VSCs improves the throughput by 0.17 Mbps when compared with the only-macrocell case. Nevertheless, the macrocell-VSC scenario provides 0.17 Mbps lower user throughput in the cell-edge than when picocells are introduced to cover the hotspot. Table 3 summarizes the results for the user throughputs. It should be noted that

when introducing the VSC, in the one hand the average SIR is not enhanced but on the other hand the user throughput is considerably improved. This is due to the fact that VSC reuses the radio resources used by the macro cell, thereby doubling the total radio resources used by the network.

Table 3. Summary of the user throughput results.

Metric	**Only macro**	**Macro-VSC**	**Macro-Picocell**
Avg	0.88	1.7	2
5%	0.3	0.47	0.64
50%	0.84	1.4	1.8
95%	1.6	4.2	4.3

2.2.1.4 Conclusions

As demonstrated by the results, the introduction of VSCs can improve the user throughputs when compared with a situation where only macrocells are deployed. Nevertheless, when comparing this solution with the classical HetNet implementation, it can be seen that the network performance can be jeopardized due to increased distance between the VSCs and the users resulting in higher pathlosses. Further studies are required to investigate if the gains in throughput introduced by the VSCs are high enough for the growing traffic demand. In any case, this solution brings significant savings for the operators in terms of CAPEX and OPEX and these benefits could eventually overcome the suboptimal performance when compared to traditional HetNets. Finally, it should be noted that VSCs are not subject to any architectural constraints since the macro cell and the VSCs are hosted in the same site and hence coordination is not required. Therefore, no inter-site synchronization is required a priori. Non-ideal backhaul between the macro-sites would not be a problem.

2.2.2 Intra-LTE offloading by middleware deployment

Moving users need to satisfy and guarantee their needs in terms of coverage and quality of service (QoS). Therefore, the network must be able to manage the changes in the serving cells without any prejudice to the customer. That means intelligently managing handover procedures so as to guarantee that each user is served in the most appropriate cell given their characteristics and network requirements. The envisioned solution consists of creating and deploying a middleware capable of helping in these processes, which will perform handovers as high level operations, forcing disconnections and reconnections but not going into the control plane details.

The proposed innovation corresponds to scenario 2.9.1 in [D.2.2]. The scenario being faced in this case is therefore similar to the one shown in Figure 13, where an offloading process is carried out between the LTE cells where the diverse users are moving from and to while carrying out active sessions.

Thus, the proposed innovation implies the creation of the aforementioned middleware, which will be deployed in the base stations and will work on an IP level, capable of managing the cell occupation, given the impossibility of obtaining valid direct information from the users on certain levels.

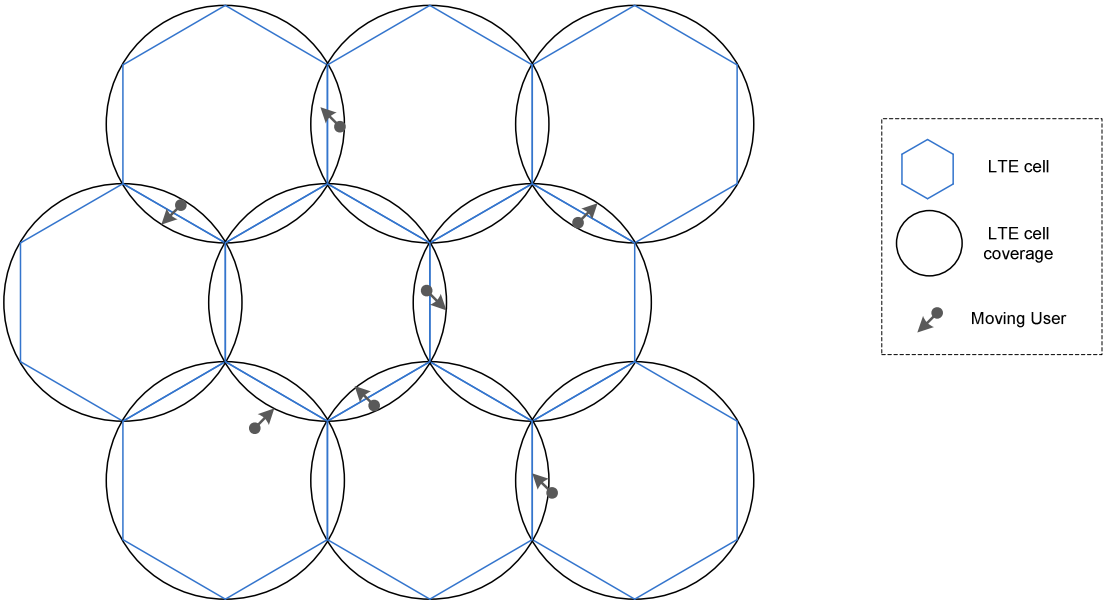


Figure 13. Intra-LTE offloading scenario

The scenario presents an added difficulty, marked by the users' mobility, which will derive in the need of maintaining an exchange of information between the lower and upper layers both at local and remote level (offering defined Service Access Points, SAP) and ona global network scale as well. Figure 14 depicts roughly the interactions taking place between the lower entities (for example the case represented by transceivers, base stations or users) and the high level entities managed at core network level.

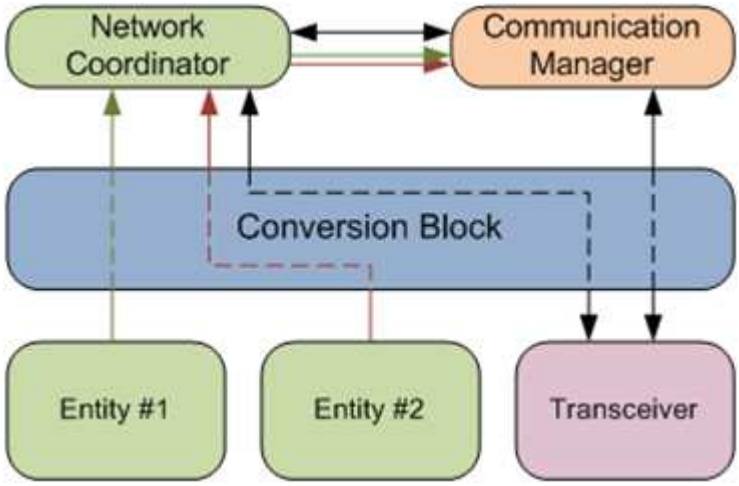


Figure 14. Basic block diagram of the envisaged middleware

The envisioned middleware will be a functional entity that will receive and manage information from different layers dispatching primitives to their final destination, trying to avoid duplicity of paths. It will also support event-subscription functionalities for higher layers and functional blocks at local and remote level.

In addition, this middleware, which will receive the name of *Conversion Block*, will enable technology-agnostic information management when necessary, presenting capacity to analyse and adapt measurements and parameters depending on the RATs below and capability to accommodate the received data to a given Common Data Model (i.e. battery level alarms, overload levels in APs or BSs, SNR border values, etc.).

All in all, the Conversion Block will imply a novel and an efficient method to perform intra-LTE duties, as required in the mobility conditions so typical nowadays. Its internal structure will be similar to the one shown in Figure 15, distinguishing two main parts: the ACCESS section, directly connected to the desired entity, and the CORE section, where a dedicated database is deployed and the information processed and conveniently dispatched.

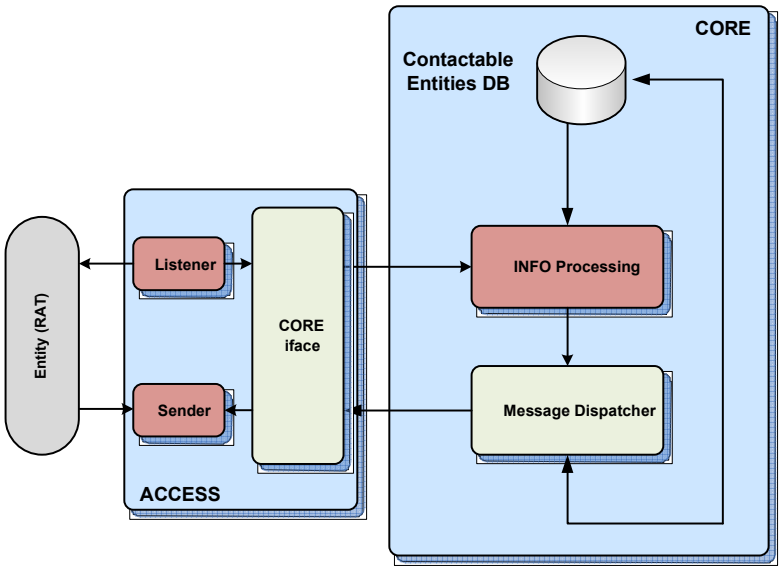


Figure 15. Internal structure of the middleware

The interaction among the diverse entities and this middleware can be roughly depicted in the Message Sequence Chart (MSC) presented in Figure 16, where a SHARING entity (namely, a user equipment) initiates a registration process, in order to make itself known within the system. The moment the middleware acknowledges its presence, proceeds to include it in its own database and to confirm the entity this event. From this moment on, the middleware is capable of sending information to the entity when necessary or interesting.

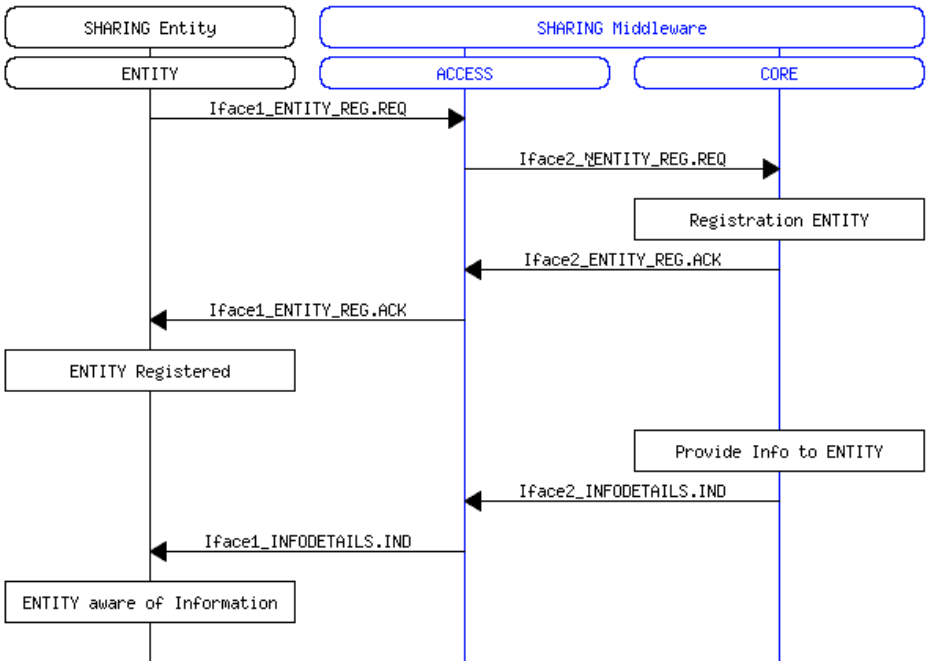


Figure 16. Interactions between Entities and Middleware MSC

2.3 Mobility Management

2.3.1 Combined cell performance within HSPA heterogeneous network deployment

The deployments of heterogeneous network with small cells in HSPA allow increasing the overall network capacity, coverage and performance. In co-channel deployment, the Low-Power Nodes (LPN) are deployed within the macro cell coverage region, where the transmission/reception points created by the low-power nodes have different cell IDs (different primary scrambling codes) as the macro cell. As shown in Figure 17, cells A, B and C have different primary scrambling codes, hence the same legacy procedure of cell selection applies for each cell and is controlled by the RNC.

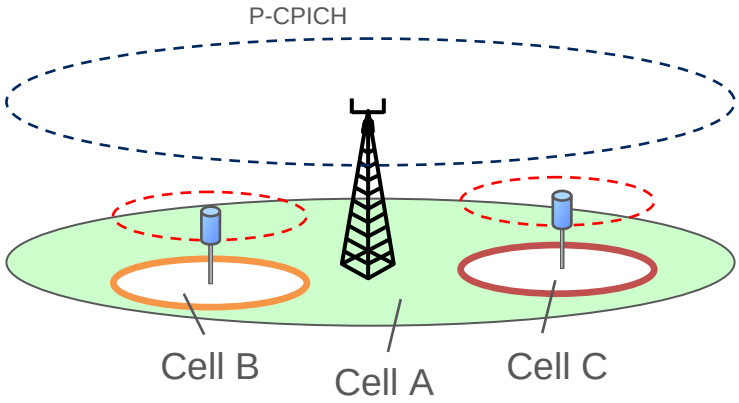


Figure 17. Low-power nodes have different cell IDs as that of the macro node in a co-channel deployment.

In a combined cell deployment, the low-power nodes are deployed within the macro cell coverage area, where the transmission/reception points created by the low-power nodes have the same cell IDs (same primary scrambling codes) as compared to the macro cell. This deployment is also referred to as soft or shared cell. As shown in Figure 18, each node belongs to the same cell and these nodes assist the macro node.

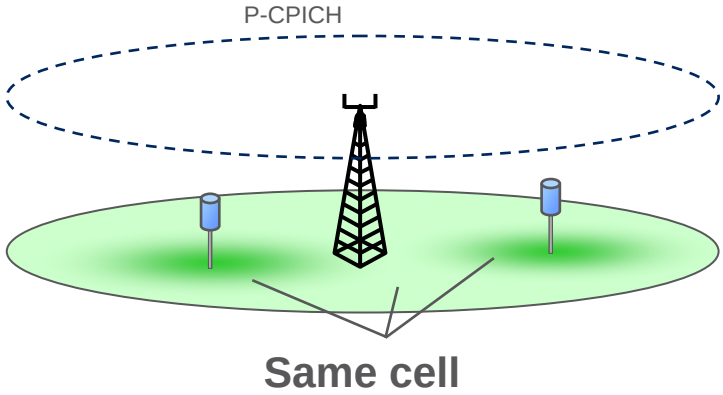


Figure 18. A combined cell deployment, where the low-power nodes are part of the macro cell, i.e. macro and low-power nodes have the same cell ID.

The co-channel deployments of small cells introduce some additional challenges that the network has to confront:

1. **Frequent Handovers and Impact on End-User Performance**: Since there are more cells in a macro node coverage area with the introduction of LPNs, the frequency

of handovers is increased. This results in more frequent RRC signalling which might impact the end-user performance negatively. For example, more dropped calls due to RRC signalling delay or failure.

2. **Neighbour Cell List Size**: In a co-channel deployment, the neighbor cell list becomes too large to cover all radio positions. The required cell planning and the system capability to identify and keep the cell update subset is complex.
3. **Intercell Interference and Pilot Pollution**: With the introduction of low power nodes, the interference structure becomes more complex than in a homogenous network. Since all the LPNs have to transmit the pilot signals continuously, irrespective of data transmission, the pilot pollution is more severe.
4. **Downlink/Uplink Imbalance**: The well-known problem of downlink/uplink imbalance where the UE is served by strong macro downlink and has a stronger uplink to the LPN. This might cause problems, both for uplink and downlink control channels.
5. **Energy Consumption**: Since the pilots and certain control channels are always transmitted in a co-channel deployment (even though no UE is served by these LPNs), the energy associated with these channels is wasted.

The main principle of combined cell is that the UE can move seamlessly within the cell coverage area without any RNC interaction. Hence in combined cell, it can avoid active set update, serving cell change and cell selection/reselection procedures. Figure 19 shows the system architecture for the combined cell deployment; where all the nodes within a combined cell are tightly coupled by high speed data link to a central unit in the combined cell. For example this central unit can be a macro scheduling unit similar to current main unit in main/remote base station implementations. Coupling between various nodes is not a requirement in combined cell deployment. Note that in combined cell deployment RNC connects to the central unit and is not aware of these different nodes. For example these nodes can be remote radio units (RRU). In co-channel deployment scheduling is done per each cell while in combined cell scheduling is performed per combined cell. Hence, the scheduler decides which nodes should transmit to the UE. In all, the operations performed in RNC in co-channel deployment will be performed by the central scheduler in combined cell deployment, i.e. the central scheduler tracks the UE between multiple nodes. The combined cell deployment avoids the loading of RNC, while at the same time the decisions and execution can be performed very fast (TTI level), improving both the overall network performance as well as the UE performance.

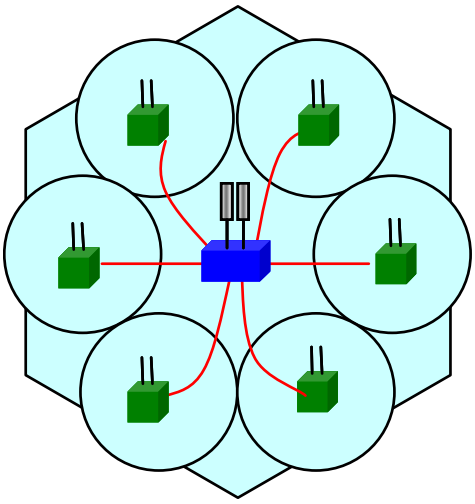


Figure 19. System architecture of combined cell deployment, where all the nodes are tightly coupled and connected to the central scheduler.

Based on the data transmission from LPN and macro nodes, we can divide the downlink transmission modes into three types:

- **Single Frequency Network (SFN) or Multicasting**. In this mode, multiple nodes (e.g. macro and LPNs) transmit the same data to a specific UE. Hence, the signal to noise ratio of the UE can be improved. The main idea of this mode is to combine signals over the air from all nodes by means of transmitting exactly the same pilot, control channels and data channel in downlink using the same carrier frequency and spreading and scrambling codes.
- **Node Selection with Spatial Reuse**. In the SFN mode, all nodes are transmitting the same downlink signal. Hence it may not give capacity gains when the traffic load is high as SFN mode is used for coverage improvement. Since many nodes do not contribute to the performance improvement, the resources from the nodes are not used efficiently. The interference pattern in combined cell deployment is similar to that of co-channel deployment; the resources from these nodes can be utilized to schedule different UEs. In this mode, the same P-CPICH signal is transmitted from all the nodes. The downlink control channel and the data traffic are scheduled to different UEs from different nodes, based on their position inside the network and the available resources (HS codes and power). Since the scheduling is done per combined cell, the central scheduler decides which nodes should transmit to the various UEs.
- **MIMO mode with spatially separated nodes**. In this mode, some of the low-power nodes act like distributed MIMO, i.e. MIMO transmission with spatially separated antennas. In this mode, MIMO gains (both diversity and multiplexing gains) can be achieved.

Since in combined cell, all the nodes are connected to the central scheduler, we envision a significant gain in uplink performance. This is due to macro diversity combining. The downlink performance might vary based on the transmission modes, but a positive impact of the reduction on mobility procedures should also be observed in the final results in realistic scenarios.

There is no impact on the legacy architecture with relation to signalling and standardized procedures. The model builds on the close network relationship between the transmitters within a cell that may be located in the same physical node or not. And for this work, the assumption is that proprietary (not standardized) signalling is used in case of any coordination between nodes is required.

Impact of combined cell deployments on mobility

The mobility performance of combined cell deployments has been proved to increase the robustness of the handovers [MIN14]. This is achieved because the deployment enables a reduction on the number of handovers triggered on a coverage area. Furthermore, the handover signalling failure rate is reduced compared with deployments of separated small cells, especially for challenging mobility scenarios. A challenging mobility scenario can be defined as one with high system load, several spots with low geometry and user's mobility that can reach high-speed. Some example of this scenario could be a train station or train tracks, a stadium hosting an event or a traffic light corner where low-power nodes have been deployed.

A macro-only scenario compared with a low-power node scenario would produce similar results in terms of number of handovers, but the deployment would suffer from their own capacity and coverage limitations. This means, that a combined cell deployment is expected to address also the capacity and end-user performance limitation in addition to the mobility.

Initial performance results for Single Frequency Network mode

The Single Frequency Network mode of combined cell is 100% legacy compatible and does not require any changes in the standard.

Initial simulation results are available for the scenarios 2.8.1 (macro only) and 2.8.2 (HetNet) in [D.2.2] for FTP traffic (2 MB object) and 4 LPNs (distributed RRU's/antennas connected to macro in addition to the main antennas) per macro sector. For FTP upload, the results confirm the initial assumption that even a SFN deployment is able to increase the uplink throughput due to the macro-diversity gain of having distributed receiving antennas. As can be seen in Figure 20, by increasing the load per square kilometre, the mean user throughput is impacted in a macro-only deployment (labelled MACRO-500m-ISD). Meanwhile, the combined cell SFN deployment shows a very low degradation (labelled CC-SFN-4LPN). Additionally, the 5-percentille user's throughput for combined cell remains virtually unchanged with the load.

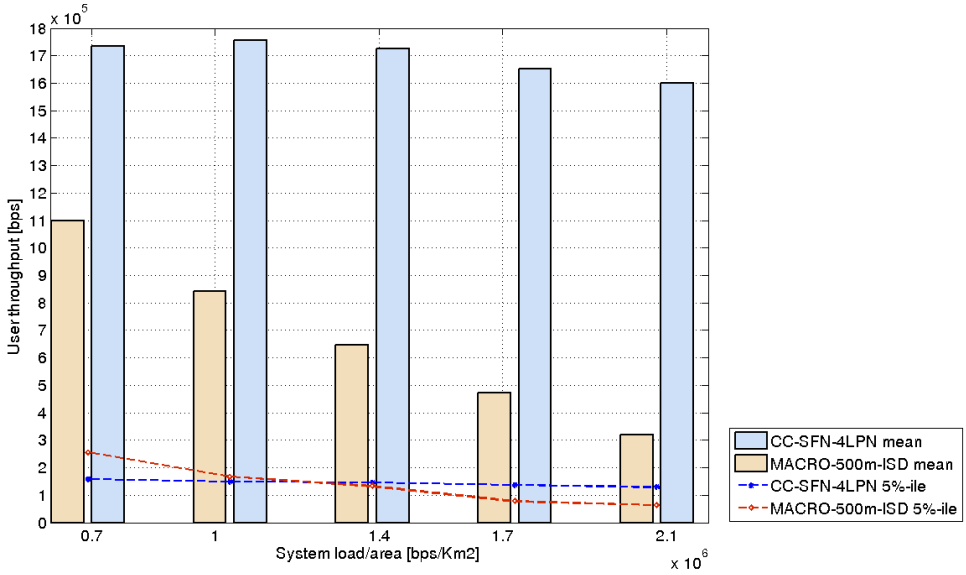


Figure 20. Comparison of UL performance of Macro-500 ms ISD deployment versus a Combined Cell SFN deployment with 4 additional distributed RRUs.

The results indicate that the combined cell increases the area capacity. This can be visualized by looking at the system throughput efficiency, defined as the ratio between the offered load and the system throughput. If the network starts to delay packets the output rate is reduced with respect to the input rate meaning that the queues in the network increases and the services start to suffer. According to the results in Figure 21, the throughput efficency of the combined cell deployment for uplink is higher than the macro-only deployment for higher loads in the same area. This is an indication that more users can be served by a combined cell deployment as expected.

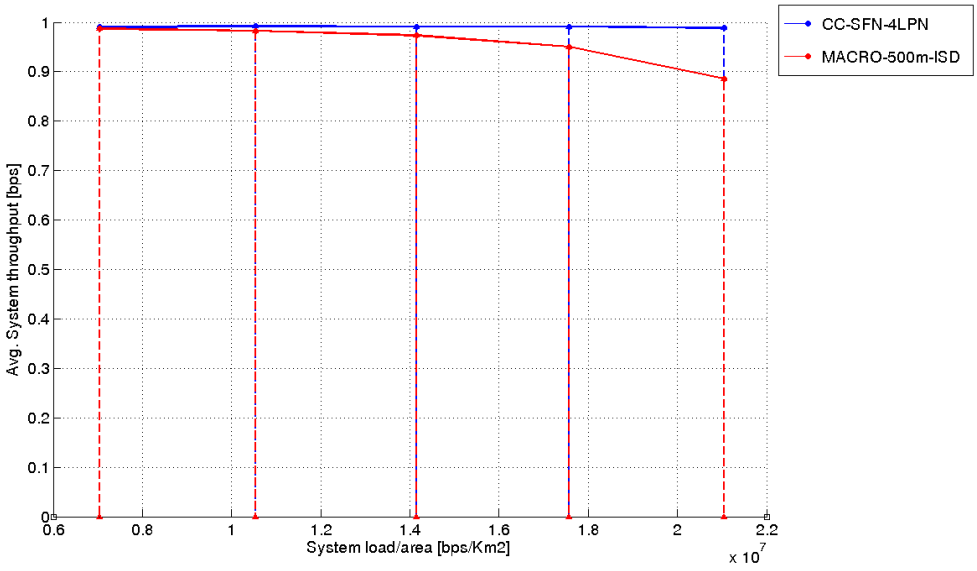


Figure 21. System throughput efficency of the simulated Combined Cell SFN deployment versus the Macro only with 500m ISD.

2.3.2 Uplink/Downlink split within a heterogeneous LTE network

Due to the nature of heterogeneous networks there is a possibility that the best cell in downlink is not the same cell as the best cell in uplink. In such situations using the traditional cell selection method, where the serving cell is the one with the strongest downlink signal strength may result in significantly sub-optimal performance, especially in uplink transmissions. This specific issue emerging with heterogeneous cells is called Uplink/Downlink imbalance. As the macro eNodeB and LPNs in heterogeneous networks have different downlink output powers and the maximum uplink transmission power of UEs is the same regardless of the serving cell, some UEs may find themselves in a UL/DL imbalance situation. Due to the much higher transmission power of macro eNodeB compared to LPN, the downlink cell border between macro eNodeB and LPN is pushed relatively close to the LPN, where both cells are perceived equally strong. In uplink on the other hand the maximum UE transmission power is constant regardless of serving cell, and therefore basically it makes no matter whether the receiving cell type is macro eNodeB or LPN. Therefore if only uplink would be considered, the cell border should be roughly in the middle of the two nodes where the path loss is the same to both nodes. Thus, optimal handover borders are different for uplink and downlink in heterogeneous networks.

The illustration of the UL/DL imbalance situation can be seen in Figure 22. In the figure, the location of the macro eNodeB and the LPN is depicted on the x-axis. The above part of the figure shows the RSRP measurement (based on downlink received signal power) result on the y-axis while the bottom part of the figure shows the path loss on the y-axis.

The RSRP measurement border can be regarded as the ideal downlink cell border since the RSRP measurements are based on downlink measurements, while the ideal uplink cell border would be the border where the path losses are equal since the UE can transmit with the same power regardless of the serving cell. As can be seen from the figure, if the downlink transmission powers of the LPN and macro node are not equal, these borders are not equal.

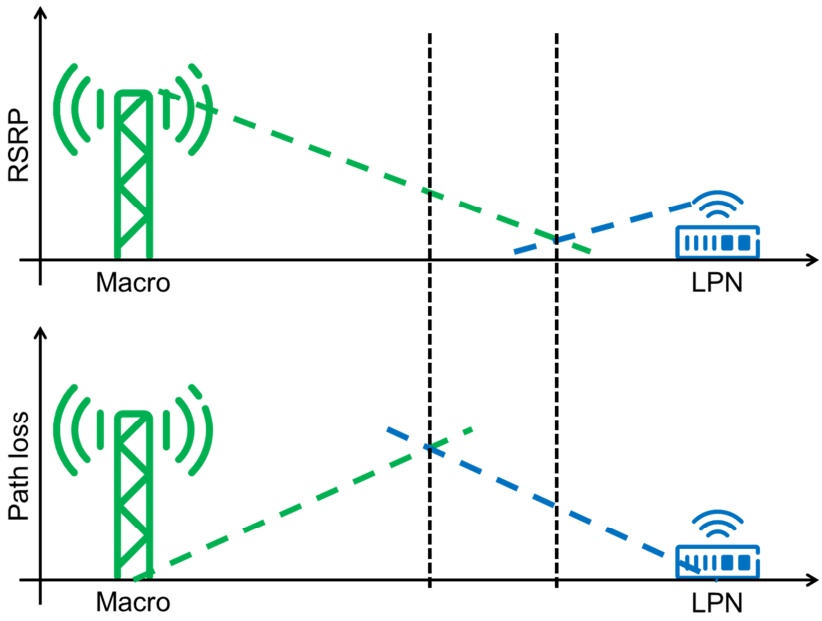


Figure 22. The Uplink/Downlink imbalance with heterogeneous networks.

To improve the efficient usage of radio resources in such situations, one method is called Uplink/Downlink split. In case of UL/DL split the UE would connect both links separately to the best serving cells, in practice the downlink to the best macro cell and uplink to the closest node, whether it is the same macro eNodeB or a LPN. This would improve especially the uplink performance and the utilization of radio resources.

As the transmission power of the UE is limited, it is beneficial to transmit to the eNodeB with the least amount of path loss, whether it is a macro eNodeB or a LPN. Transmitting to the eNodeB with least amount of path loss would increase the transmission capacity due to higher SNR. It could also decrease the interference caused to other UEs, since the users in the cell border areas would be the most probable users to activate the split mode. The power control of those users in the cell border areas still connected to the macro eNodeB in uplink adjusts the transmission power close to maximum due to the long distance and high path loss to the macro eNodeB. In co-channel case this could of course cause interference to other users in neighbouring cells. When such a user activates the UL/DL split and connects the uplink to the LPN, which is much closer, it can adjust the transmission power much lower and its transmissions are now scheduled together with those users it used to interfere, resulting in much less interference.

UL/DL split would also help in load balancing by offloading users located within the CURE (Cell Uplink Range Expansion) area from the highly congested macro cell into the low-power cell.

The user plane split can be done on several protocol levels. In our study the UL/DL split is based on dual-connectivity, currently being specified in 3GPP for LTE Release-12. The separation is done at the PDCP layer, as shown in Figure 23, so there would be separate RLC and lower layers for both nodes, when the split is activated. Control signalling over X2 is used to setup and release the UL/DL split.

The performance of the UL/DL split feature on protocol level will be evaluated by means of system-level simulations. The simulation setup is based on the 3GPP Case 1 specified in [3GPP10a] and on Scenario 2.3.5 described in [D.2.2]. The network deployment consists of 21 macro cells and 84 outdoor pico cells co-located with traffic hotspots. Furthermore, the users are generated according to Poisson process with several arrival intensities, and the users are assumed to be moving with a speed of 3 km/h towards a random direction. Finally, the data traffic is mainly uplink FTP traffic with fixed packet size, but also downlink is studied.

The aim of the simulations, results and their analysis is to provide understanding about the usefulness of the UL/DL split in real world systems and whether or not the feature would actually provide gains especially in the uplink in LTE heterogeneous networks. The usefulness is evaluated according to the achieved gains in, for example, user throughput and delay compared to the case where no UL/DL split is used.

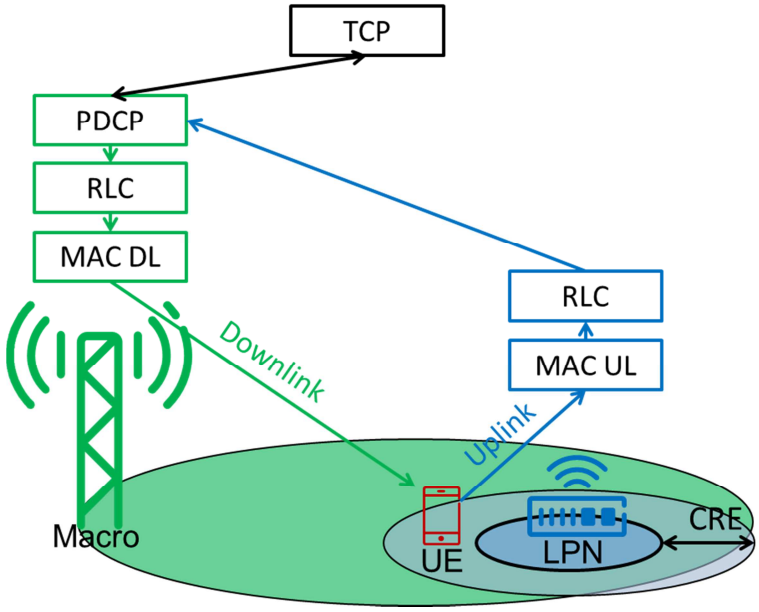


Figure 23. Assumed realization of the uplink/downlink split.

2.3.3 Dynamic Uplink-Downlink optimization in TDD-based small cell networks

We consider a wireless communication system consisting of a set of small cell base stations (SCBSs) $\boldsymbol{B} = \{1, \dots, B\}$. We assume that a user equipment (UE) arrives at location x within the considered geographical area according to a Poisson arrival process with rate $\lambda(x)$. Each UE requests either a DL or UL file whose size follows an exponential distribution with mean $1/\mu(x)$. A closed-access policy is assumed in this work, meaning that each SCBS has its own subscribed UEs, and hence no handover is considered. We further assume L_b to be the coverage area of an SCBS b, where a UE at location x is served by an SCBS b if $x \in L_b$. We assume that the system operates in TDD mode. A time frame consists of a number of N_f subframes. A frame is divided into two portions, UL portion and DL portion. Each portion consists of a group of subframes dedicated to serving either UL or DL traffic. A switching point w_b is defined as the point in which SCBS b switches from UL mode to DL mode. There are a number of $N_f - 1$ possible switching points for a frame length of N_f, then $w_b \in \{1, \dots, W_f\}$, where $W_f = N_f - 1$. For any of these possible switching points, there will be at least one subframe for UL and for DL in each frame.

We define the vector $w = [w_1, w_2, \dots, w_B]$ as the vector of switching points for all SCBSs in the system. Varying switching points asynchronously in different cells may cause opposite transmission directions in different cells which leads to cross-link interference (i.e. UL-to-DL interference and DL-to-UL interference). Consequently, the Signal-to-Interference-plus-Noise-Ratio (SINR) for UL and DL, respectively, for a receiving node at location $x \in L_b$ is given by:

$$\Gamma_b^{UL}(x) = \frac{p_b^{UL} h_{b,b}(x)}{\sigma^2 + \sum_{j \in B_{UL} \setminus \{b\}} p_j^{UL} h_{j,b}(x) + \sum_{k \in B_{DL}} p_k^{DL} h_{k,b}(x)} \quad (40)$$

$$\Gamma_b^{DL}(x) = \frac{p_b^{DL} h_{b,b}(x)}{\sigma^2 + \sum_{j \in B_{UL}} p_j^{UL} h_{j,b}(x) + \sum_{k \in B_{DL} \setminus \{b\}} p_k^{DL} h_{k,b}(x)} \tag{41}$$

where p_b^{UL} (p_b^{DL}) is the UL (DL) power from the serving node b, p_j^{UL} (p_j^{DL}) is the UL (DL) power from the interfering node j, $h_{m,b}(x)$ is the channel gain, including pathloss, between the transmitting node in SCBS b and the receiving node in location $x \in L_b$, B_{UL} and B_{DL} are the sets of cells operating in UL and DL, respectively, and σ^2 is the noise variance. Furthermore, the data rates of a UE at location $x \in L_b$ for UL and DL, respectively, are given by:

$$c_b^{UL}(x) = f_b \log_2\left(1 + \Gamma_b^{UL}(x)\right) \tag{42}$$

$$c_b^{DL}(x) = f_b \log_2\left(1 + \Gamma_b^{DL}(x)\right) \tag{43}$$

where f_b is the bandwidth allocated to that UE. The system-load density at location x is defined as

$$\varrho_b^{(l)}(x) \coloneqq \frac{\gamma^{(l)}(x)}{c_b^{(l)}(x)} \tag{44}$$

where $l \in \{UL, DL\}$ and $\gamma^{(l)}(x) = \lambda^{(l)}(x)/c_b^{(l)}(x)$ is the load density at location x. The cell load density for cell b is defined as the time delay needed to serve the UL and DL traffic as follows:

$$\rho_b^{(l)}(w_b) = \frac{1}{\delta^{(l)}(w_b)} \int_{x \in L_b} \varrho_b^{(l)}(x) dx \tag{45}$$

where $l \in \{UL, DL\}$, $\delta^{(l)}(w_b)$ is the UL or DL duty cycle, which is the fraction of time frames dedicated to either UL or DL service within a frame, and is expressed as follows:

$$\delta^{(l)}(w_b) = \begin{cases} \dfrac{w_b}{W_f} & l = UL \\ \dfrac{W_f - w_b}{W_f} & l = DL \end{cases} \tag{46}$$

Here, dividing each cell load by its respective UL or DL duration is done in order to account for the UL/DL effective traffic. Therefore, lower duty cycles lead to higher delays and vice versa. Our objective is to find the vector of switching points w that minimizes the overall average flow delay by minimizing $\sum_b \frac{\rho_b}{1-\rho_b}$ over the entire time frame. Therefore, we define a cost function that reflects the flow delay average over the whole subframes within a timeframe, calculated as follows:

$$J(\boldsymbol{w}) = \sum_{b=1}^{B} \left(\frac{1}{w_b} \sum_{j=1}^{w_b} \frac{\rho_{b,j}^{UL}(w_b)}{1 - \rho_{b,j}^{UL}(w_b)} + \frac{1}{W_f - w_b} \sum_{j=w_b+1}^{w_f} \frac{\rho_{b,j}^{DL}(w_b)}{1 - \rho_{b,j}^{DL}(w_b)} \right) \tag{47}$$

Thus, we can define the following cost optimization problem:

$$\begin{aligned} &\underset{w}{\text{minimize}}\, J(\boldsymbol{w}) \\ \text{subject to}\quad &0 < \rho_{b,j}^{UL}(w_b) < 1,\ \ \forall b \in B \\ &0 < \rho_{b,j}^{DL}(w_b) < 1,\ \ \forall b \in B \end{aligned} \tag{48}$$

To solve this problem, we develop a distributed algorithm which dynamically optimizes the UL/DL configuration. The goal is to design a decentralized algorithm that selects a vector of switching points w that minimizes the cost function. With the lack of global network information, the algorithm must rely only on the local information available at each SCBS to optimize an individual cost function rather than the global cost. However, the cost function for each SCBS depends not only on its own traffic load but also on the interference experienced from neighboring cells. Therefore, each SCBS b should learn to estimate its cost function and use this estimated cost function to update its strategy. Here, an SCBS's strategy is essentially the selection of a switching point. In view of the interference coupling between neighboring cells, the performance of each SCBS depends not only on its choice of switching points, but on other SCBSs' choices as well. Therefore, we model this problem as a strategic noncooperative game where the set of players are the SCBSs, in which each of them selects its action $a_b^{(n_b)}$ where N_b is the number of possible actions, which corresponds to the number of switching points W_f in our problem. For each base station $b \in B$, the corresponding cost function can be expressed as follows:

$$J_b\left(a_b^{(n_b)}, \boldsymbol{a}_{-b}\right) = \frac{1}{w_b}\sum_{j=1}^{w_b}\frac{\rho_{b,j}^{UL}}{1-\rho_{b,j}^{UL}} + \frac{1}{W_f - w_b}\sum_{j=w_b+1}^{w_f}\frac{\rho_{b,j}^{DL}}{1-\rho_{b,j}^{DL}} \tag{49}$$

Each player b chooses an action following a mixed strategy profile π_b, which is a vector of probability distributions over the set of possible actions A_b. Let the strategy of choosing an action $a_b^{(n_b)}$ by player b at a time frame t be the probability that this action is selected $\pi_{b,a_b^{(n_b)}}(t) = \Pr\left(a_b(t) = a_b^{(n_b)}\right)$. Then, by randomizing the action selection following their mixed strategies, players aim at minimizing their long-term (expected) cost functions given by:

$$\bar{J}_b(\boldsymbol{\pi}_b, \boldsymbol{\pi}_{-b}) = \sum_{a\in A} J_b\left(a_b^{(n_b)}, \boldsymbol{a}_{-b}\right)\prod_{j=1}^{B}\pi_{j,a_j^{(n_j)}} \tag{50}$$

In this game, each SCBS will choose the action that can lead to minimizing its cost function J_b, given other players' actions. We propose an algorithm that captures this behaviour by adopting the Gibbs Sampling-based probability distribution, in which the probability of playing an action $a_b^{(n_b)}$:

$$\Lambda_{b,a_b^{(n_b)}}(\boldsymbol{a}_{-b}) = \frac{\exp\left(-\beta_b J_b\left(a_b^{(n_b)}, \boldsymbol{a}_{-b}\right)\right)}{\sum_{m=1}^{N_b}\exp\left(-\beta_b J_b\left(a_b^{(m)}, \boldsymbol{a}_{-b}\right)\right)} \tag{51}$$

where β_b is a Boltzmann's temperature coefficient. In what follows, we evaluate the performance of the proposed dynamic TDD algorithm. To illustrate the gains of the proposed scheme, we compare it against two baseline schemes: 1) fixed TDD frame, in which small cells are assumed to have the same synchronous TDD frame, with equal UL and DL duty cycle, and 2) random TDD frame, in which the switching point is varied randomly. We consider an arbitrary number of SCBSs distributed randomly, and underlaying the macrocell coverage area. In this work, we focus only on the SCBS-to-SCBS co-channel interference scenario. The bandwidth is assumed to be shared between all SCBSs. The bandwidth is assumed to be divided equally between all UEs transmitting or receiving in a given subframe.

Both SCBSs and UEs are assumed to transmit with their maximum power, and hence, no power control is considered here. We use the average packet throughput as the performance measure for different schemes, which is defined as the packet size divided by the delay encountered to complete its transmission. The motivation behind this is that it captures both packet rate and delay, which is the objective of the proposed scheme. To investigate the asymmetric UL/DL traffic, we conduct simulations for different mean UL-to-DL ratios. For example, UL-to-DL ratio of 0 dB means that the average rate requirement λ_b/μ_b is the same for UL and DL. Each SCBS uses a sequence of time frames, no more than a maximum of 200 frames to learn its load and update its UL/DL configuration accordingly. In general, the performed simulations are based on scenario 2.2.1 defined in [D.2.2].

In Figure 24, we compare the packet throughput performance of our scheme against the two baseline schemes for different UL-to-DL ratios. All SCBSs are assumed to have the same average UL-to-DL ratios while the instantaneous traffic is different. Figure 24 shows that the proposed scheme achieves significant gains reaching up to 200% at −20 dB compared to the random scheme in all traffic conditions. Moreover, the figure also shows that our approach outperforms the fixed scheme in case of asymmetric traffic conditions. The gain increases as the level of asymmetry increases, since the SCBSs are able to learn their UL and DL loads and adapt their transmissions accordingly. Figure 24 also shows that the proposed algorithm achieves up to 97% gain at −20 dB over the fixed assignments. However, the gain becomes smaller in the symmetric traffic case in which the fixed scheme is shown to achieve the same performance since it allocates equal resources to UL and DL and hence it is suitable for symmetric traffic.

Figure 25 shows the average packet throughput for the case in which half of the cells have opposite UL-to-DL ratios compared to the other half. For example, if the first half has a ratio of 10 dB, the second half has a ratio of −10 dB. This scenario is challenging in the sense that it is associated with high cross-link interference. In Figure 25, we can see that the proposed scheme achieves considerable gains over both the random and fixed schemes. Clearly, the proposed algorithm is able to find a balance between selecting the switching point that matches the SCBS load and avoiding configurations that are associated with high cross-link interference. Figure 25 shows that the proposed approach achieves gains reaching up to 145% and 53% over the random and fixed schemes, respectively in the case of UL-to-DL ratio of 20 dB.

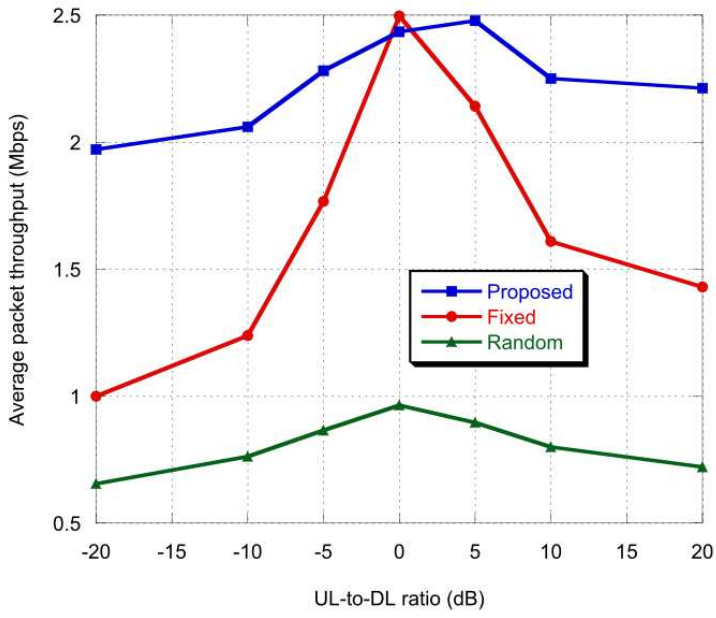


Figure 24. Packet throughput performance in case of cells having the same UL-to-DL ratio for a network with 4 SCBSs.

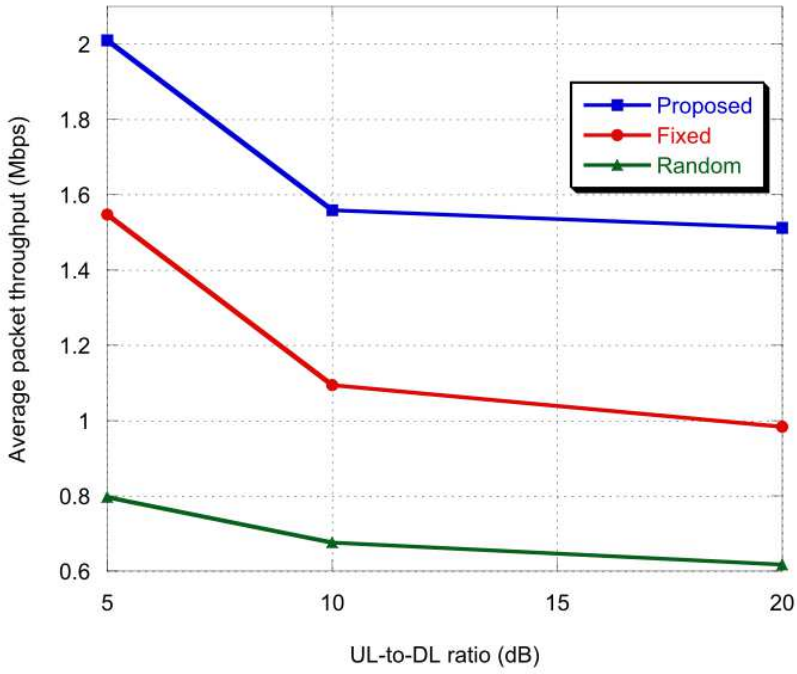


Figure 25. Packet throughput performance in case of cells having opposite UL-to-DL ratio for a network with 4 SCBSs.

2.4 Backhaul offloading by proactive caching

The rapid proliferation of smartphones has substantially enriched the mobile experience, leading to new wireless services (e.g. multimedia streaming, web-browsing applications), and content owners (Telco's, users, and OTT). Currently, mobile video streaming accounts for 50% of mobile data traffic and is expected to skyrocket to a 500X increase over the next ten years. These new paradigms urge mobile operators to redesign their networks cost-effectively by deploying intelligence at the network edge. Despite these exponential traffic growths, results have shown that multiple users actually request similar contents. At the same time, dynamic content caching has recently attracted much attention and is considered as one of the most disruptive technology directions for beyond 4G networks. By harnessing recent advances in storage and computing, dynamic caching can help alleviate backhaul congestion, reduce loads at peak times and minimize latency, by pre-caching contents at strategic network edge locations. If smartly coupled with meta-data analytics, network operators can further exploit the vast amount of users' context information (location, speed, etc.) for a better predictability of future demands, to proactively cache popular contents before users actually request them. We first assume that small base stations are deployed with storage capabilities but have limited capacity backhaul links. We then describe the model and define a Quality-of-Experience (QoE) metric in order to satisfy a given file request. Figure 26 illustrates the concept of backhaul offloading by proactively fetching contents during low-peak demands.

An optimization problem is formulated in order to maximize this QoE metric for all users' requests under backhaul and storage constraints. The proposed proactive caching procedure relies on the popularity statistics of the requested files. Since not all requested files can be cached due to storage constraints, the algorithm selects the files with the highest popularities until the total storage capacity is achieved. The results will be evaluated by means of synthetic traces and possibly real data traces.

This innovation is related to scenario 2.3.9 in [D.2.2].

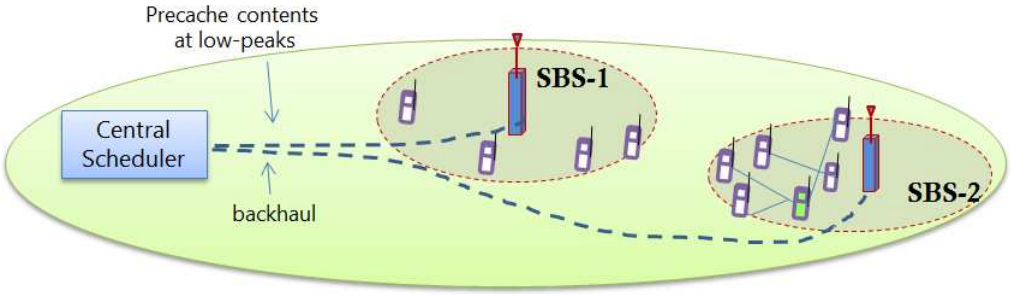


Figure 26. Illustration of backhaul offloading via caching.

3 INTER-SYSTEM RADIO ACCESS OFFLOADING

An efficient and cost-effective integration of cellular (for example WCDMA and LTE) and WiFi technologies, referred to as inter-RAT (or inter-system) offloading, has recently attracted significant interest from academia, industry, and standardization bodies alike, whereby the complementary benefits of both RATs can be leveraged. On one hand, due to the uncontrolled, unlicensed nature of WiFi, the competition for resources among a potentially large number of hotspot users and other devices transmitting on the same unlicensed band can yield dramatically poor throughput. In such a scenario, offloading some of this traffic to a well-managed small cell network operating over the licensed spectrum can improve the performance. On the other hand, the inherent constraints of small cell networks, particularly due to cross-tier and co-tier interference, motivate offloading some of the traffic to the WiFi band both to reduce the interference and to ease the congestion. Recently, with the advent of dual-mode small cell base stations operating on both cellular and WiFi bands, smart traffic offloading strategies harnessing the benefits of both RATs is seen as instrumental in taming the 1000X traffic demands.

Thus far, both cellular and WiFi RATs/systems have been in constant competition, until recently when a tighter integration of both technologies has emerged as a necessary paradigm. Indeed, when deployed along with each other, operators can not only perform classical offload (through WiFi), but also a smart fine-grained offload, whereby operators can decide which traffic flows over which RAT, while leveraging users' QoS requirements, latency, and backhaul conditions. In addition, with the advent of dual-mode small cell base stations, operators can reduce their site acquisition costs by combining several RATs into a single device. This can further lead to a reduction in their capital expenditures.

In order to reap the benefits of small cell's multimode capabilities, operators need to devise dynamic offloading strategies and make intelligent decisions aiming at enriching users' QoS and avoiding user churn. For instance, when a user equipment (UE) discovers the presence of WiFi in its vicinity, delay-tolerant traffic (for example web browsing) should be offloaded to WiFi, whereas traffic with more stringent data requirements (for example multimedia) would remain on the cellular 3G/LTE RAT. Noteworthy is that while traffic offloading at the access level is important, backhaul offloading is yet another important component of the cellular-WiFi integration. Here, operators need to jointly take into account the backhaul conditions and congestion level in their offloading policy before deciding to which RAT a user is offloaded, in order to ensure a seamless user experience.

In SHARING Task 4.2, smart traffic offloading and steering strategies taking into account users' QoS requirements (rate and delay requirements), indoor/outdoor locations, terminal speed, users' contexts, interference levels, and network loads will be investigated. Both UE-assisted and network-assisted offloading scenarios taking into account network parameters such as latency, reliability, user traffic and mobility patterns, network capacity, congestion, content type (video, picture, voice, data, etc.), packet loss, channel characteristics, will be explored. A decentralized and dynamic traffic offloading framework, in which small cells seamlessly steer their traffic between cellular and WiFi RATs, will also be proposed. Of interest is quantifying the improvements in terms of cell throughputs and cell-edge throughput, for the "macro-only", "HetNet," "WiFi-only" and "HetNet+WiFi" offloading strategies. Task 4.2 will also study efficient WiFi-LTE interconnections for fast handover and propose traffic steering strategies between LTE and WiFi, leveraging the prediction of QoS and power consumption on WiFi and LTE, to enable QoS-aware and energy-efficient load balancing. Finally, Task 4.2 will look into the use of ANDSF for load balancing, procedures for interference control for multi-RAT, mobility management, backhaul-aware load balancing, and UE-specific load balancing policies.

3.1 3GPP-WiFi offloading

3.1.1 Inter-LTE offloading to WiFi by middleware deployment

WiFi technology offloading can alleviate saturation of almost 60% of mobile data traffic from smartphones and tablets, which in recent times are the devices monopolizing the access to the network. In 2013 WiFi accounted for more than half of global Internet traffic, and it was estimated that there existed over 6 million public WiFi hotspots. The forecast announces an increase estimated at 12.2 million of these hotspots (93% increase) by 2017. Last year, this kind of traffic was distributed on a daily basis tending to the pattern shown in Figure 27, provided by the studies and analysis carried out by GOWEX [GOW1].

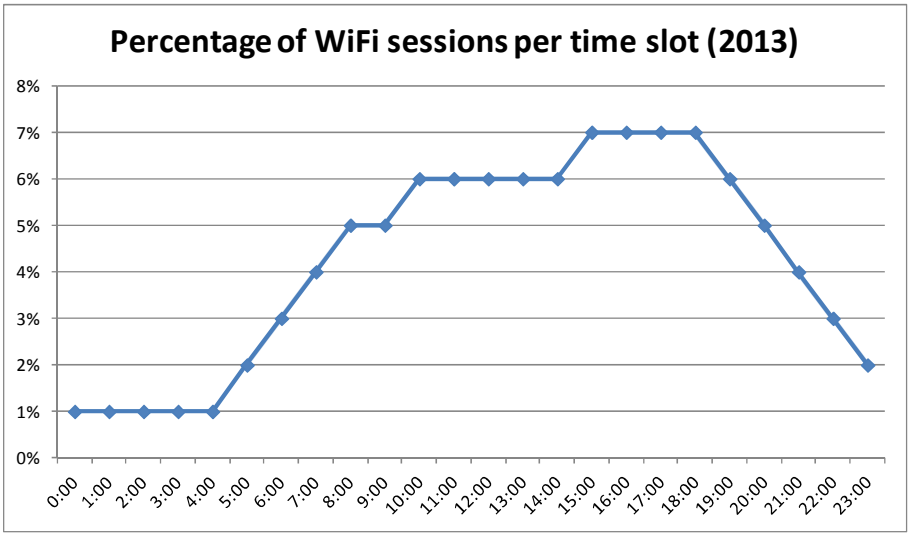


Figure 27. Percentage of WiFi sessions per time slot

It is a proven fact that WiFi users require mobility, since almost 9 out of 10 connections are taking place from mobile devices, a trend that is unlikely to change in the future, but rather the contrary.

The most important advantages of the WiFi offloading include the involved cost savings, as it entails a saving of 75% per byte downloaded in WiFi networks and a 50% reduced OPEX costs, not having to buy any additional equipment. WiFi offloading also improves the user experience, allowing a better connection to be ensured in areas of high mobile saturation. And finally, WiFi offloading helps to increase the coverage. Today different mobile operators complement their mobile broadband networks with WiFi and in that way extend the wireless coverage.

Knowing these facts and looking to the way forward, the goal pursued within this innovation consists of developing a valid technique to manage user mobility between the LTE cell and small WiFi cells, making the process totally seamless and transparent to the customer and assuring the maintenance of a notable Quality of Service (QoS).

Thus, the envisioned scenario to work for the design and development of these load balancing procedures is similar to the one presented in Figure 28. This scenario is tightly related to the one shown in Section 2.2.2, but adding an additional degree of complexity. The studies will follow the path taken in Section 2.2.2, implementing a middleware to be deployed in the different entities involved in the process which will be capable of handling the switching process. In Figure 28, WiFi deployments coincide with the cell edges, but they can be placed at any point inside the map. The middleware will be in charge of determining at each point and for each user which is the best base station to be attached to from the list of potential candidates. The introduction of WiFi enables a plethora of new considerations, regarding the availability of WiFi interfaces at each UE, the need of SLA/latency associated to each user or even the available battery at each device.

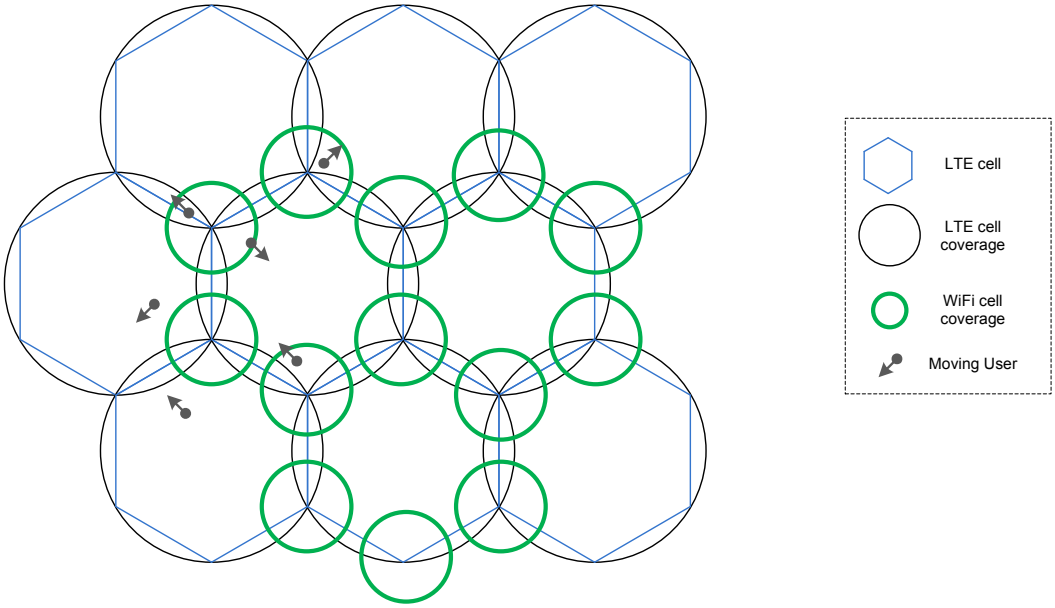


Figure 28. 3GPP-WiFi load balancing scenario

As depicted in the Section 2.2.2, the middleware which will be designed and deployed will be able to handle the needed interactions among the diverse actors present in the scenario, having the internal architecture there specified and following a communication process similar to the one depicted in Figure 16.

This innovation is related to scenario 2.9.1 in [D.2.2].

3.1.2 Seamless offloading in heterogeneous wireless networks

The increase on smartphone usage has brought the burden of data traffic with it. Operators are looking for cost-effective solutions to overcome the problem of 3G infrastructure for high contention traffic scenarios. Several schemes were offered to save the moment, and they brought some extra costs, including deploying femtocell or WiMax, LTE, LTE-Advanced systems along with their expensive equipment. On the other hand, operators are expanding their networks with 802.11 technologies such that they can exploit the free-band communication. There are several solutions that can be applied to improve operators' network performance including traffic management, backhaul and infrastructure upgrades. However, network operators realize that such options provide short-term relief and only apply as long as its user stays inside the operator’s network. They would like to have a comprehensive solution that addresses user behaviour in the real-world, i.e. which supports roaming from one network type to another, and allows users to receive and enjoy quality services from their operator regardless of their location and choice of network access.

At this point, mobile data offloading provides a solution with real business value. Significant cost reductions and improvement of the operator's reach to its customers gives a competitive edge to it. In this study, we will evaluate and simulate collaborative type of mobile data offloading scheme for real-time network (RTN) systems. In this scenario, we will develop a seamless offload scheme between 3GPP and WLAN networks for real-time packet traffic. This solution would carry heterogeneous mobile broadband networks' fundamental identifiers such as application targets, user's mobility and usage characteristics, network capacity and density, network management, terminal's power state, network congestion status, physical channel parameters to a mathematical plane.

With this model, by taking into account the base protocol functions such as access, error and congestion control, network and channel capacity, data transmission latency and reliability

parameters will be improved based on the upper and lower limits of them. The scenario will allow us to contribute to the following points:

1. User preference as a Quality of Experience (QoE) metric during handover decision making will be integrated into the developed multiple attribute decision making (MADM) algorithm.
2. Handover execution will be handled not only based on link-quality including quality of Service (QoS) values, but also based on subjective measures.
3. Network traffic of the heterogeneous wireless networks (HWNs) will be improved and managed based on user experience.

The typical scenario is depicted in Figure 29, and it consists of UTRAN/GERAN access networks along with a WiFi access point in a tightly-coupled heterogeneous network architecture.

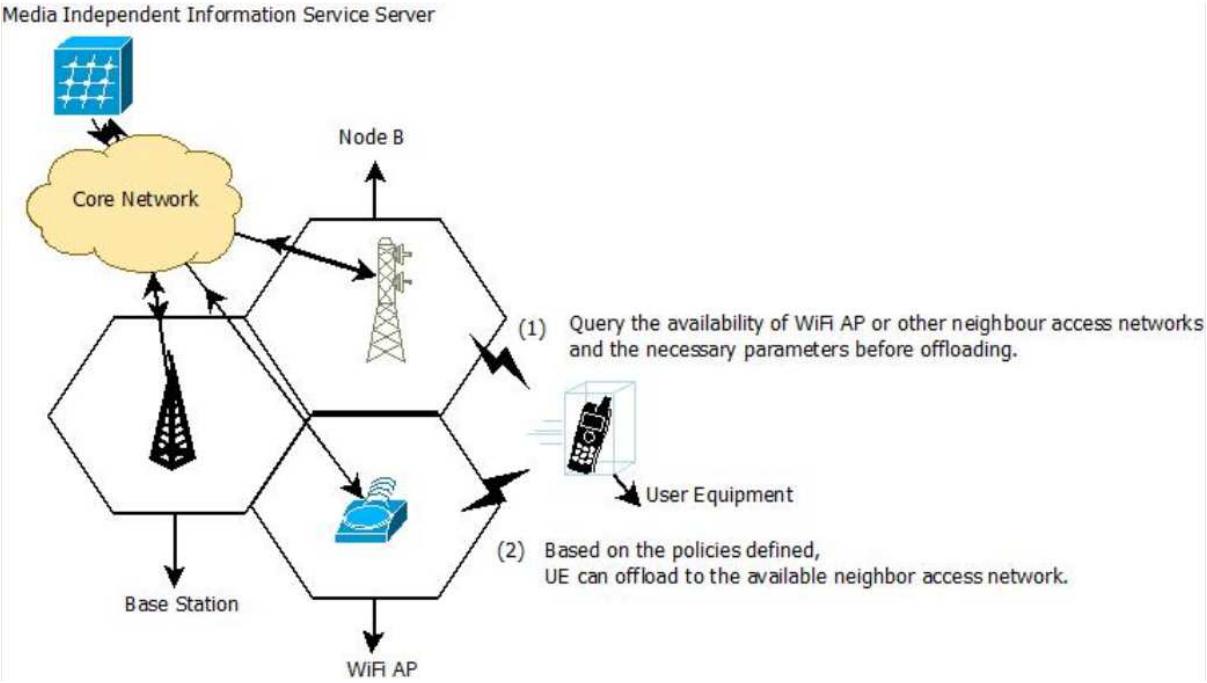


Figure 29. IEEE 802.21 and user experience based offload scenario.

Essentially, the resource monitoring helps decision maker to identify and discover the access networks along with a QoS and/or QoE mechanism when a user changes its location. The decision maker could be user-equipment (UE) or network, meaning either each cell will broadcast its connection information to UEs or UE will retrieve the cell information from both 3GPP and non-3GPP networks. In 4G networks, access network discovery could be controlled by Evolved Packet Core (EPC) along with access network discovery and selection function (ANDSF). However, 3G networks are missing such a core system and require either an additional device or an additive functionality to the existing serving nodes in the infrastructure or a user-centric approach. 3GPP or trusted/untrusted non-3GPP (WiFi) networks could be discovered and monitored by this functionality.

As for decision making functionality, UE or Mobile Network Operator (MNO) selects the access networks by considering probabilistic demands ideally. Network related, terminal related, user related and application related metrics need to be considered pertaining to vertical handover decision. However, the paramount elements amongst them are the user-related ones as access network alters, users should approve the changes such as throughput, energy consumption of the terminal, security etc. of the suggested network. Otherwise, it would infringe on their right to privacy because for the very same application an adult's decision

would differ from that of a young person, for instance, security-wise an adult might not prefer to watch videos through WEP or WPA on WiFi networks but EAP-SIM on 3GPP network. Maybe this choice could be trivial for a young person and actually he would prefer a free communication band, but considering the rising security challenges of today's world, operators need to pay attention tothe subject for each subscriber. We can classify the handoff decision criteria as below:

1. **Network-related**: it represents network conditions: coverage, bandwidth, latency, link quality (CQI), BER (Bit Error Rate), cost, security level.
2. **Terminal-related**: velocity of the user, battery power,
3. **User-related**: user profile (i.e. student or businessman) and preferences (i.e. Cost or security),
4. **Service-related**: service capabilities, QoS, QoE, security level.

The handoff decision algorithm aims at selecting a network for a particular service that can satisfy objectives based on some criteria (such as low cost, good RSS, high MOS, optimum bandwidth, low network latency, high reliability and long life battery) and taking into account the preferred access network of user. Some techniques used for network-centric solutions such as stochastic programming, game theory and utility function could be performed in this respect. In client-centric solutions, analytical hierarchy process helps ranking the networks based on QoS by checking user's requirements and network conditions. Consumer surplus is an economical model, and could be used to find the network benefits the best-generally used for non-real-time traffic. In profit function, the difference between bandwidth gain and handoff cost for each network is computed, and the most appropriate network is found. As for the collaborative models, fuzzy logic controller ranks the candidate networks based on the user's selection criteria, network data rate and SNR.

Some other problems related to user-centric models are: a device would connect at Layer 2, but not at network layer. Also, UEs would connect to one of APs available based only on signal strength, and end-up with wrong assignment such as application class or QoS requirement are not met. Adding the increasing number of interfaces such as WiMax, WiFi and cellular network, the burden of UE would extend to cover multiple interfaces.

IEEE 802.21 Working Group was formed to overcome the diversity in the handover mechanisms and to eliminate user-centric methods' drawbacks, and therefore, a common MIHF (Media Independent Handover Function) was introduced. MIHF is an abstraction layer between layer 2 and layer 3. All interfaces on L2 (cellular, WiMax, LTE, WiFi) could communicate with MIHF, and MIHF could transmit the necessary messages to L3 and above such as SIP, MIPv4, MIPv6 or HIP. In IEEE 802.21 framework, there are several factors to determine the handover decision including service continuity, application class, QoS, network discovery and selection, security, power management, and handover policy. In Figure 30, MIHF architecture is shown. The most important advantage MIHF brings is not only to provide both L2 and L3 handover but also to allow running make-before-break soft handover mechanism which in return supply us Mean Opinion Score (MOS) calculation time before the final break execution with the connected access network.

Service delivery and user experience are strongly related items for operators in terms of radio resource management. Technical aspects targeting this issue relate to QoS parameters that can be handled by the platform, at least partially. Subjective psychological issues and human cognitive aspects are typically unconsidered aspects and they directly determine the QoE.

It does not matter how smoothly packets move through the network, if the users find out that services and applications don't meet expectations. Planning must address the factors that underlie QoE for each service that runs on the network, as well as any interactions or inconsistencies between them. For this work, we do not consider the problems due to the application's packetization and/or encoding/decoding schemes which could also affect QoE, but only focus on the optimization of the network traffic for the heterogeneous networks.

This innovation corresponds to scenario 2.4.3 described in [D.2.2].

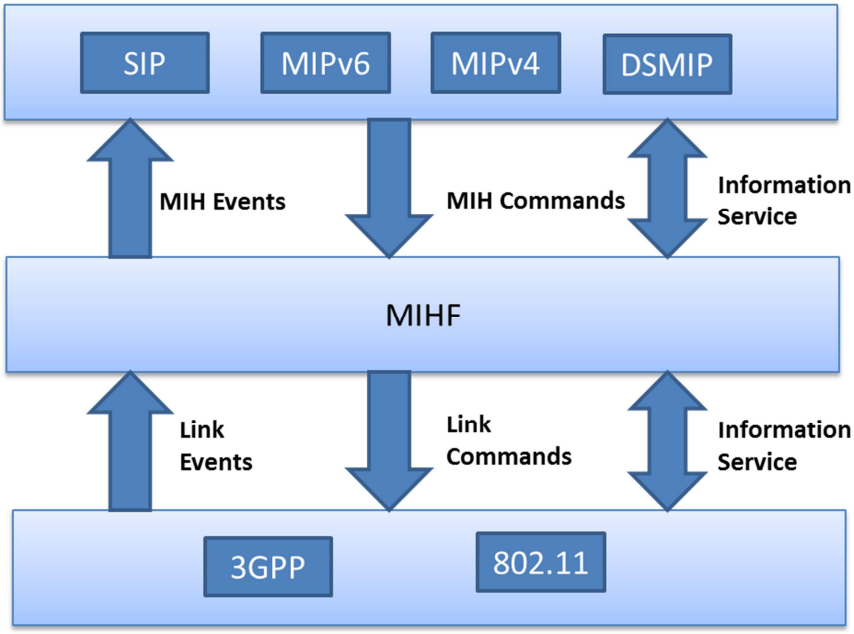


Figure 30. MIHF Architecture

QoE Model

In this study, we used TOPSIS (Technique for Order Preference by Similarity to Ideal Solution), due to its easy implementation, as a way of selecting the best target network for a given user's video application. The decision to use this algorithm was made based on the other multiple attribute decision making (MADM) algoritms' performance comparison results. By using this algorithm, we trigger the MIH events such as connect-link or disconnect-link to execute the handover seamlessly. For this purpose, first, we created a decision matrix:

$$A_{ij} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \quad (52)$$

In A_{ij} matrix (i-row, j-column), n refers to decision points such as link quality, MOS of the target network for the given application, user preference (cost security), and m refers to the target networks which are UMTS or WLAN. In second step, we formed a normalized decision matrix by using the following equation:

$$r_{ij} = \frac{a_{ij}}{\sqrt{\sum_{k=1}^{m} a_{kj}^2}} \quad (53)$$

We obtained the R matrix:

$$R_{ij} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix} \quad (54)$$

Then, we created a weigted normalized decision matrix by multiplying each column of the matrix by corresponding weight w_i where $\sum_{i=1}^{n} w_i = 1$ by using the following equation:

$$v_i = w_i * r_i \tag{55}$$

TOPSIS method assumes that each evaluation factor has a monotonically increasing or decreasing tendency. Next, we formed the positive (A^*) and negative (A^-) solutions by using the following formula:

$$A^* = \left\{\left(\max_i v_{ij} \,\middle|\, j \in J\right), \left(\min_i v_{ij} \,\middle|\, j \in J'\right)\right\} \tag{56}$$

$$A^- = \left\{\left(\min_i v_{ij} \,\middle|\, j \in J\right), \left(\max_i v_{ij} \,\middle|\, j \in J'\right)\right\} \tag{57}$$

For both formula, J refers to benefit (max), and J' refers to lost (min). For both solution set, the number of evaluation factor consists of m elements.

We ended up with the respective sets of $A^* = \{v_1^*, v_2^*, \dots, v_n^*\}$ and $A^- = \{v_1^-, v_2^-, \dots, v_n^-\}$.

Then we calculated the Euclidean distance S_i^* of each alternative a_i from the positive point and S_i^- of each alternative a_i from the negative point A^-. Both positive and negative set consists of the number of evaluation factors, that is, m elements. The calculations are shown respectively as:

$$S_i^* = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^*\right)^2} \tag{58}$$

$$S_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2} \tag{59}$$

In the next step, we calculated the relative similarity of the alternatives from the positive and negative point which is done in the following manner:

$$C_i^* = \frac{S_i^-}{S_i^- + S_i^*} \tag{60}$$

where $0 \le C_i^* \le 1$. If C_i is close to value 1 the solution is closer to ideal.

Performance Results

For our simulation, we used NS 2.29 simulator integrated with EURANE, NIST and EVALVID packages to evaluate the video performances in a heterogeneous network during a handover execution where the TOPSIS algorithm results were utilized. Decision parameters of TOPSIS are as follows:

(i) MOS: Mean Opinion Score is considered as a subjective measure. Currently, it is more often used to refer to one or another objective approximation of subjective MOS. Although all "MOS" metrics are intended to quantify QoE performance and they all look very similar (values between one and five with one or two decimal places), the various metrics are not directly comparable to one another. ITU P.800 and P.830 define the MOS scale as showed in Table 4.

(ii) PSNR (dB): The peak signal-to-noise ratio is used as an objective measurement of the restored image quality. PSNR is most commonly used to measure the quality of reconstruction of lossy compression codecs which is in our case MPEG-4.

$$PSNR = 20\log\frac{Vpeak}{MSE} \quad (61)$$

$Vpeak = 2^k - 1$, where k is equal to the number of bits per pixel (luminance component). MSE is the mean squared error.

(iii) CQI: Channel quality indicator is reported by UE and is calculated using BLER and SNR values. It is a vital parameter to estimate the UMTS air interface quality. The UE type that is assumed in the simulator is 3GPP UE category 1 to 6. In our simulation, the highest CQI value was accepted as 22. However, it varies between 1 and 22.

(iv) QoS: Quality of service level of the access point (AP) is utilized in the algorithm to determine the link-quality of WiFi network. Voice = Platinum = 6, Video = Gold = 5, Best Effort = Silver = 3, Background = Bronze = 1

(v) Security Policy used in WiFi network: WPA or WPA2 cannot be used for a seamless solution. EAP-SIM is required to do so.

(vi) Channel Utilization: It is a WiFi network parameter, and is monitored for a stable traffic level and to prevent under or over utilization.

(vii) Client SNR: Signal-to-noise ratio is a critical and widely used metric to obtain the experienced WiFi quality per user.

(viii) User Preference: For a businessman security and quality level could be extremely important whereas for a student the cost is of the utmost importance.

Table 4. ITU-R Quality and Impairment Scale

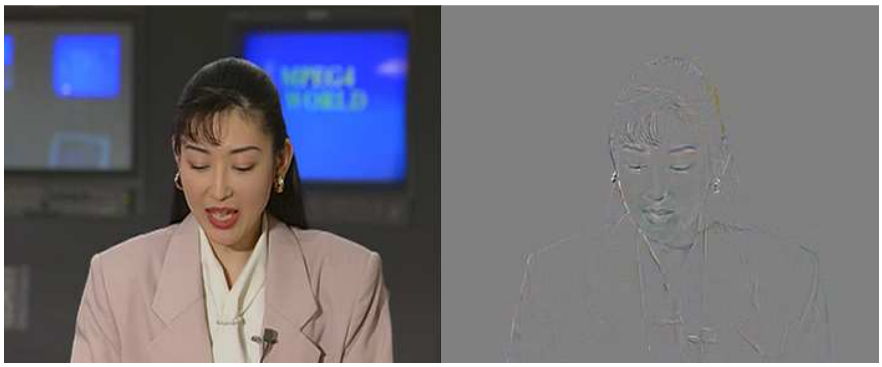
Scale Impairment	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

QoE assessment could be performed with subjective tests with humans, but by using this scheme we cannot make a handover execution in real-time, other approaches to the problem of QoE assessment includes utilizing objective testing to predict the MOS value of a service. These solutions need original signals (for real time applications e.g., ITU-T objective measurement standards like PESQ (P.862), E-model (G.107) etc.) and are computationally complex [Rec_2001][Rec_2003]. Therefore, we calculated the PSNR frame by frame and map it to the corresponding MOS value as in Table 5.

Table 5. PSNR to MOS mapping

PSNR [dB]	MOS
> 37	5 (Excellent)
31 - 37	4 (Good)
25 - 31	3 (Fair)
20 - 25	2 (Poor)
< 20	1 (Bad)

When simulating our heterogeneous network in tight-coupling architecture, we used a case where a video is downloaded in the beginning. Our user was connected to an UMTS network and in this network throughput was 45 kb/s which is not even acceptable for voice networks since 64 kb is used for bearer payload whereas 16 kb needs to be used for signaling purposes in each direction, making a total of 80 kb/s at least for a good quality voice traffic. Considering our traffic, video requires a lot more throughput for an acceptable communication. As a result, we obtained the following frames in Figure 31.



(a) (b)

Figure 31. (a) Transmitted Common Intermediate Format (CIF) 352x288 MPEG-4 XviD (b) Received CIF 352x288 MPEG-4 XviD

The speed of the UE was 1 m/s, and after 2 seconds where video is transmitted with a rate of 30 fps, the MIH module discovered the WLAN network and requested the target WLAN network's metrics regarding QoS (jitter, delay, packet loss) and also user preference such as security (i.e. EAP-SIM or WPE) along with MOS. Based on the TOPSIS algorithm, user was attached to the target WLAN network by using MIH functionality at frame of 60. Between frames 60 and 300, user experienced MOS values between four and five. The graphical representation of the scenario can be found in Figure 32.

Furthermore, we calculated the packet loss rates in percentage for the same scenario while user experiences different throughput rates in UMTS network. The loss rates were measured to check the impact of an UMTS network before the handover execution and the impact of these losses for the full transmission. Our encoded video is composed of two distinctly different types of frames; Intracoded (I) frames, predictive (P) frames with a group of pictures (GOP) length of 30 frames with no B-frames. The frame losses are shown in Figure 33.

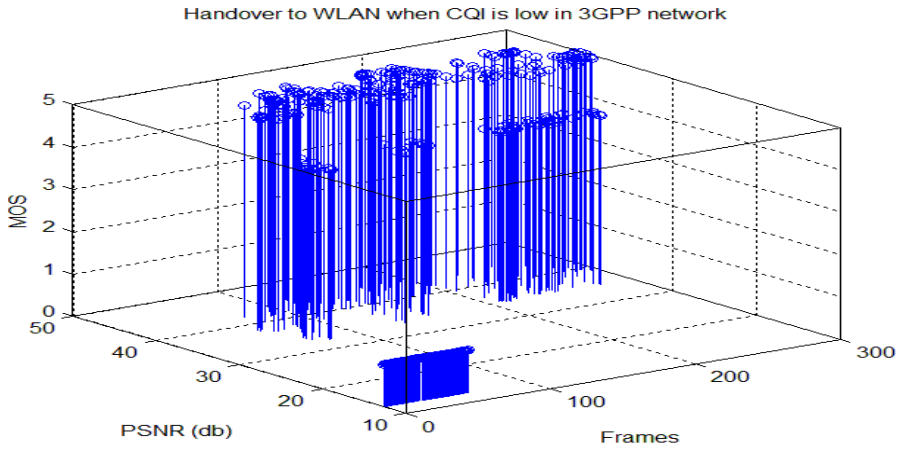


Figure 32. Handover to WLAN when CQI is low in 3GPP network

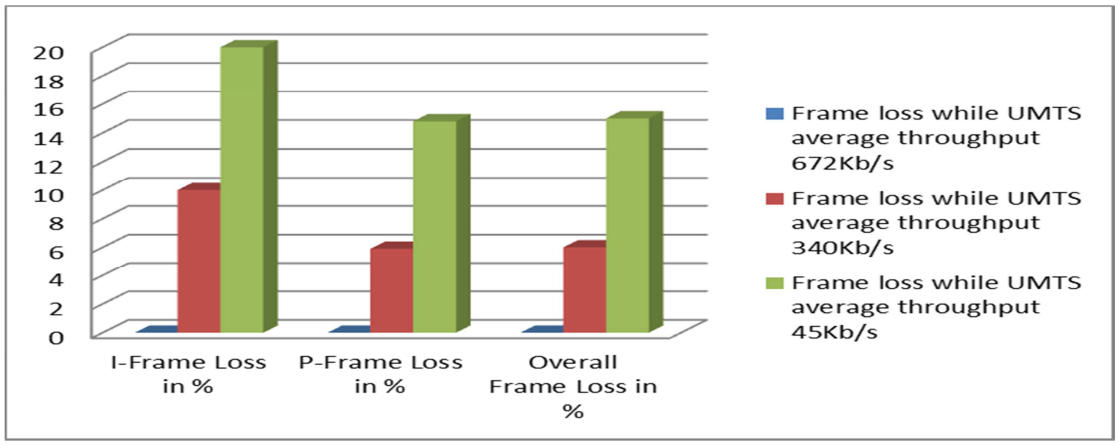


Figure 33. Impact of the frame losses due to UMTS network on the full transmission

3.1.3 Joint offloading and scheduling strategies for dual-mode small cells

The deployment of small cell base stations, SCBSs, overlaid on existing macrocellular systems is seen as a key solution for offloading traffic, optimizing coverage, and boosting the capacity of future cellular wireless systems. The next generation of SCBSs is envisioned to be multimode (i.e. capable of transmitting simultaneously on both licensed and unlicensed bands). This constitutes a cost-effective integration of both WiFi and cellular radio access technologies that can efficiently cope with peak wireless data traffic and heterogeneous quality of service requirements. To leverage the advantage of such multimode SCBSs, we discuss the novel proposed paradigm of cross-system learning by means of which SCBSs self-organize and autonomously steer their traffic flows across different RATs. Cross-system learning allows the SCBSs to leverage the advantage of both the WiFi and cellular worlds. For example, the SCBSs can offload delay-tolerant data traffic to WiFi, while simultaneously learning the probability distribution function of their transmission strategy over the licensed cellular band. The basic building blocks of cross-system learning are first introduced and then a preliminary performance evaluation in a Long-Term Evolution simulator overlaid with WiFi hotspots is provided. Remarkably, it is shown that the proposed cross-system learning approach significantly outperforms a number of benchmark traffic steering policies.

In the context of cellular and WiFi integration, the goal of every SCBS is to devise an intelligent and online learning mechanism to optimize its licensed spectrum transmission, and

at the same time leverage WiFi by offloading delay-tolerant traffic. The developed procedure, dubbed cross-system learning, is rooted in the fact that every small cell optimizes its long-term performance metric, as a function of its traffic load, interference levels, and users' heterogeneous traffic requirements. In addition, unlike standard reinforcement learning (RL), the cross-system learning procedure allows players to implicitly coordinate their transmissions with no information exchange, as well as to leverage the coupling between LTE and WiFi, which, as shown below, increases the overall network performance and significantly speeds up the convergence. The cross-system learning framework is composed of the following interrelated components. Subband selection, power level allocation, and cell range expansion bias: Every SCBS learns over time how to select appropriate subbands with their corresponding transmit power levels in both licensed and unlicensed spectra, in which delay-tolerant traffic is steered toward the unlicensed spectrum. In addition, every SCBS learns its optimal CRE bias to offload the macrocell traffic to smaller cells. Traffic-aware scheduling: Once the small cell acquires its subband, the scheduling decision is traffic-aware, taking into account users' heterogeneous QoS requirements (throughput, delay tolerance, and latency).

During cross-system learning, every SCBS minimizes over time its regret of selecting strategies yielding lower payoffs, while experimenting with other strategies to improve its long-term utility estimation. The considered behavioral assumption is that small cells are interested in choosing a probability distribution over their transmission strategies that minimizes the regret, where the regret of SCBS k for not having played action $q_k^{(l,s,b)}$ from $n = 1$ up to time t is defined as

$$r_{k,q_k^{(l,s,b)}}(t) = \frac{1}{t}\sum_{n=1}^{t} \hat{u}_k\left(q_k^{(l,s,b)}, p_{-k}(n)\right) - \hat{u}_k(n) \tag{62}$$

where $u_k(n)$ is the instantaneous utility observation (i.e. feedback) of SCBS k at time n, obtained by constantly changing its strategy. In addition, to calculate its regret, every SCBS k estimates its utility function $\hat{u}_k(.)$ when taking a given action based on local information. The rationale is as follows: If the regret is strictly positive, SCBS k would have obtained a higher average utility by playing action $q_k^{(l,s,b)}$ during all previous time instants; thus, SCBS k "regrets" not having done so. In contrast, if the regret is negative, SCBS k does not regret its strategy selection. Therefore, each SCBS needs to strike a balance between choosing actions that yield lower regrets (more often than those with higher regrets) and playing any of the other actions with a non-zero probability. The behavioral rule of every SCBS can be modeled by the following probability distribution subject to the maximum transmit power constraints where

$$\beta_k\left(\boldsymbol{r}_k^+(t)\right) \in \arg\min_{\pi_k}\left[\sum_{\boldsymbol{p}_k \in A_k} \pi_{k,\boldsymbol{p}_k} r_{k,\boldsymbol{p}_k}(t) + \frac{1}{\kappa_k} H(\pi_k)\right] \tag{63}$$

where $r_k^+ = \max\left(0, r_k(t)\right)$ denotes the vector of positive regrets, and $H(.)$ represents the Shannon entropy function of the mixed strategy p_k. The temperature parameter κ_k represents the interest of SCBS k in choosing other actions. The unique solution to the right side of the continuous and strictly convex optimization problem

$$\beta_{k,q_k^{(l,s,b)}}\left(\boldsymbol{r}_k^+(t)\right) = \frac{\exp\left(\kappa_k r^+_{k,q_k^{(l,s,b)}}(t)\right)}{\sum_{\boldsymbol{p}_k \in A_k} \exp\left(\kappa_k r^+_{k,\boldsymbol{p}_k}(t)\right)} \tag{64}$$

Furthermore, given users' different QoS requirements, the cross-system learning framework leverages WiFi, in which the learning process carried out over WiFi is faster (from a timescale perspective) than that on the cellular band. More concretely, inspired by the well-known turbo

principle, the output (i.e., feedback) from the WiFi learning process is used to update the cellular learning process. As shown later, this notion of timescale significantly reduces the convergence time of the traffic steering algorithm compared to standard RL, and improves the overall performance.

Once the SCBSs select their subbands using cross-system learning, they engage in a proactive and traffic-aware scheduling procedure on the selected subband's resource blocks. The scheduling algorithm is proactive and traffic-aware in nature as it incorporates users' traffic requirements. Notably, the scheduling decision is not only based on the instantaneous channel condition but also on the completion time (delay) and service class of each transmission. For that, within every small cell, all users are sorted in ascending order as a ratio of their remaining file size and estimated average data rate. Then SCBS k computes a metric $D_{k_i}(t)$, which is a function of the position of UE k_i and the number of UEs served by SCBS k at time t. Finally, UE k_i^* is scheduled such that:

$$k_i^* = \arg \min_{k_i} D_{k_i}(t) \tag{65}$$

This innovation corresponds to scenario 2.4.2 defined in [D.2.2]. Figure 34 plots the convergence behavior of the cross-system learning procedure in terms of the ergodic transmission rate (i.e., average cell throughput). Here, we consider 10 UEs per macrocell sector, and 1.4 MHz bandwidth in the licensed band. In addition, we plot the standard RL algorithm, in which learning is carried out independently over both licensed and unlicensed bands, without any sort of coordination. Quite remarkably, it is shown that the cross-system learning approach converges within less than 50 iterations, while the standard approach needs several hundreds of iterations to converge. Furthermore, the standard RL procedure exhibits an undesirable oscillating behavior (i.e. ping-pong effects between the licensed and unlicensed band, which can be detrimental in mobility scenarios). Finally, it is worth noting that convergence can be proven using tools from RL.

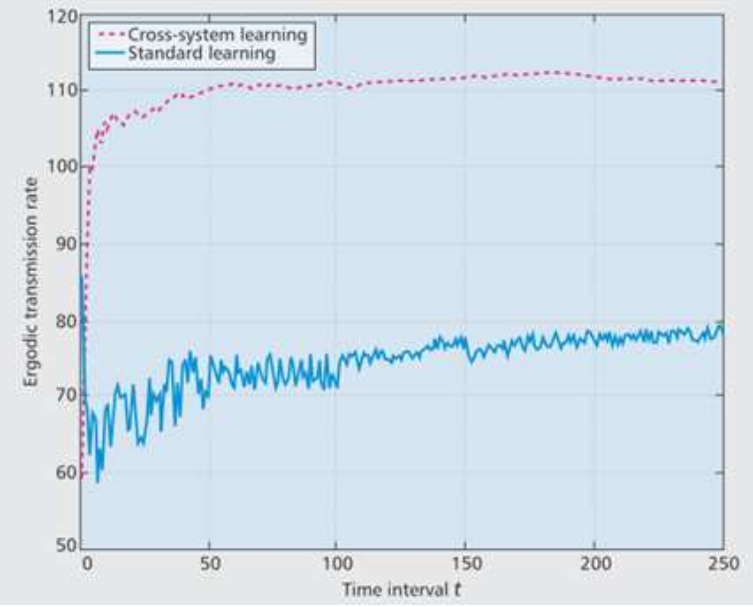


Figure 34. Convergence of the cross-system learning algorithm vs. the standard independent learning, K = 2 SCBSs per macrocell sector.

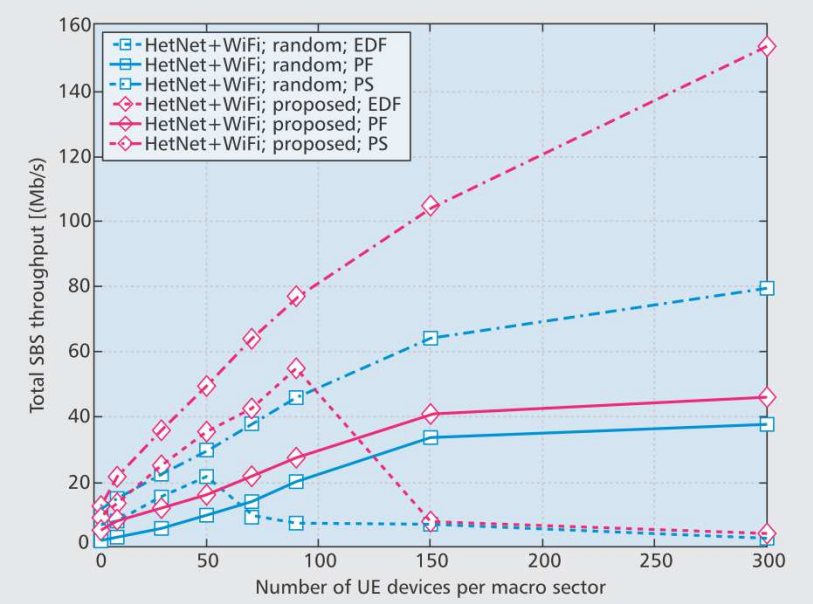


Figure 35. Total cell throughput vs. number of users for different traffic offloading and scheduling strategies, K = 2 SCBSs/macrocell sector.

Figure 35 shows the total cell throughput as a function of the number of UEs in the network, for the earliest deadline first (EDF), proportional fair (PF), and proactive scheduling (PS) strategies, respectively. While the standard PF scheduler cannot cope with the increasing number of UEs, the traffic-aware scheduling approach judiciously steers users' traffic in an intelligent and dynamic manner over both licensed and unlicensed spectrum, with a 160-fold increase for 300 UEs. These significant gains are rooted in the fact that unlike the proactive scheduler, both EDF and PF schedulers fall short of accounting for the heterogeneous traffic and delay-tolerant nature of their users.

4 ENERGY SAVING MECHANISMS

Currently deployed base stations need to continuously signal their presence and availability, and to listen to the radio environment to detect incoming users even during the low-traffic periods (e.g. nights). At the same time, the massive large scale deployment of small cells can be ecologically worthwhile only if armed with smart energy efficiency and power saving mechanisms. Motivated by this, energy-efficient sleep mode strategies for small cell base stations are required in order to reduce the power consumption of cellular networks. These strategies allow the hardware components in the base station to be smartly switched off in idle conditions, which means that the total cell, and further the total network energy consumption will become highly load-dependent. Furthermore, in case of deployments with a large number of overlaid cells, serving cell selection can be a one key mechanism to reduce the number of activated cells. Indeed, in this case, the classical approach where a UE is connected to the strongest Access Point (AP) may no longer be an optimal strategy in terms of interference and energy efficiency. Hence, there is a need to investigate new ways to select the serving cell for each UE. One solution lies in enabling a dynamic activation of local APs limiting the load of the worst (for example from the energy or interference point of view) APs. A possible improvement is to leverage on predictive RRM aspects, based on the estimation of traffic evolution, offering the required amount of capacity in both time and space.

The scope of SHARING Task 4.3 is to investigate the performance of various types of energy saving mechanisms both at the network and the base station hardware level. At the network level, the proposed solutions will be based on load balancing and on RRM reconfiguration. More specifically, the common factor is to propose and evaluate the performance of (self-optimizing) energy saving mechanisms based on the use of fast cell DTX (e.g. micro sleep), eNodeB sleep mode and/or dynamical cell switch-off. The investigations will consider both the QoS of the users and the overall network performance (e.g. capacity).

Regarding the base station hardware, the innovations will focus on enhancing the energy efficiency of the power amplifier (PA). For example, the studies will investigate how Peak to Average Power Ratio (PAPR) standards compliant methods can help to reduce the overall power consumption of the network. Furthermore, real-time tuning and optimization of RF modules will be investigated. Through external parameters such as global and local load demand, different power levels can be defined and radiated power can be adjusted. For different radiated power values, the PA can be in several working points with optimal. These optimal can even go to low power modes of operation and contribute to energy saving.

4.1 Energy saving via RRM reconfiguration

4.1.1 Utilizing eNodeB fast cell DTX and sleep mode to save energy

One of the problems related to the state-of-the-art eNodeBs is that they consume a considerable amount of power even when they are idle, i.e. not serving any users. As concluded by the EARTH project [EAR12], an efficient way to reduce this kind of fixed power consumption of idle eNodeBs is to apply some form of fast cell DTX, e.g. micro sleep or antenna muting. The main principle of micro sleep is to switch off the eNodeB power amplifier for the time periods when the eNodeB has nothing to transmit. In practise this means the time periods in between the cell-specific reference signals (CRS), broadcast (PBCH) and synchronization symbols (PSS/SSS), see Figure 36. The main advantage of micro sleep is that it can exploit the very short idle periods of the eNodeB. Another benefit of micro sleep is that the cell stays accessible, and hence, the overall system performance will not be affected.

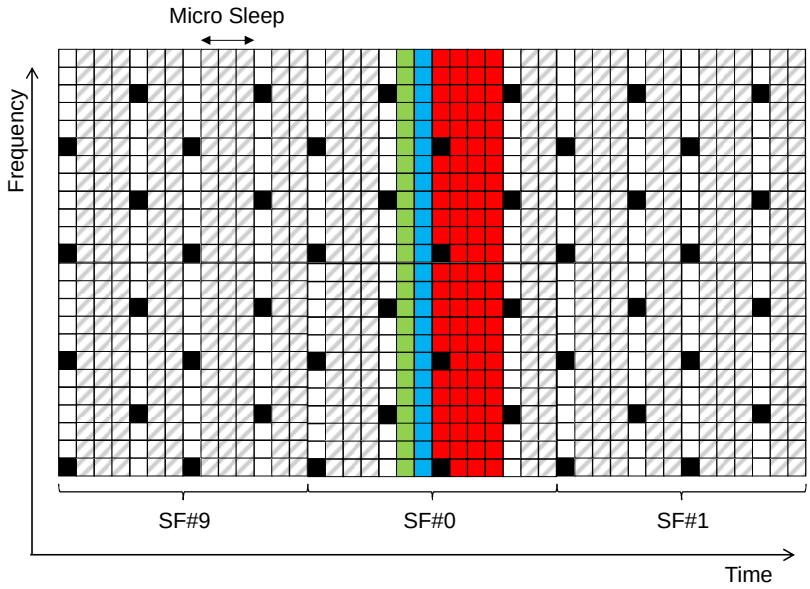


Figure 36. In case of micro sleep the eNodeB power amplifier is switched off in between the downlink control signals.

In case of antenna muting, all transmit antenna branches except of the first one are switched off to reduce the power consumption of an idle eNodeB, see Figure 37. The active antenna branch is needed to keep the cell accessible by transitting the required downlink control signals (CRS, PBCH and PSS/SSS). The muted antenna branches are then re-activated when needed to serve the traffic. A more detailed discussion of the antenna muting mechanism can be found for example in [EAR12] and [Ski12].

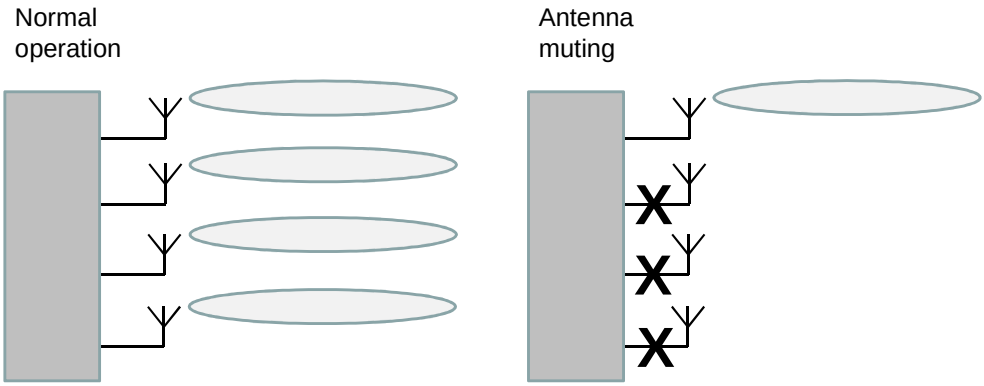


Figure 37. In case of antenna muting, all transmit antenna branches except the first one are switched off.

Since the fast cell DTX mechanisms do not affect the accessibility of a cell, they are applicable for all cells.

In addition to micro sleep, the pico cells can be subject to a sleep mode mechanism, where not only the power amplifier, but also the RF and BB components of the eNodeB are switched off. In case of the sleep mode the inactive periods are expected to be in the order of seconds. Since all the eNodeB transmissions are deactivated during the sleep mode, the cell will not be accessible to the users, and hence, the maximum achievable system performance will

potentially be affected. That is also the reason, why sleep mode is not applicable for (macro) cells, which are needed to provide the overall system coverage.

When it comes to the sleep mode, the main challenge is to maximize the sleep periods, without affecting the system performance too much. Hence, the eNodeB should be deactivated as soon as it is not needed to serve the offered traffic. Similarly, a sleeping eNodeB should be reactivated as soon as it is needed to serve close-by users.

Here, an eNodeB sleep mode mechanism is proposed, where the pico eNodeB is switched to sleep mode as soon as it has been idle for δ_{idle} seconds. In this initial scheme the decision to enter the sleep mode is made by the pico eNodeB itself, without the control of the neighboring macro eNodeB.

Two different schemes are proposed for the reactivation of sleeping pico eNodeBs, see Figure 38. In case of the first method (reactivation scheme A), the reactivation is based on the monitored macro cell load [Fal12]: if the load, averaged over δ_{macro} seconds, is higher than the predefined threshold (λ_{macro}), the macro eNodeB reactivates the sleeping eNodeBs, which are under its control. Suitable measures for the macro cell load can consist of for example the cell throughput or the average utilization of the cell resources. Reactivation scheme A assumes that the macro eNodeB is able to transmit the reactivation command to the desired pico eNodeBs. However, since there are no strict delay or bandwidth requirements for this message, it can be transmitted over the normal X2-interface.

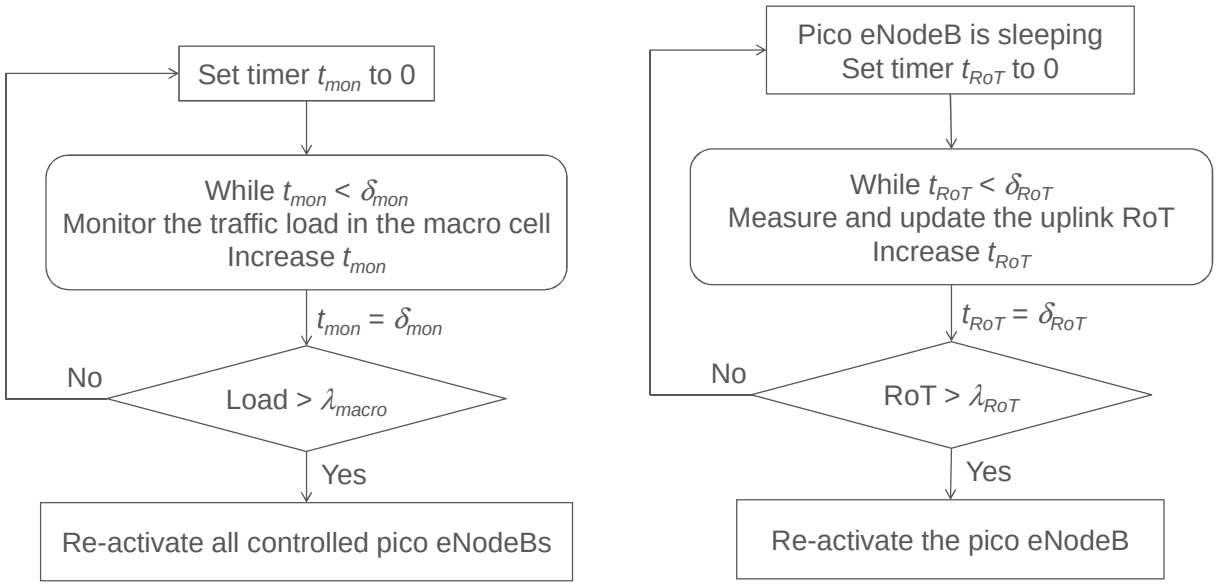


Figure 38. Two cell reactivation schemes are proposed: reactivation scheme A (left) and reactivation scheme B (right).

The second scheme (reactivation scheme B) is based on the uplink measurements performed by the sleeping eNodeB. Hence, in that case it is assumed that either the pico eNodeB includes a separate network listening module, capable of performing uplink measurements, or that the pico eNodeB uplink receiver is left active when entering the sleep mode. Now, the principle is that if the measured uplink power, averaged over δ_{RoT} seconds, is greater than the predefined threshold (λ_{RoT}), the sleeping pico eNodeB is reactivated. This also means that there is no need to transmit any additional control signalling between the eNodeBs.

The performance evaluation of the proposed energy saving mechanism is based on Scenario 2.3.3 described in [D.2.2]. Hence, the performance of a heterogeneous LTE network deployment consisting of 21 macro and 66 outdoor pico cells is simulated over a period of 24 hours. During that period both the average hourly area traffic volume and the geographical

traffic distribution is assumed to vary. For example, the lowest area traffic volume, 28 Mbps/km^2, is experienced during the early morning hours, while the peak traffic equal to 200 Mbps/km^2 is experienced around 09pm-10pm. Similarly, during the office hours most of the traffic is concentrated within the commercial areas, while during evenings, nights and morning hours most of the traffic is generated within the residential areas.

A dynamical system-level simulator is utilized, and for each of the 24 hours the system is simulated for 10 minutes. The output of the system simulation consists of the observed user performance and the total network power consumption measured as kW/km^2.

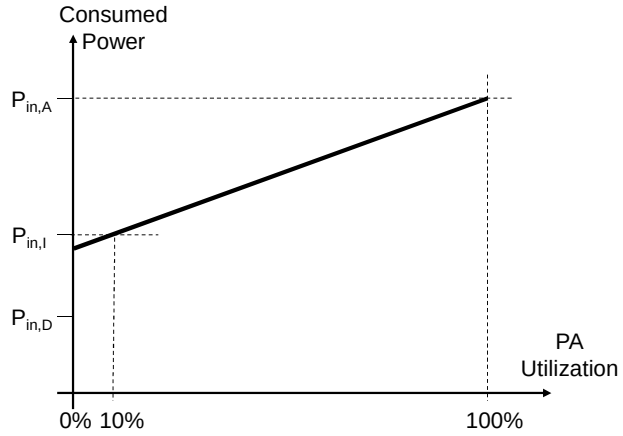


Figure 39. Assumed model for the eNodeB power consumption.

The impact of fast cell DTX is modelled by defining different eNodeB power consumption values for the different activity states, see Figure 39. In general, an eNodeB is assumed to be either in active state or in idle state. In case of an active state, i.e. when the eNodeB is serving at least one user, the eNodeB is transmitting at the maximum output power and the power consumption is equal to $P_{in,A}$. In case of an idle state, the eNodeB is transmitting only the downlink control signals, and the average power consumption is assumed to be equal to $P_{in,I}$ or $P_{in,D}$ depending on whether fast cell DTX is applied or not. The assumed P_{in} values have been derived from [3GPP11a] and are listed in Table 6. Yet another assumption is that the power consumption during the sleep mode ($P_{in,S}$) is equal to 10 W.

Table 6. Assumed eNodeB power consumption values (per cell).

Type of eNodeB	Maximum Output Power	$P_{in,A}$	$P_{in,I}$	$P_{in,D}$	$P_{in,S}$
Macro	40 W	280 W	174 W	144 W	N/A
Pico	1 W	97.7 W	96.3 W	78.1 W	10 W

The aggressiveness of the sleep mode algorithm can be adjusted via the timer (δ_{idle}) and the load thresholds ($\lambda_{macro}$and λ_{RoT}). For example, the algorithm can be made more aggressive by reducing the value of δ_{idle}. Similarly, the algorithm becomes more aggressive if higher values of λ_{macro} or λ_{RoT} are applied. By increasing the aggressiveness of the sleep mode algorithm,

the total network power consumption can be reduced. However, at the same time the risk for an unacceptable system performance will increase.

Results for the observed cell-edge user performance as a function of time are presented in Figure 40, assuming the reactivation scheme A. Four curves are shown, one for a scenario where all the pico cells have been completely switched off, and another one for the assumed heterogeneous network scenario without any sleep mode algorithm. The additional two curves assume different sleep mode algorithm setups: aggressive sleep mode ($\delta_{idle} = 1$ second and $\lambda_{macro} = 10$ Mbps) and guaranteed performance ($\delta_{idle} = 4$ seconds and $\lambda_{macro} = 6$ Mbps). If the minimum acceptable cell-edge throughput is equal to 10 Mbps, the aggressive sleep mode algorithm is able to fulfill it in between 01am and 11am, whereas during the rest of the day the observed cell-edge throughput is less than 10 Mbps. One can also see that with a more relaxed algorithm setup, the observed cell-edge user throughput can be kept above the desired threshold throughout the day.

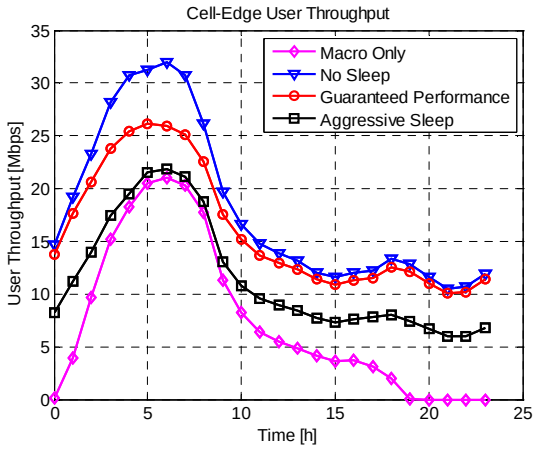


Figure 40. Observed cell-edge user throughput for the different sleep mode algorithm setups.

Results for the total network power consumption as a function of time are presented in Figure 41. Now, five different curves are shown: macro only, without sleep mode, aggressive sleep mode, guaranteed performance and ideal sleep mode. The curve for the ideal sleep mode assumes a scheme, where all idle capacity cells are automatically assumed to be in sleep mode [Hil13]. Furthermore, the figure on the left presents the results without fast cell DTX, while the results on the right present the results with fast cell DTX.

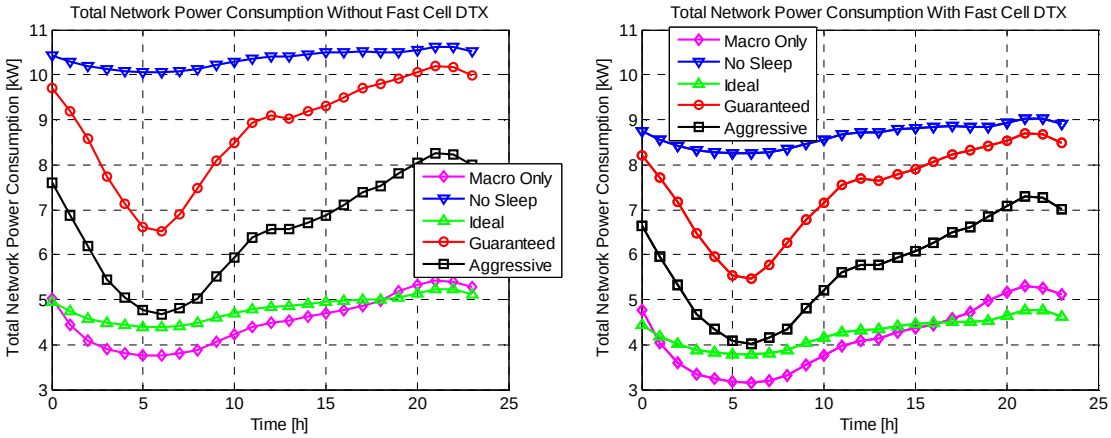


Figure 41. Total network power consumption as a function of time without fast cell DTX (left) and with fast cell DTX (right).

As demonstrated by the results for the low-traffic hours, the aggressive sleep mode algorithm is able to reduce the total network power consumption by more than 50%. If the guaranteed performance is assumed instead, the total network power consumption can be reduced by 35%. One can also notice that the gain of fast cell DTX alone is equal to 18%. Hence, the utilization of sleep mode can provide larger energy saving gains compared to the utilization of fast cell DTX. Finally, a combination of fast cell DTX and aggressive sleep mode results in a 60% lower power consumption, and the combination of fast cell DTX and sleep mode with guaranteed user performance results in a 46% lower power consumption compared to a deployment without any energy saving mechanisms.

The inefficiency of the evaluated sleep mode algorithm becomes visible when looking at the results for the peak traffic hour. Now, the guaranteed performance alternative can offer only a 4% reduction in the total network power consumption. The aggressive sleep mode algorithm reduces the power consumption considerably more, but results at the same time in an unacceptable user performance. One can also notice that neither of the two sleep mode algorithm setups are even close to the performance of the ideal sleep mode. Yet another thing to note is that the gain of fast cell DTX alone is equal to 15%, while the gain of combined fast cell DTX and sleep mode is equal to 18%. Hence, during the peak traffic hours it is more efficient to utilize fast energy saving actions compared to sleep mode.

Based on the initial evaluation results, pico eNodeB reactivation scheme B utilizing uplink measurements shows very similar performance to the re-activation scheme A described above.

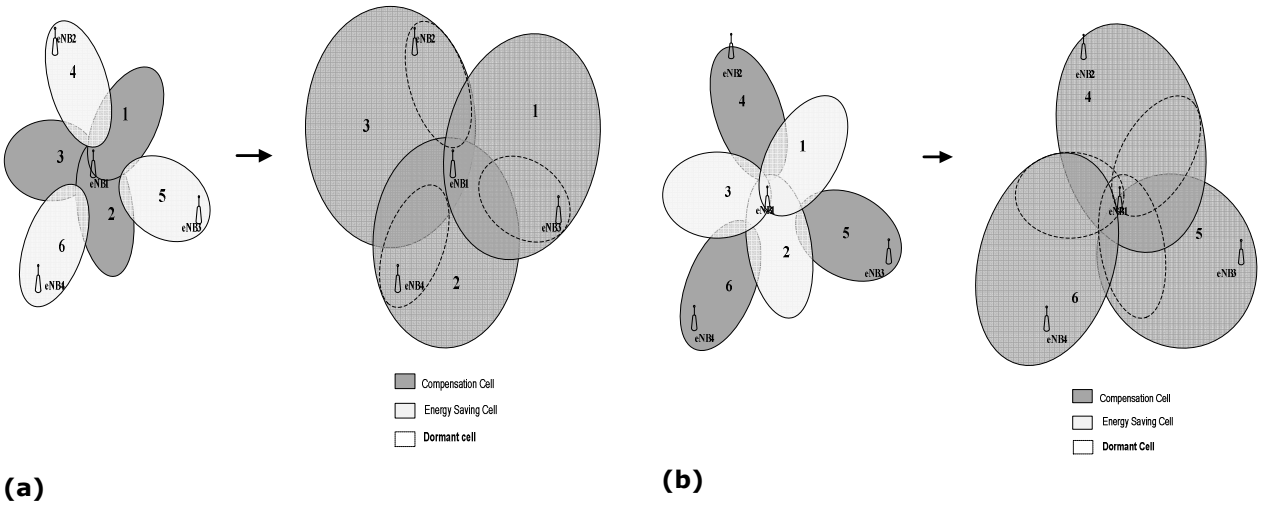
The reason why the evaluated sleep mode algorithms are so inefficient during the peak traffic hours is that even though the pico cells are idle in 95% of the time, the idle periods are so short that the sleep mode algorithms are not able to utilize them. At the same time the life-time of a user is very short due to the assumed packet transmission traffic model. In practice this means that when the macro cell is serving multiple simultaneously active users, the cell load exceeds the pre-defined threshold and the pico cell reactivation is triggered. However, at the time when the pico cells have been reactivated, the users have already finished their packet transmissions, and have left the system.

4.1.2 Centralized techniques for ON/OFF energy saving in HetNet campus scenario

Compensation based ON/OFF energy saving is an important energy saving technique that is currently discussed in 3GPP normalization [3GPP887]. In compensation based energy saving, a compensation base station set is determined by network planning such as the base stations not allowed to be turned OFF are in the compensation set. Two compensation based energy saving strategies are foreseen in the current status of the standardization discussions as a means for energy saving in interference limited scenarios discussed in [3GPP887].

- Single compensating base station energy saving: where one base station compensates for the coverage of neighbouring base stations.
- Multiple compensating base stations energy saving: where multiple base stations compensate for the coverage of neighbouring base stations.

Figure 42 illustrates the single base station compensation and multiple base stations compensation use cases.



(a) **(b)**

Figure 42. Compensation based energy saving [3GPP887] (a) single compensating eNB, (b) multiple compensating eNBs.

When triggering the system state into energy saving state, the compensating base stations are transmitting on the maximum power and the non-compensating (energy saving base stations) are set OFF. So, the determination of the compensating base stations is an important task when considering such energy saving mechanisms. Currently two level network planning is foreseen as a means for determining the compensating base station set. The first network planning level is considered when the network is not in energy saving, then a lower target capacity is set for another network planning of the network when it is in energy saving state. This network planning effort may be relevant for homogeneous macro base stations network but is clearly irrelevant for heterogeneous deployment and especially for clusters of randomly deployed heterogeneous networks that are used to cover shopping mall/campus areas (campus scenario). This scenario, presented in [D.2.2] as scenarios 2.2.2 and 2.3.10 will be the benchmark scenario for the techniques we propose in this study for the determination of the compensating base stations and the associated evaluations. In this study we propose methods for finding the compensating base station set for cluster of base stations randomly deployed in a coverage region *C* and connected to a central node by using the graph representation of its coverage as shown in the Figure 43 (a) for the campus deployment and (b) for the graph representation of the campus.

The graph representationof the campus deployment is defined as follows: the nodes of the graph represent the nodes of the campus (eNB, HeNBs etc). The edges of the graph representation are present if the neighbouring base station is reported with sufficiently high signal to interference and noise ratio (above a predetermined threshold) to the current, attachement base station of a given active user terminal UE in the scenario. For example two nodes i and j of the graph representation are connected by an edge if the signal to interference and noise ratio (SINR) of at least one UE attached to the base station i, of base station j is above -10dB and the SINR reported by at least one UE attached to the base station j of the base station i is above -10dB. Another possible method for the graph construction is to link the two nodes i and j if the downlink interference level seen in the two base stations is above a predetermined threshold. Intuitively, this graph representation is used to find the compensation set as the set of nodes that is a neighborto every node in the graph representation, so the SINR of the UEs dropped in the coverage region of the campus are optimized even when the energy saving base stations are set OFF. So, the compensation base station set is a dominant set in the graph representation of the campus.

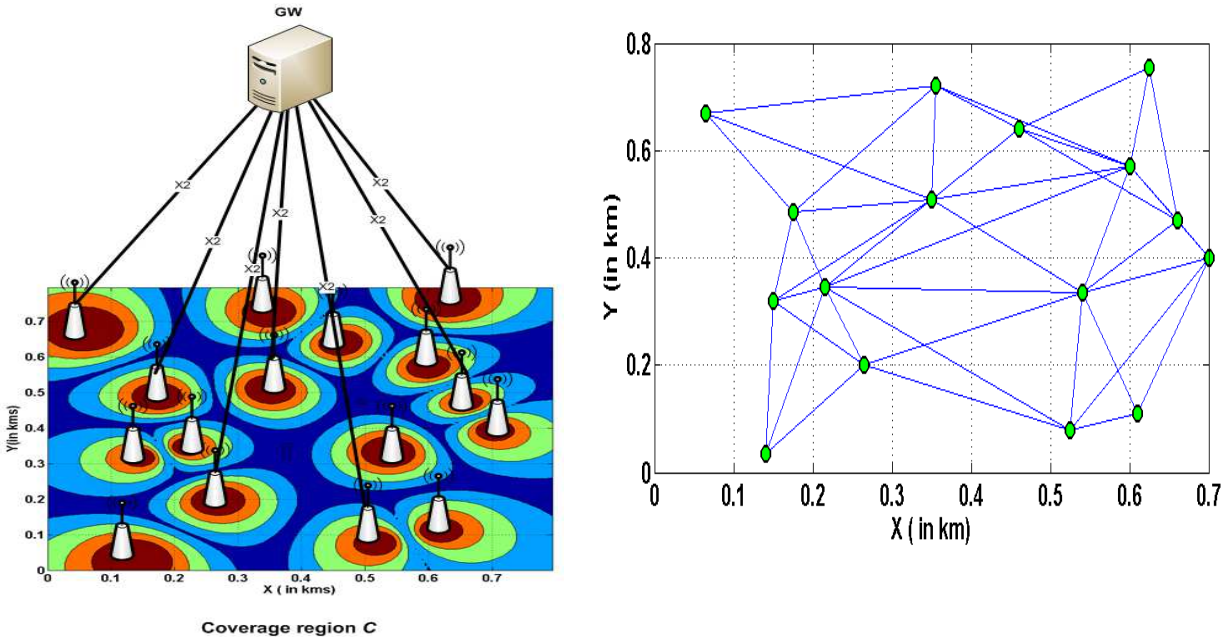


Figure 43. Campus deployment scenario and coverage map (a) and corresponding graph representation (b)

The dominant/dominating set may be either a set of neighbouring nodes in the graph or a set of independent, nonneighbouring nodes but such that every node in the graph is reachable from the dominating set nodes. This leads to the following dominant set classes:

1. Minimum connected dominant set (CDS): where the nodes of the dominant set are allowed to be connected by an edge in the graph.
2. Maximum independent set (MIS): where the nodes of the dominant set are not connected in the graph of the campus and the coverage of the dominant set is maximized.

In the Figure 44, examples of the two dominant sets are shown for the example of the 16 nodes campus of the Figure 43: the blue nodes are CDS compensation nodes and red nodes are MIS comepnsating nodes. CDS and MIS based compensation sets corresponds roughlyto the single and multiple compensating base stations sets.

Initial performance evaluation of the compensation based energy saving is shown to improve the overall performance of active UEs in the campus coverage region and reduce the radiated power of the base stations in the campus.

Figure 45 illustrates the minimum spectral efficiency and the mean spectral efficiency for the 16 base nodes campus scenario. We have assumed that the user terminals are sampled from spatial poisson point process (SPPP) with an intensity of 80 users. This spatial traffic intensity is corresponding to an average of 10 users per base station. The radio parameters of the base stations of the simulation scenario corresponds to the 3GPP pico base stations radio parameters as described in [3GPP10a] and the performance is averaged over 1000 base stations positions.

The average radiated power gain of the compensation technique is around 75% which corresponds to an average number of 4 base stations in the compensation sets. The simulation results were done for a homogeneous campus scenario (only pico base stations are considered). In future studies we will further evaluate the performance of the compensation techniques for a heterogeneous campus scenario (pico base stations and macro eNBs). Another important evaluation will be to translate the radiated power gain into energy saving gain by using the power models provided by the EARTH project [EAR12].

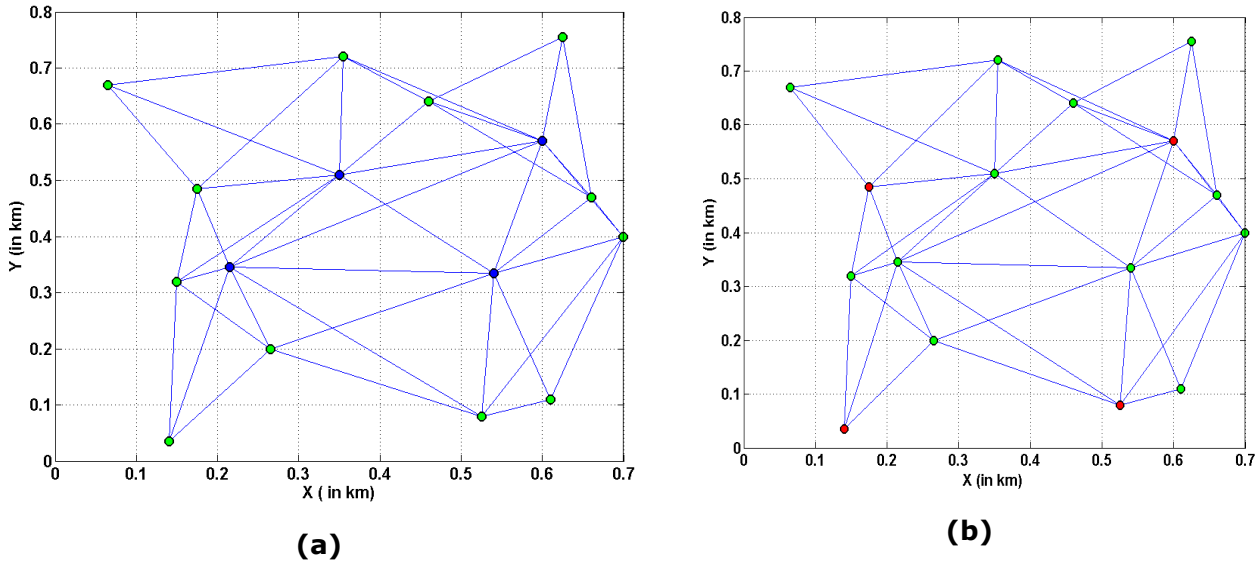


Figure 44. Dominant sets based compensating base stations (a) connected dominating set compensation (b) maximum independent sets compensation.

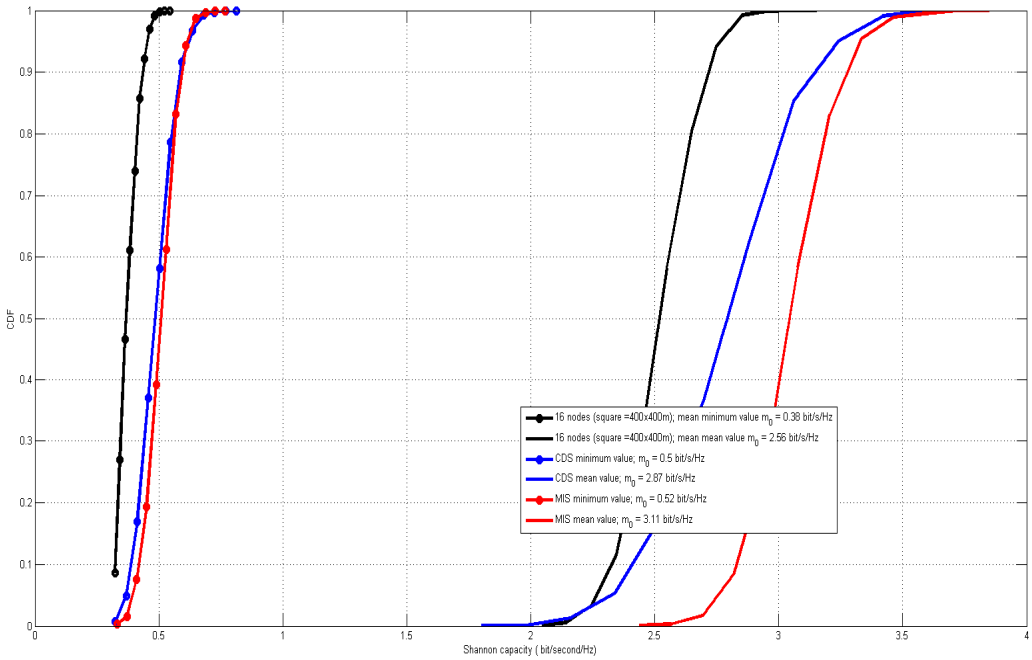


Figure 45. Initial performance evaluation of the dominating sets based compensation.

4.1.3 Decentralized techniques for base station power setting in HetNet campus scenario

In this study we will focus on downlink power setting for sum radiated power minimization in the case of heterogeneous network campus scenario. We will discuss in this section the initial thoughts already provided in [KG11] where group of N base stations is deployed in the campus coverage region. The radiated sum power minimization may be formulated as the following classical optimization problem:

$$\min \sum_{i=1}^{N} P_i$$
$$st \ \frac{\alpha_{ii} P_i}{\sum_{j \neq i} \alpha_{ij} P_j + N_0} = \gamma(i); \ \ i = 1, \dots, N \tag{66}$$

This optimization problem is the problem of minimizing the sum of the radiated powers of the base stations such as the cell edge signal to interference ratio of each base station is above a given quality threshold $\gamma(i)$ that is set by operation and maintenance center (OMC). In [KG11] we propose to solve this optimization problem through well known log barrier interior point method where the optimization problem is reformulated as the following minimization problem, parametrized by the common offset to the constraints variable t.

$$\min_{\underline{P}} \sum_{i=1}^{N} \left[P_i - \frac{1}{t} \log \left(P_i - \frac{\gamma(i)}{\alpha_{ii}} \left(\sum_{j \neq i} \alpha_{ij} P_j + N_0 \right) \right) \right] \tag{67}$$

where the parameters α_{ii} and α_{ij} are respectively the path gains from the base station i to the cell edge UE of the base station i and the path gain from the base station j to the cell edge UE of the base station i. This centralized optimization can be solved through the central path technique:

1. Starting with initial low value of the parameter t, the central path technique solves the parametrized optimization problem, i.e. find the power vector $\underline{P}^*$ that is solution to the optimization problem above as function of parameter t.
2. Increase the value of the parameter t and go to step (1)
3. The algorithm stops if the precision N/t is below a predifined precision.

In [KG11] we have developed a decentralized algorithm for solving the log barrier interior point optimization problem with the following main ingredients:

1. Each base station evaluates its coordination parameter $t(i)$ and broadcasts it to the neighbouring base stations in the campus.
2. The base stations of the campus evaluate through consensus averaging step the average of the coordination parameters in the campus and this value is stored in each base station as t.
3. Each base station of the campus sets autonomously it's downlink radiated power such as to achieve the reference SINR $\gamma(i)$ up to an SINR margin defined by a power parameter $\lambda(i)$ that is a function of the average coordination parameter of the network t.
4. Each base station tests the value of the average coordination parameter of the network and the algorithm stops if the precision N/t is below a predefined precision. Otherwise, the algorithm goes to step (1).

The increase of the parameter t may be triggered by any base station of the campus that have enough power margin to achieve the constraint and the initial value of the coordination parameter is obtained when all the base stations are transmitting at the maximum power.

It was shown in [KG11] that the proposed technique doubles the energy efficiency for fairly large randomly deployed campus and shows the additional feature of being scalable and resistant to topology variations in the campus.

We will extend and refine this technique in future work within SHARING WP4 to HetNET campuses with macro base station interference and provide additional evaluations of the performance of the technique for the dual uplink power saving optimization problem. In this

case the distribution of the uplink interference will be taken into account in the setting and the adaptation of the coordination parameter t.

This innovation corresponds to scenario 2.3.10 defined in [D.2.2].

4.1.4 Dynamic cell ON/OFF power saving

A careful view of typical daily traffic data models reveals that due to different reasons long periods appear where the total load decreases and thus the network infrastructure can be considered as oversized.

Developing methods which can help achieving notable energy savings in cellular network infrastructure are of interest. One of them consists of studying a network management approach where different ON-OFF schemes are applied for providing considerable energy savings. The idea behind all different ON-OFF schemes is to consider different network layouts (as shown in Figure 46). Depending on the scenario considered, in case the low traffic periods are longer or deeper, bigger cell groups can be considered, with higher gains. In those cases, it is also possible to combine several schemes, for instance applying 1/4 (switching off 3 out of 4 cells) schemes at the beginning, and the 1/7 or even 1/9 afterwards.

The method of operation is to select certain base stations to shut down in periods of low load, dynamically reducing the number of active base stations so that the network has fewer nodes in use at any time. The service carried out by these switched-off nodes is replaced by the base stations that remain active, considered able to adjust and set their power dynamically so as to increase their range according to the requirements of the network. It is also appropriate to note that the primary objective being to achieve energy savings, this will be achieved without compromising the Quality of Service (QoS) expected by the user.

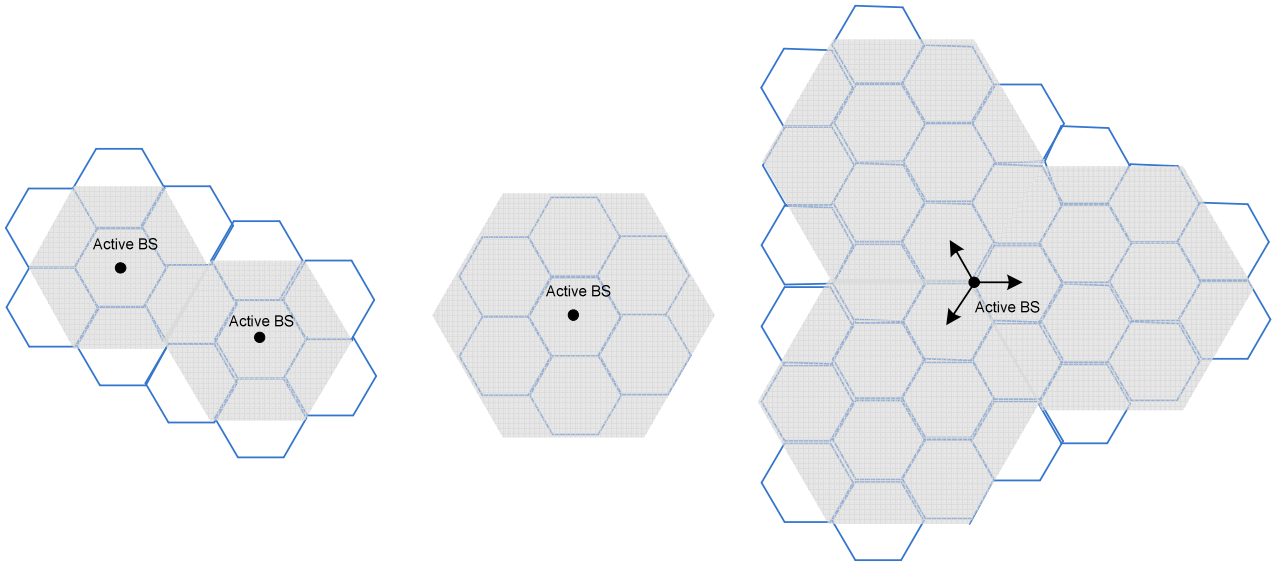


Figure 46. Typical real world cellular network schemes

There are certain principles that need to be applied to this study, starting from switching off base stations only when the load conditions are such that the remaining nodes can serve user's demands, thus new load requirements are never bigger than the network capability. In addition, viability relates to a dense base station deployment, done to cope with peak-traffic. The developed management mechanism should be easy to apply in base stations where the limiting factor is capacity and not the power budget. In this way, the required transmit power increase of a few watts increase would appear negligible compared to the gain of switching off

base stations. And finally the designed strategy must consider a set of 'a priori' defined number of BSs put into idle mode, based on load curves and previous experiences.

This study should not end in a theoretical plane but be translated to a real cellular network deployment instead, where it's not possible to switch off any random number of base stations, but it's restricted by their configuration. Thus, the work to be carried out will employ selected typical schemes for geometry and site positioning, knowing that on a common network configuration like the hexagonal layout the number of cells that can be switched off respond to 1-out-of-4 or 1-out-of-7 proportion when using omni-directional antennas, and to a 1-out-of-9 scheme when employing tri-sectorial ones, as reflected in Figure 46.

Different models for the daily traffic load (L) in cellular networks have been considered. First of all from a static perspective, where just a layout is applied (for instance, just 1/2 or 1/3). By $1/X$ notation, the intention is to express that just 1 out of X base stations from one cluster remains active when activating the scheme (due to the detection of low traffic). Then, dynamic combinations of schemes are considered. This extends the time where savings are available, making them in addition higher. The first option implies using just one scheme all 24 hours, while dynamic models allow several transitions depending on the instantaneous load.

The performance of the proposed innovation has been evaluated for scenario 2.3.4 described in [D.2.2].

Figure 47 (left hand side) tries to evaluate the differences between these two strategies, using the simplified model on the right plot. Static 1/2 and 1/4 schemes are compared against a dynamic combination of both. This model can be split into three states, corresponding to peak traffic, low demanding period and transitions from both states (denoted as slope-d linear transitions). Depending on the value of this d parameter, the results varysignificantly. For $d = 2/T$ (T being the time period corresponding to one day), the lowest savings are obtained (more or less 30%). On the other hand, using $d = 0.5$, results on top energy savings (up to nearly 70%).

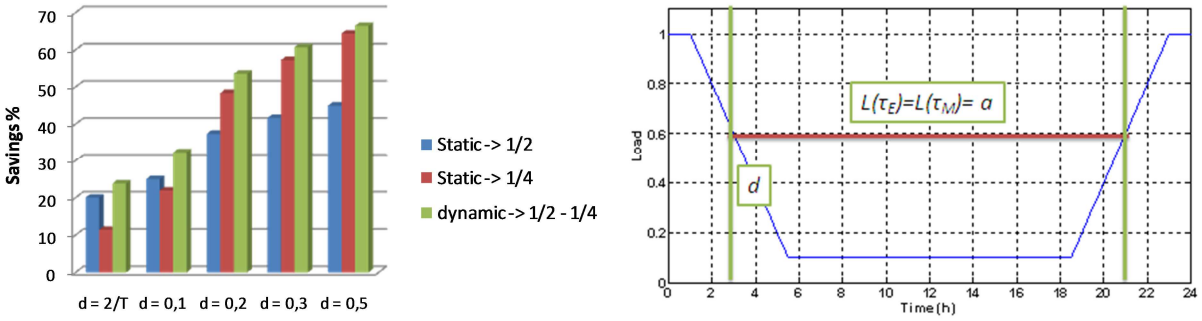


Figure 47. Comparison between static and dynamic ON-OFF schemes in case of simplified linear traffic models.

This is just a first step, as this linear simplified assumption cannot be extrapolated to the real world. But it acts as an upper bound for real results envisaged with the use of real network data. Figure 48 includes the same study as in the previous plot but using just static ON/OFF models and realistic traffic curves extracted from the literature (see Figure 49). Uniform omnicell deployments are considered. Using this next step, the savings limit seems to be around at least 25%.

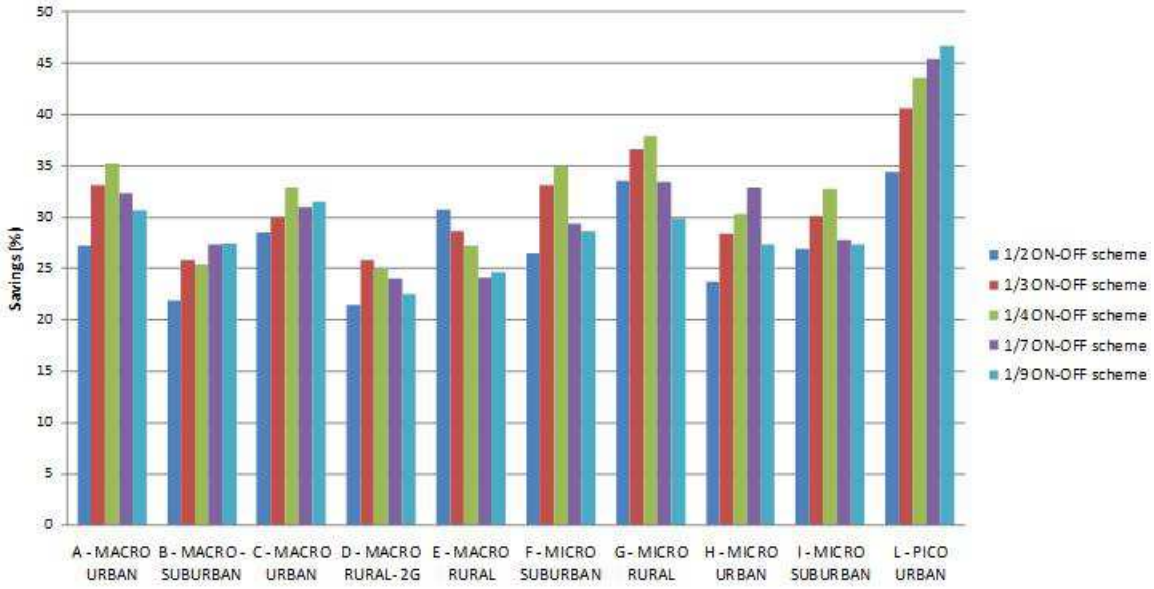


Figure 48. ON-OFF schemes applied in omnicell deployment scenario.

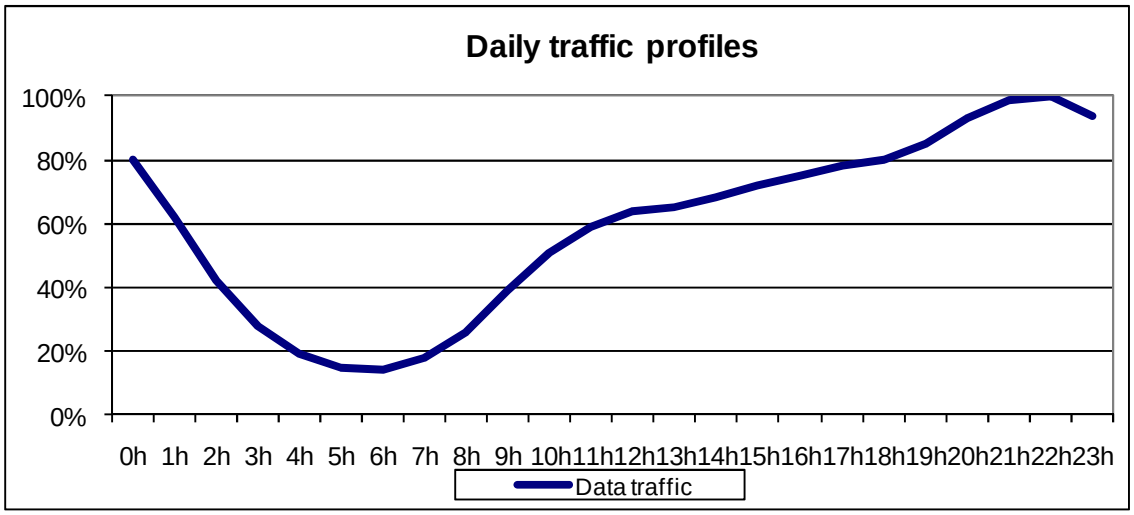


Figure 49. Daily averaged traffic curves extracted from EARTH project [EAR12].

But, when consideringthe real world deployment in each cell, the traffic variation is more bursty, with peaks and valleys happening during the whole day, so a good interpretation of the future trend is critical to determine the best option. Figure 50 shows two potential time slots to activate an ON/OFF scheme based on the fact that the load is under a given threshold. In the right hand side slot, applying the scheme is a good idea, as there is plenty of time to take advantage of savings. On the other hand, the first slot on the left hand side may be not enough to apply a scheme, as it must be canceled soon and the process may consume more energy in the transition than the energy needed to maintaining the cell as it is. This will be taken into account in future steps of the study, as well as more advanced dynamic schemes to be applied to clusters of cells with centralized and distributed management.

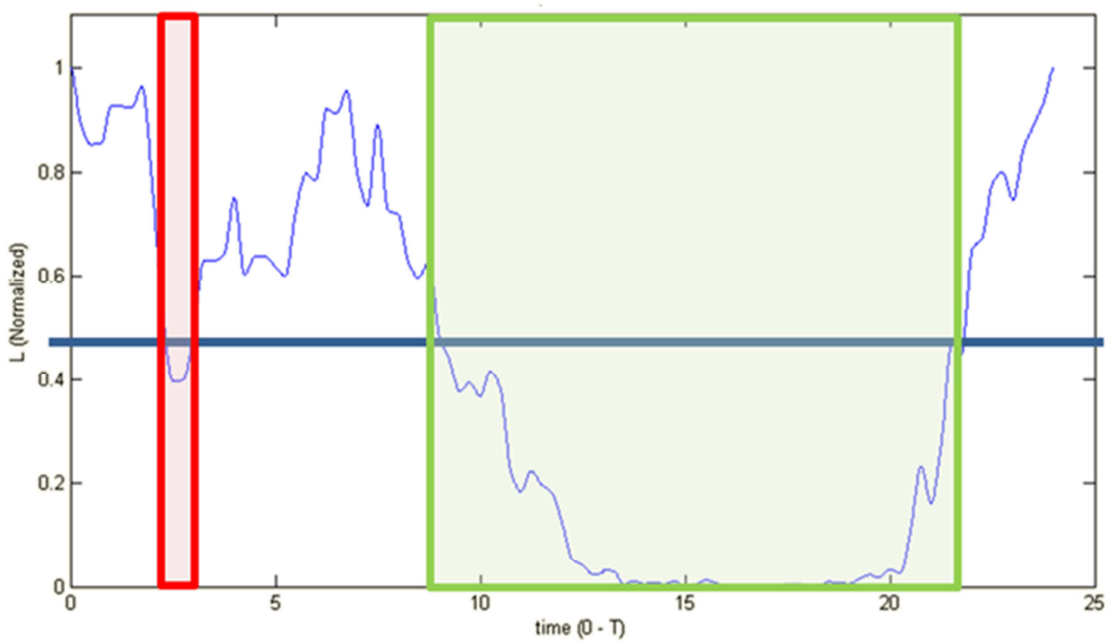


Figure 50. Time slots for ½ ON-OFF schemes

4.1.5 Opportunistic sleep mode strategies for small cells

The design of energy-efficient mechanisms is one of the key challenges in emerging wireless small cell networks. In this subsection, a novel approach for opportunistically switching base stations on and off to improve the energy efficiency in wireless small cell networks is proposed. The proposed approach enables the small cell base stations to optimize their downlink performance while balancing the load among each other and satisfying the users' quality-of-service requirements. The problem is formulated as a non-cooperative game among the base stations that seek to minimize a cost function which captures the tradeoff between energy expenditure and load. To solve this game, a distributed learning algorithm is proposed where the base stations autonomously choose their optimal transmission strategies.

Consider the downlink transmission of a heterogeneous wireless network with a set of base stations $\boldsymbol{B} = \{1, .., B\}$. The set $\boldsymbol{B}$ consists of a set of small cell base stations (SBSs) $\boldsymbol{B_S} = \{1, \dots, B_S\}$ underlaid on a macro cellular network with a set of macro base stations (MBSs) $\boldsymbol{B_M} = \{1, .., B_M\}$. Without loss of generality, we assume that the MBS 1 is located at the origin of the two-dimensional network layout and we let x be any location on the plane measured with respect to the origin. Moreover, let L_b be the coverage area of base station b such that any given user equipment (UE) at a given location x is served by base station b if $x \in L_b$. Let S_b be the transmission indicator of base station b such that $S_b = 1$ indicates the active state while $S_b = 0$ reflects the idle or sleep state. Here, we assume that, in active state, each base station will serve all UEs in its coverage area. From an energy saving perspective, some base stations might have an incentive to switch into sleep mode. Note that during the idle state, a base station consumes a non-zero power to sense the UEs in its vicinity. Consider that base station b uses a cell range expansion bias (CREB) ζ_b to absorb additional UEs (expand its coverage area L_b) along the transmission power P_b. The concept of CREB has been proposed in small cell networks due to the disparate cell sizes between different base stations. Since the CREB concept is used only by SBSs, we let $\zeta_b = 0$. Moreover, we assume that all base stations transmit on the same frequency spectrum (i.e. co-channel deployment). Therefore, the received signal to interference and noise ratio (SINR) from base station b at location $x \in L_b$ is given by:

$$\gamma_b(\boldsymbol{x}) = \frac{P_b S_b h_b(\boldsymbol{x})}{\sum_{\forall b' \in B/b} P_{b'} S_{b'} h_{b'}(\boldsymbol{x}) + N_0} \tag{68}$$

Further, the data rate at a given location x from base station b is given by:

$$R_b(\boldsymbol{x}) = \omega \log_2\big(1 + \gamma_b(\boldsymbol{x})\big) \tag{69}$$

We assume that the UEs connected to base station b are heterogeneous in nature such that each UE has a different QoS requirement based on its individual packet arrival rate. In this respect, let $\lambda(x)$ and $1/\mu(x)$ be the packet arrival rate and the mean packet size of any UE at location $x \in L_b$. The data rate offered to the UE at location x from base station b is $R_b(x)$ and thus, the load density of base station b becomes

$$\varrho_b(\boldsymbol{x}) = \frac{\lambda_b(\boldsymbol{x})}{\mu_b(\boldsymbol{x}) R_b(\boldsymbol{x})} \tag{70}$$

Consequently, the base station load ρ_b of base station b is given by:

$$\rho_b = \int_{\boldsymbol{x} \in L_b} \varrho_b(\boldsymbol{x}) d\boldsymbol{x} \tag{71}$$

Here, for each base station $b \in B$, we define a cost function that captures both energy consumption and load, as follows:

$$\begin{aligned} &\underset{\rho}{\text{minimize}} \sum_{\forall b \in B} \Gamma_b(\boldsymbol{\rho}) \\ &\text{subject to} \quad 0 \le \rho_b \le 1, \ \ \forall b \in B \\ &\qquad\qquad P_b^{Total}(\boldsymbol{\rho}) \le P_b^{Max}, \ \ \forall b \in B \end{aligned} \tag{72}$$

At time instant t, each base station b advertises its estimated load $\rho_b(t)$ via a broadcast control message along with its transmission power $P_b(t)$ and CREB $\zeta_b(t)$. Considering both the received signal strength and load, at time t the UE at location x connects to base station $b(x,t)$, $x \in L_b(x,t)$, according to the following UE association rule:

$$b(\boldsymbol{x}, t) = \arg\max_{b \in B} \big\{ (\hat{\rho}_b(t) + \epsilon_b)^{-\delta} P_b^{Rx}(t) \big\} \tag{73}$$

Due to fact that the base stations need to estimate their loads beforehand, the estimations must accurately reflect the actual load. In order to obtain an accurate estimation for the load of the base station b, we compute the load estimation $\hat{\rho}_b(t)$ at time t based on history as follows

$$\hat{\rho}_b(t) = \hat{\rho}_b(t-1) + \nu(t)\big(\rho_b(t-1) - \hat{\rho}_b(t-1)\big) \tag{74}$$

where $\nu(t)$ is the learning rate of the load estimation. Leveraging different time-scales, we assume that $\nu(t)$ is selected such that the load estimation procedure is much slower than the UE association process.

The innovation presented above is related to scenario 2.3.8 defined in [D.2.2].

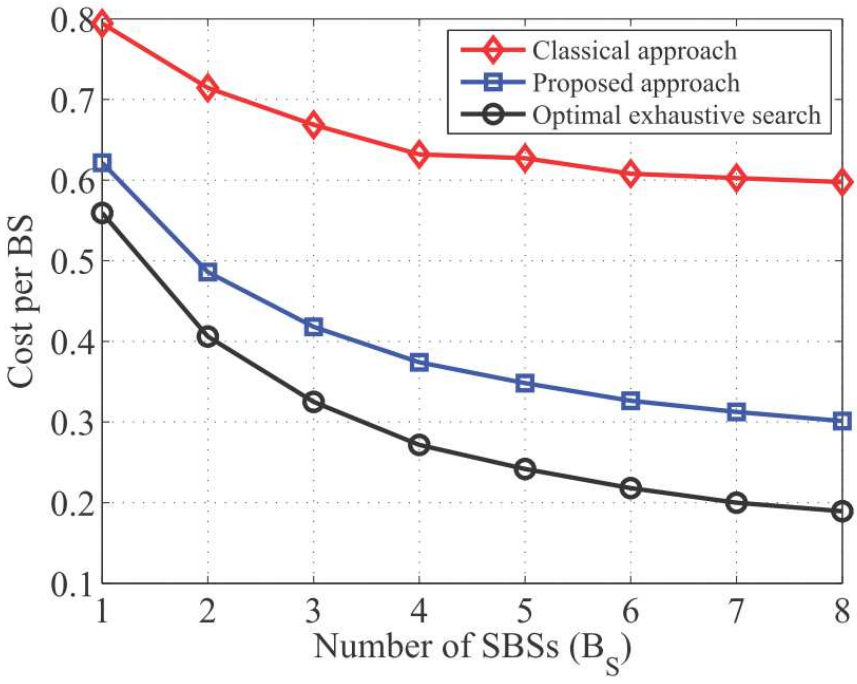


Figure 51. Variation of the cost per base station with respect to the number of SBSs. The number of UEs is fixed to 100.

Figure 51 shows the average cost achieved per base station as the number of SBSs varies. The cost captures the tradeoff between load and energy consumption. As the number of base stations increases, the total energy consumption of the network increases. However, the load from a fixed number of UEs is distributed among the various base stations. Therefore, the energy required to handle the load decreases per base station resulting in a cost decrease per base station for all three approaches as seen in Figure 51. In the proposed approach, the base stations switch to a sleep state when there are no UEs in their vicinity. Consequently, Figure 51 shows that the proposed approach exhibits a considerable cost reduction compared to the classical model (where all base stations are ON). We also see that for a single SBS, the proposed approach exhibits a cost reduction of 21.8% and it reaches up to 49.5%, relative to the classical approach, with eight SBSs. Figure 51 also shows that the difference in the average performance between the proposed approach and the optimal exhaustive search solution reaches about 18.8% at $B_S = 8$ SBSs. However, the optimal solution requires an exhaustive centralized search which yields significant overhead. Indeed, the gap between the exhaustive search and the proposed approach is a byproduct of the uncoordinated decision making processes and the selfish behavior of the players (base stations).

4.2 Hardware efficiency optimization for energy saving

4.2.1 Efficiency and linearity enhancement in power amplifiers

Wireless communication standards impose stringent requirements on linearity performance of power amplifiers (PA). In addition, since the PA consumes most of the energy in telecommunication equipments (almost 60% for macro base stations), its power-efficiency becomes a primary concern. In fact, with the growth in voice and data communication usage, a high efficiency will contribute to the reduction of both the UE energy consumption resulting in a longer battery lifetime, and the base station energy consumption resulting in power savings and reduced environmental pollution. However, these two requirements, linearity and

power efficiency, tend to be mutually exclusive in traditional PA design, so that any increase of the PA linearity by amplifying in linear region is usually achieved at the expense of the efficiency and conversely. This PA issue is a primary concern when considering multi-carrier modulations like Orthogonal Frequency Division Multiplexing (OFDM) which are prone to high Peak-to-Average Power Ratio (PAPR) resulting in low power efficiencies.

In such conditions, techniques including linearization and PAPR reduction have been proposed separately in the literature in order to improve the performance of the transmitters, including PA. The linearization ensures high linearity of the PA in order to avoid carrier inter-modulations and to respect the power mask. In wireless communication systems, the most promising and cost-effective linearization method is predistortion which guarantees an acceptable linearity level of the PA over its intended power range. In case of the PAPR reduction, the dynamic range of the signal is reduced, which allows the PA to operate closer to the saturation power more efficiently. PAPR reduction includes some techniques like clipping and filtering, coding and tone reservation. Since PAPR reduction is associated to linearization and improves its effectiveness, the methodology of PA design focuses on a trade-off between linearity and power efficiency represented by linearization and PAPR reduction respectively.

Several works in literature consider the association of PAPR reduction technique and digital predistortion as linearization method aiming to improve the performance of the PA in terms of linearity and efficiency. They can be classified in two approaches. The first approach consists of combining a PAPR reduction technique followed by predistortion; the second one takes into consideration the mutual effects of PAPR reduction and predistortion in order to propose an optimal combination. Obviously, the second approach will achieve better performance and a good trade-off between the PA linearity and the power efficiency.

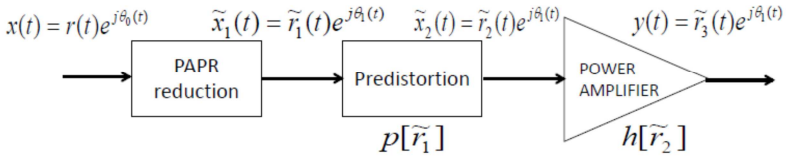


Figure 52. PAPR and linearization processes

We consider the simplified transmission chain presented in Figure 52. The OFDM signal $x(t)$ becomes $\tilde{x}_1(t)$ after PAPR reduction technique and $\tilde{x}_2(t)$ after predistortion. The amplified signal is $y(t)$. The PA is a memoryless Solid State Power Amplifier (SSPA). Under the assumption that the power amplifier has no phase distortion, signals $\tilde{x}_1(t)$, $\tilde{x}_2(t)$ and $y(t)$ have the same phase $\theta_1(t)$. Consequently AM/AM and AM/PM power amplifier characteristics are given by equations

$$h(r) = \frac{r}{\left(1 + \left(\frac{r}{A}\right)^{2b}\right)^{\frac{1}{2b}}} \tag{75}$$

$$p(r) = \frac{r}{\left(1 - \left(\frac{r}{A}\right)^{2a}\right)^{\frac{1}{2a}}}, \quad r \in [0, A[\tag{76}$$

$$\phi(r) = 0 \tag{77}$$

where a and b are the knee factors of the amplifier and predistorter respectively. We define a predistortion error $\epsilon(\tilde{r}_1)$ that compares the amplified signal $y(t)$ to the signal before predistortion $\tilde{x}_1(t)$. This error quantifies the performance of predistortion and is expressed by:

$$\epsilon(\tilde{r}_1) = |\tilde{r}_1 - \tilde{r}_3| \tag{78}$$

It can be shown that the upper bound of $\epsilon(\tilde{r}_1)$ is given by

$$\epsilon(\tilde{r}_1) \leq \tilde{r}_1 \left|1 - 2^{\frac{b-a}{2ab}}\right| \tag{79}$$

In the rest of this section, this approximation will be considered to determine the second order moment of $\epsilon(\tilde{r}_1)$ depending on the distribution of the signal after the PAPR reduction technique.

We assume that the OFDM signal $x(t)$ is characterized by a complex stationary Gaussian process. Therefore, its amplitude $r(t)$ converges to a Rayleigh distribution. Our objective is to study the distribution of the signal after PAPR reduction in order to calculate the second order moment of the predistortion error firstly for probabilistic PAPR reduction methods and secondly for amplitude clipping methods.

PAPR reduction methods can be classified in two categories: the ones that modify the PAPR distribution and the ones that do not, i.e. where the Gaussian distribution remains Gaussian. In the case where the distribution of the signal $\tilde{x}_1(t)$ remains Gaussian (with Selective Mapping or Partial Transmit Sequences methods for instance) with mean power denoted $P_{\tilde{r}_1}$, the second order of the distortion error is given by

$$\widehat{m}_{2\_max}^{(prob)} = P_{\tilde{r}_1}\left(1 - 2^{\frac{b-a}{2ab}}\right)^2 \left(1 - \left(PAPR_{\tilde{r}_1} + 1\right)e^{-PAPR_{\tilde{r}_1}}\right) \tag{80}$$

In the second case, we have considered the clipping method which is the most used due to its simplicity and its reduction gains. In this case, the second order of the predistortion error is given by

$$\widehat{m}_{2\_max}^{(clip)} = \frac{P_{\tilde{r}_1}}{\gamma}\left(1 - 2^{\frac{b-a}{2ab}}\right)^2 \left(1 - e^{-\gamma PAPR_{\tilde{r}_1}}\right) \tag{81}$$

The Error Vector Magnitude (EVM) of the amplified signal is defined as the ratio of the Root Mean Square (RMS) of the predistorsion error to the root of the mean power of the signal after PAPR reduction. As a result, the EVM of $y(t)$ is given by eq. (82) for the probabilistic case and by eq. (83) for the clipping case. γ is the ratio of the signal power after and before PAPR reduction.

$$EVM_{max}^{(prob)} = \left|1 - 2^{\frac{b-a}{2ab}}\right| \sqrt{1 - \left(PAPR_{\tilde{r}_1} + 1\right)e^{-PAPR_{\tilde{r}_1}}} \tag{82}$$

$$EVM_{max}^{(clip)} = \left|1 - 2^{\frac{b-a}{2ab}}\right| \sqrt{\left(\frac{1}{\gamma}\right)\left(1 - e^{-\gamma PAPR_{\tilde{r}_1}}\right)} \tag{83}$$

Each simulation considers 5000 randomly generated OFDM symbols with 64 sub-carriers each 16-QAM modulated. The difference between theory and simulations comes from the aforementioned upper bound. The output PAPR has been fixed to 5 dB.

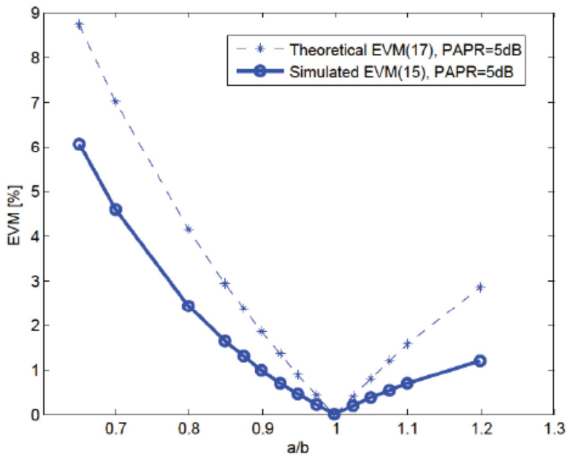


Figure 53. EVM vs a/b ratio (for probabilistic PAPR reduction methods).

Environmental or green issues, such as the need to reduce both direct and indirect CO_2 emissions, and equipment size, are nowadays up-to-date concerns. Several operators have pledged to work with suppliers to increase the energy efficiency of their networks but this is mainly feasible if PA efficiency is greatly improved. Hence, an improvement in PA efficiency will boost the battery lifetime of mobiles. Improving PA efficiency for high PAPR systems like OFDM is particularly challenging due to the need to use linear PAs to meet the critical RF performance criteria. PA is a key component of wireless RF transmitters and consumes the major part of the total power. The main characteristic of the PA is that it dissipates power whatever the amplitude of the input signal. Figure 54 illustrates a basic PA power budget.

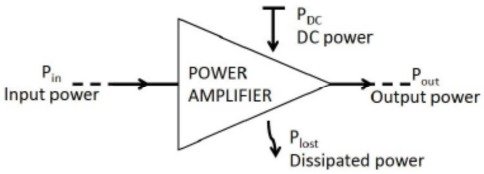


Figure 54. Simplified PA power budget.

Referring to Figure 54, the PA efficiency can be defined as

$$\eta_{DC} = \frac{P_{out}}{P_{DC}} \tag{84}$$

The maximum efficiency is achieved at the maximum amplitude of the linear output signal. This efficiency mainly depends on the input signal PAPR as shown in the next subsection but also on the PA class. η_{DC} cannot exceed 50% and 25% for class A and B power amplifiers respectively when operating in linear conditions.

As explained before, PA's linearity and its efficiency are mutually exclusive and cannot be achieved simultaneously. So the main challenge of designers is to achieve an optimal tradeoff by applying PAPR reduction and linearization techniques. Predistortion can be used to improve PA's linearity while PAPR reduction increases its efficiency. Nevertheless, linearization and PAPR reduction methods are studied and applied independently. After some derivations, it is shown that EVM is a function of the PA efficiency and the predistortion quality as well, as illustrated in

$$EVM_{max}^{(prob)} = \left|1 - 2^{\frac{b-a}{2ab}}\right| \sqrt{1 - \left(\frac{\eta_{DC}}{G}\right)^{\frac{1}{g}} \left(1 - \frac{1}{g} \ln\left(\frac{\eta_{DC}}{G}\right)\right)} \quad (85)$$

This dependancy is illustrated in Figure 55 for different values of a/b ratios.

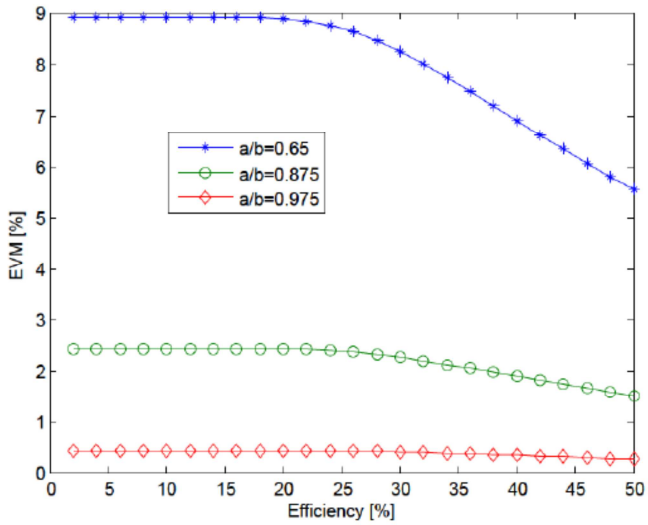


Figure 55. EVM vs power efficiency.

Figure 56 shows the evolution of the EVM metric measured between the amplified signal and the PAPR reduced signal for different IBO values when predistorter is optimal ($a = 2$). To give an idea, we observe that the IBO = 5 dB, corresponding to the targeted PAPR, could be a good trade-off between the linearity and the power efficiency. When IBO < 5 dB, the EVM increases from 0 to 17% for both Selective Mapping (SLM) and Amplitude Clipping. This means that the linearity is degraded while the power efficiency is improved. When IBO > 5 dB, the EVM is minimal, 0%, what that means is that the highest linearity is achieved while the efficiency is decreasing.

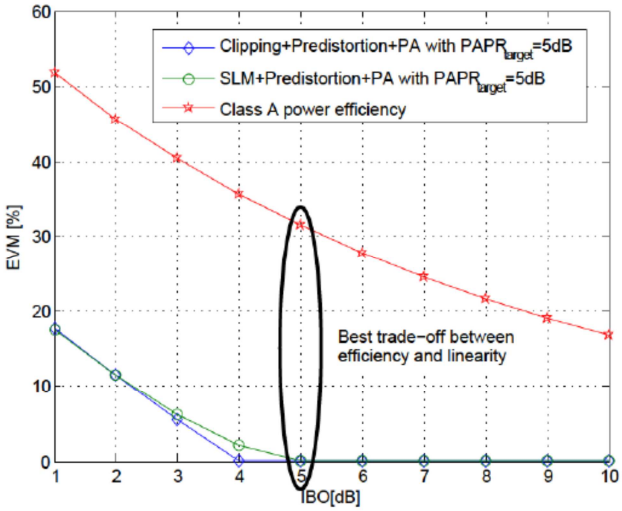


Figure 56. Efficiency and linearity trade-off illustration.

This study, corresponding to scenario 2.9.9 in [D.2.2], has provided a theoretical analysis of the tradeoff between linearity measured by EVM metric and the efficiency in OFDM context. Analytical expressions of the EVM and the power efficiency for a memoryless SSPA have been formulated based on systems where PAPR reduction technique is followed by a predistortion before the nonlinear PA. The validity of the theoretical expressions has been verified through simulations for Amplitude Clipping and SLM techniques. Thus the performance of an OFDM system with nonlinear PA where a PAPR reduction technique is followed by a predistortion can be estimated theoretically without the need to perform extensive simulations. Some other configurations can be considered taking into account the memory effects of the PA or considering other linearity metrics like Adjacent Channel Power Ratio (ACPR). This will be the subject of our future work.

4.2.2 Joint optimization methods between PAPR reduction and linearization

In this work, we propose a new vision of predistortion as an adding signal technique. Traditional predistortion consists of applying a non-linear function $P(.)$ to an input signal, which is the inverse of the PA characteristic so that the concatenation of this function and the PA is equivalent to a linear function. In addition, if the memory effects of the PA are negligible, predistortion is therefore based on non-linear and non-memory function. Thanks to Bussgang theorem and given that multi-carrier signal converges to a complex Gaussian process, it is possible to express the output of the predistortion as the sum of its input and an additional term. Let $s[n]$ denote the input signal and $c^{dpd}[n]$ be the additional signal for predistortion. This predistortion signal $c^{dpd}[n]$ is in fact the signal for PA non-linearity compensation. The output of the predistortion $\tilde{s}[n]$ is then expressed as follows:

$$\tilde{s}[n] = s[n] + c^{dpd}[n] \tag{86}$$

Similarly, it can be proved that all PAPR reduction methods based on a non-linear function (like clipping and companding) can be formulated as adding signal techniques. Generally speaking, the principle of adding signal techniques supposes a peak cancellation signal $c^{papr}[n]$ that, added to the input multi-carrier signal $s[n]$, reduces its PAPR. Then, the peak reduced signal $\hat{s}[n]$ can be expressed as follows:

$$\hat{s}[n] = s[n] + c^{papr}[n] \tag{87}$$

Performing PAPR reduction and predistortion by adding a signal allows a common vision of these two operations. Let us consider $s[n]$ as an input multi-carrier signal, this supposes that it can generate a peak reducing signal $c^{papr}[n]$ and a predistortion signal $c^{dpd}[n]$. Then, the output signal is expressed as follows:

$$\check{s}[n] = s[n] + c^{papr}[n] + c^{dpd}[n] \tag{88}$$

As a consequence, a new additional signal can be defined:

$$c[n] = c^{papr}[n] + c^{dpd}[n] \tag{89}$$

for PAPR reduction and predistortion simultaneously. Then it can be concluded that a single additional signal can be generated to jointly perform PAPR reduction and predistortion. This joint approach by adding a signal is of particular interest for systems for which a dynamic adaptation of PAPR and predistortion is necessary in real time.

Most of the works dealing with the combination of PAPR reduction and predistortion considers a PAPR reduction technique such as clipping and tone reservation followed by predistortion. This is quite a straightforward approach. In this work, we propose new combination scenarios

based on the adding signal technique. The first scenario presents the series combination, where the PAPR reduction and predistortion techniques are combined in cascade and the predistortion signal is generated according to the peak reduced signal. In the second scenario, we propose to perform PAPR reduction and predistortion in parallel. For the following, let us define $c_1[n]$ as the time additional signal for PAPR reduction and $c_2[n]$ as the time predistortion signal.

The idea of serial combination is basically to combine PAPR reduction and predistortion in series. This has already been done previously but the main difference here is that PAPR reduction and predistortion are performed by means of adding a signal. The block diagram of this combination scenario is depicted in Figure 57.

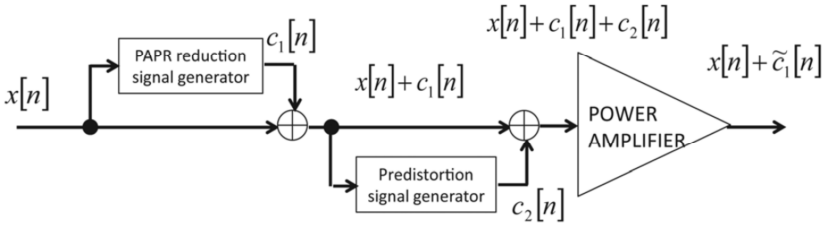


Figure 57. Serial approach for joint PAPR and predistortion by adding signal strategy.

Unlike the general trend, parallel combination performs PAPR reduction and predistortion simultaneously. The block diagram of the scenario is presented in Figure 58. From the OFDM signal $x[n]$, the additional signals $c_1[n]$ and $c_2[n]$ are generated and added before passing through the PA. Since the signal before PA corresponds to the peak reduced and predistorted signal, it can be written as follows

$$y[n] = x[n] + c_1[n] + c_2[n] \tag{90}$$

Ideally, the predistortion signal $c_2[n]$ entirely compensates the PA's non-linearities so it is removed after the PA so that the amplified signal corresponds to the peak reduced signal $x[n] + c_1[n]$.

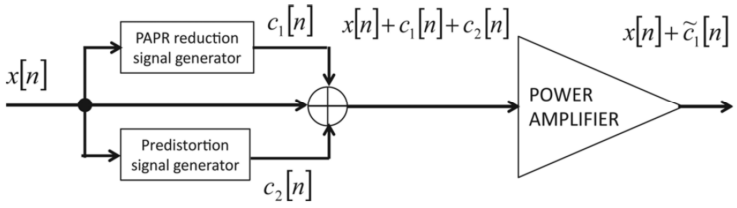


Figure 58. Parallel approach for joint PAPR and predistortion by adding signal strategy.

Compared to the series combination, the main advantage of parallel combination is that the PAPR reduction and predistortion are performed at the same time, either in the time or frequency domain, by means of adding a signal. Indeed, the computing processes can be shared (mutualised) in order to reduce the implementation complexity. In both combination scenarios, the additional signal for PAPR reduction is generated thanks to the gradient algorithm for tone reservation method proposed by Tellado [Tell99]. Next, the simulation results including the details of this algorithm are discussed.

Simulation results

The performance evaluations in this section are related to scenario 2.9.9 defined in [D.2.2].

Simulations are based on the IEEE 802.11a/g standards, where the OFDM system has 64 sub-carriers and a 16-QAM modulation. We consider a SSPA amplifier modelledby a Rapp model with a 'knee factor' $b = 2$. Its corresponding predistortion function $P(.)$ is given by eq. (75), eq. (76) and eq. (77), see Section 4.2.1. We will illustrate the proposed predistortion combination scenarios of PAPR reduction and predistortion by EVM and PSD as well.

The additional signal for the PAPR reduction is generated thanks to the gradient algorithm for Tone Reservation (TR) proposed in [Tell99]. The TR PAPR reduction signal uses the reserved sub-carriers of a given standard to reduce the PAPR and to avoid by the way bit error rate degradation. According to [Tell99], the PAPR reduction can be expressed as a convex optimisation problem as follows

$$\min_{\hat{C}_1} \left\| x + \hat{\boldsymbol{Q}}\hat{C}_1 - g\left(x + \hat{\boldsymbol{Q}}\hat{C}_1\right) \right\|_2^2 \tag{91}$$

where x is the input OFDM signal and $g(.)$ is a clipping function, $\hat{C}_1$ represents the frequency symbols of the PAPR reduction signal c_1. $\hat{Q}$ is the discrete Fourier transform (DFT) matrix, whose coefficiencts are given by:

$$\hat{Q}(a,b) = e^{\frac{2j\pi}{NL}ab}, a \in [0, NL-1], b \in [0, NL-1] \tag{92}$$

In eq. (92), N is the total number of OFDMs sub-carriers and L the oversampling factor. As a solution of eq. (91), [Tell99] proposed a gradient iterative algorithm that updates the PAPR reduction signal thanks to a subset of reserved sub-carriers in a finite number of iterations.

Here, we propose an iterative error compensation algorithm that can be derived in time or in frequency domain. Its concept is based on the compensation of the error between the amplified signal and the OFDM signal. Figure 59 presents the block diagram of the algorithm in the frequency domain and its equivalencein the time domain is depicted in Figure 60. The two versions have the same performance so we focus on the frequency version for several advantages. The main one is that PA estimation and characterization in the frequency domain are more precise. As shown in Figure 59, the PA's characteristic is directly learned thanks to the feedback. Initial frequency symbols $X[k]$ are taken as reference and $C^{dpd}[k]$ represents the frequency domain symbols of the predistortion signal $c^{dpd}[n]$. $C^{dpd}[k]$ is initialized to zero. The predistorted symbols are obtained by adding symbols $C^{dpd}[k]$ to $X[k]$ such as:

$$\hat{X}[k] = X[k] + C^{dpd}[k] \tag{93}$$

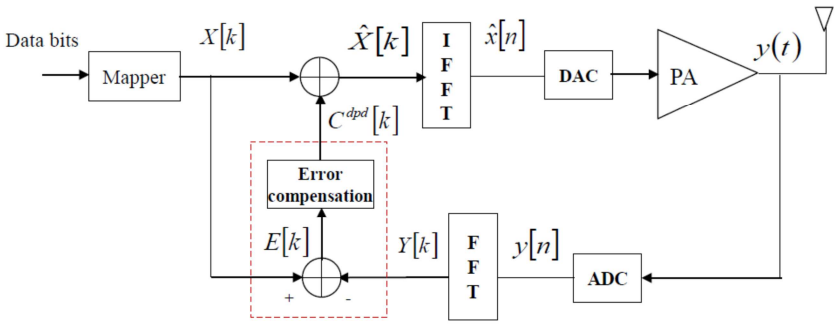


Figure 59. Frequency compensation error algorithm.

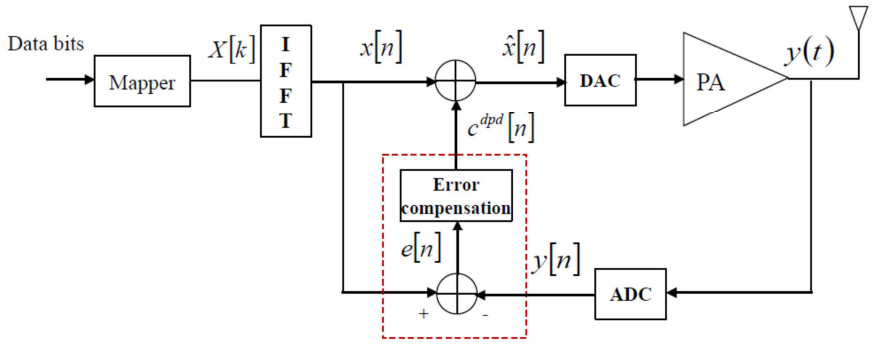


Figure 60. Frequency compensation error algorithm.

As a result, the proposed algorithm for digital predistortion in the frequency domain is:

Require: $X[k]$ input OFDM symbols.
1: $\mu \Leftarrow$ convergence rate.
2: $i \Leftarrow 0$ (initiate index of iterations)
3: $C^{dpd(i)}[k] \Leftarrow 0$ (initiate the additional signal)
4: **repeat**
5: $\hat{X}^{(i)}[k] \Leftarrow X[k] + C^{dpd(i)}[k]$ (predistorted symbols)
6: $Y^{(i)}[k] \Leftarrow \mathbf{FFT}\left\{\mathbf{PA}\left(\mathbf{IFFT}\left\{\hat{X}^{(i)}[k]\right\}\right)\right\}$ (feedback)
7: $E^{(i)}[k] \Leftarrow X[k] - Y^{(i)}[k]$ (Calculate the error)
8: $C^{dpd(i+1)}[k] \Leftarrow C^{dpd(i)}[k] + \mu * E^{(i)}[k]$ (adaptation)
9: $i \Leftarrow i + 1$ (increment iteration index)
10: **until** $i \geq$ Maximum number of iterations or $\left|E^{(i)}[k]\right| \leq \epsilon$

The first metric used is EVM (already defined). The other figure of merit is the PSD (Power Spectral Density) that allows to evaluate the OOB (Out Of Band) effects of the PA. A spectral mask is defined for each standard and the PSD signal must remain below that threshold.

We performed some simulations on the EVM and PSD for the proposed combination scenarios of PAPR reduction and predistortion obtained by means of adding a signal. There are two combination scenarios presented in the previous section: in series and in parallel. Their performance are analysed and compared with the traditional predistortion when combined with a PAPR reduction technique. All simulation results in this section are based on 5000 randomly generated OFDM symbols.

We choose 10 iterations for the TR and 3 iterations for the proposed predistortion algorithm. To avoid bit error rate regrowth, the additional signal for PAPR reduction is added on the 12 null carriers of the IEEE 802.11a/g standards. The clipping ratio of TR is fixed to 5 dB. Figure 61 presents the evolution of the EVM metric in the case of parallel and series combination scenarios compared to the combination of TR and predistortion in their classical formulations. We can note that the EVM value becomes null after IBO = 7 dB. In addition, simulation results show that the EVM performance of parallel and series combination scenarios are very close to that of the combination of TR and traditional predistortion.

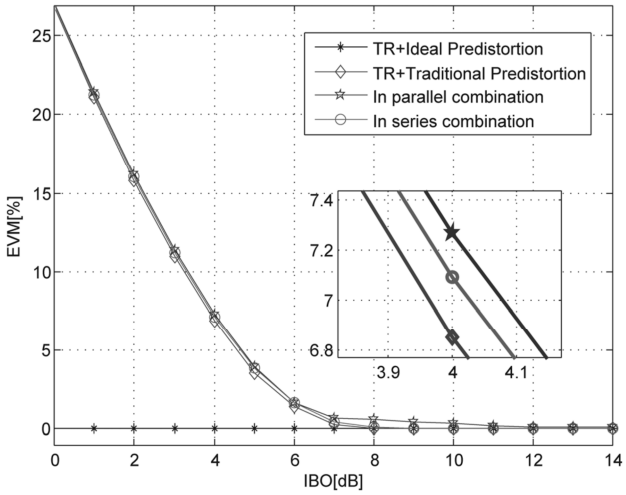


Figure 61. EVM for series and parallel combination.

We also simulate the PSD's performance for both parallel and series combination scenarios in Figure 62. The results confirm that the parallel and series combination scenarios have similar performances and both are close to the traditional combination of TR and predistortion. Thanks to the predistortion by means of adding a signal technique, the parallel and series combination scenarios have a slight advantage for the OOB interferences mitigation when compared with the association of TR and traditional predistortion for low IBO (≤ 4 dB).

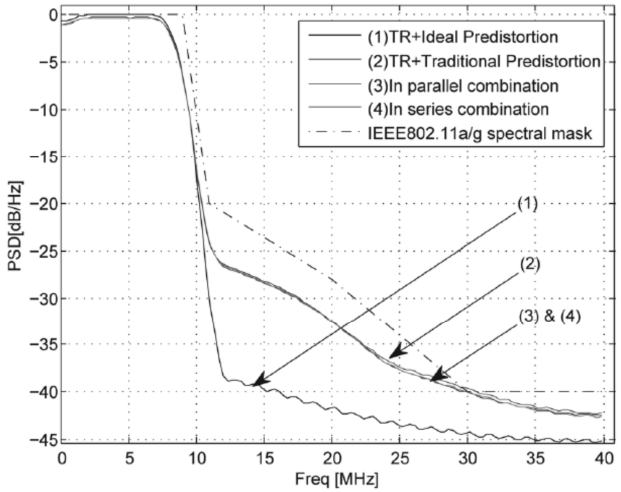


Figure 62. PSD for series and parallel combination.

As a conclusion, we have proposed a new predistortion formulation by means of adding a signal and two different ways to generate the PA's non-linearity compensation signal. The first way is based on the Bussgang theorem and needs a preliminary step of PA estimation. Since the predistortion is expressed as an adding signal technique, we have proposed afterwards two combination scenarios of PAPR reduction and predistortion by means of adding a signal: series and parallel. The performances in terms of EVM and PSD of both predistortion by means of adding a signal and combination scenarios are analysed and compared. Simulation results show that the error compensation algorithm is very close to the traditional predistortion. However, a large number of iterations slightly degrades the PSD. We showed that predistortion can be expressed as an adding signal technique, this allows us to have a common vision of PAPR reduction and predistortion. In order to further develop what has been presented in this work, future research work will be investigating the combination of

PAPR reduction and predistortion formulated as a single optimisation problem which generates only one additional signal for both PAPR reduction and predistortion.

4.2.3 Energy efficiency strategies over different power levels in the power amplifier

Base station (BS) radio equipment comprises different blocks such as an antenna interface, a power amplifier (PA), a Radio Frequency (RF) small-signal transceiver section, a baseband (BB) stage, a DC-DC power supply, an active cooling system, and a AC-DC unit (Main Supply) for connection to the electrical power grid.

The PA is one of the most power-consuming components. Furthermore it is usually designed for maximum traffic load, performing the highest efficiency at maximum RF output power. Nevertheless when the traffic load decreases, lower RF output power levels are required and the energy efficiency gets worse due to the PA power characteristics.

This section is focused on an innovative solution to be applied at PA level. The idea is to make it work in different operating points, providing the required RF power level according to the traffic load and important energy savings in the base station radio equipment.

This technique could be applied in different base station types because it is defined at component level. In 3GPP standard [3GPP12], different base station types are defined: Wide Area BS, Medium Range BS, Local Area BS and Home BS. Concerning the PA, Table 7 presents the mean power level per carrier for base stations operating in single carrier available at the antenna connector in different scenarios.

Table 7. Mean power level per carrier for BS type [3GPP12]

BS type	Scenario	Rated Output Power
Widea Area BS	Macro cell	46 dBm (typ)
Medium Range BS	Micro cell	< +38 dBm
Local Area BS	Pico cell	≤ + 24 dBm
Home BS	Femto cell	≤ + 20 dBm

Macro and micro base station PAs usually exploit energy efficiency techniques, such as digital pre-distortion or crest factor reduction, to lead a better performance. Thanks to these techniques, the typical back-off in PA for LTE downlink signals is around 8 dB. On the other hand, these techniques are not applied in pico and femto base stations due to their cost and complexity, so the typical back-off increases up to around 12 dB. In all cases, the use of a signal load adaptive PA could bring important reduction in the power consumption.

Figure 63 presents a DC power consumption breakdown for different types of base stations at maximum load [EAR12a]. This analysis could be used as a reference to assess the potential improvement using this technique. In macro base stations, the PA dominates the total power consumption. Meanwhile, the breakdown is more balanced in micro base stations. And finally, in smaller base stations like pico and femto, baseband parts dominate the overall power consumption. Anyway, the PA is a relevant power-consuming component in all base stations.

In absolute terms, a femto PA might consume about 100 times less power than a macro PA but it also serves a limited number of users and covers a limited amount of area. An optimal trade-off should be found at network level in HetNets to optimize the overall energy efficiency. The proposed technique could be applied to different base stations. However, the potential improvement will be related to the base station type. Therefore, in a HetNet, the potential improvement will be related to grouping of different base stations.

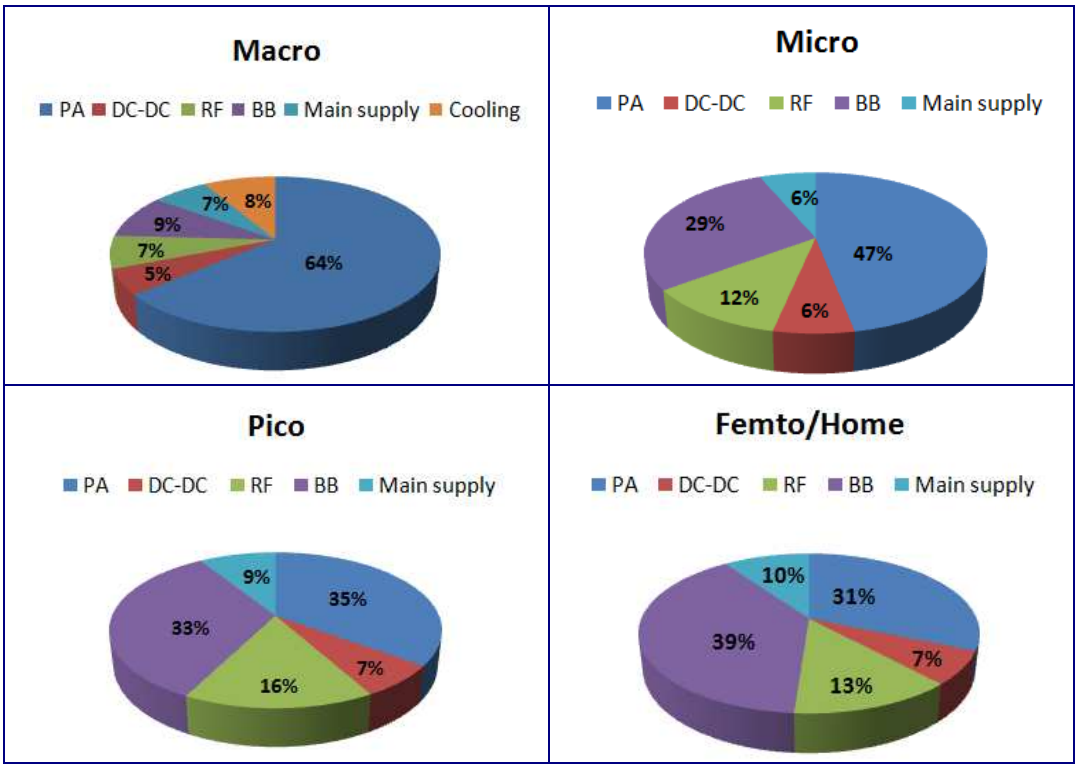


Figure 63. Power consumption breakdown for different base station types [EAR12a].

As it is shown in Figure 63, the PA power consumption could reach up to 64% in a macro base station and 47% in a micro base station. This power consumption is related to maximum traffic load and decreases at medium and low traffic load because lower RF output power levels are required. Nevertheless the PA energy efficiency at these lower traffic load levels is poor. The proposed solution will improve the PA energy efficiency at medium and low traffic load by dynamically changing the PA operating points to adjust the RF output power to the required level according to the traffic load.

Analyzing a 24 hours traffic load profile for urban areas [EAR12a], a time ratio for different traffic load ranges can be estimated (see Figure 64). Medium and low traffic load situations occur during most of the day, and consequently there is a significant potential to improve energy efficiency in a base station. During 13.7% of time, the traffic load reaches values close to the maximum which is related to peak traffic hours. As an example, 36.6% of the time, the signal load is lower than 50% of maximum traffic.

The proposed signal load adaptive PA will work in different states to enable the required RF output power according to the traffic load. These intermediate states are performed applying different operating points to the PA instead of a unique one, reducing the energy consumption in medium and low traffic loads. Figure 65 represents these PA intermediate operating points, showing their performance in terms of maximum output power and power added efficiency (PAE). State 1 represents the PA typical operating point, designed for maximum signal loads showing poor efficiency at low and medium traffic load. The concept is to dynamically adapt the different PA operating points (or states) to the traffic load demand.

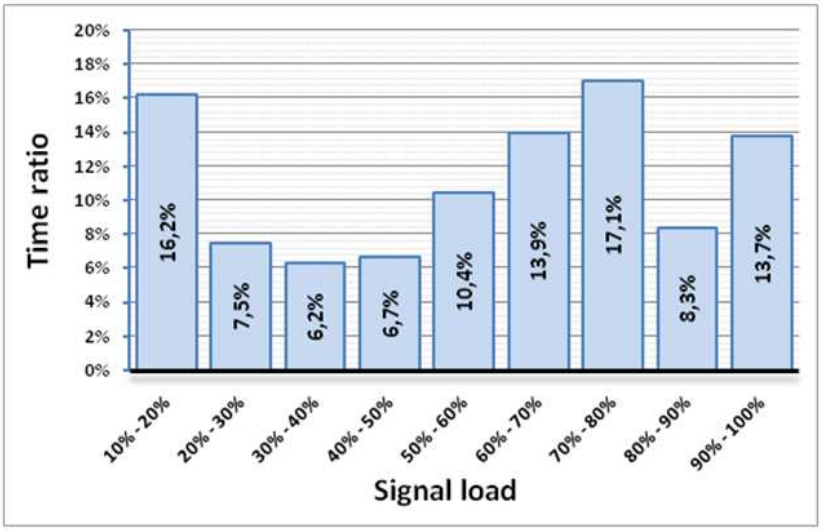


Figure 64. Time ratio for different signal load ranges in an urban area during a day.

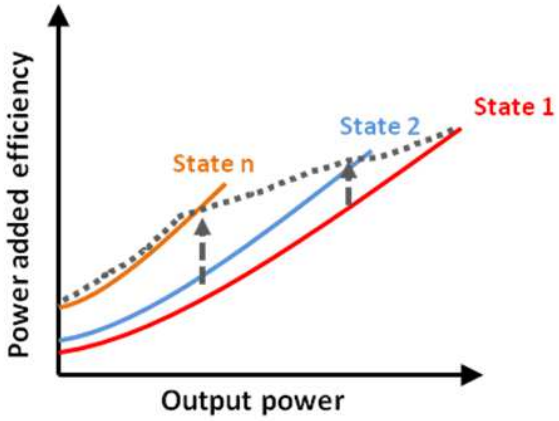


Figure 65. Power added efficiency of the signal load adaptive PA for different operating points (states).

Table 8 presents a relation between the signal load ranges and the required RF output power. At maximum traffic load, the required RF output power ($Pout$) will reach the highest value, while the required RF $Pout$ will decrease when the signal load reduces.

A simplified block diagram of the proposed signal load adaptive PA is shown in Figure 66. The implementation of this technique could be addressed in different ways:

- Drain voltage control: modifying the drain voltage between the level needed for highest energy efficiency at maximum traffic load and zero volts.
- Drain current control: modifying the gate voltage from its quiescent value (constant operating point) to a voltage close to the pinch off one (which effectively turns off the drain current).
- Coordinated drain voltage and current control.

Table 8. Mean power level per carrier according to the signal load ranges

Signal load	Required RF Pout (dBm)
90 – 100	$Pout_{MAX}$
80 – 90	$Pout_{MAX} - 0.5\ dB$
70 – 80	$Pout_{MAX} - 1\ dB$
60 – 70	$Pout_{MAX} - 1.5\ dB$
50 – 60	$Pout_{MAX} - 2.2\ dB$
40 – 50	$Pout_{MAX} - 3\ dB$
30 – 40	$Pout_{MAX} - 4\ dB$
20 – 30	$Pout_{MAX} - 5.2\ dB$
< 20	$Pout_{MAX} - 7\ dB$

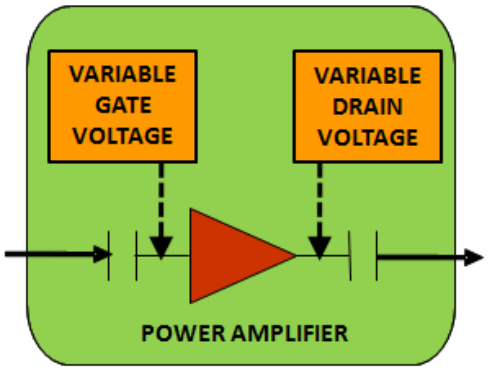


Figure 66. Block diagram of the signal load adaptive PA.

Apart from an adequate bias optimization, a selection of the most suitable transistor technology is required. Laterally Diffused MOS (LDMOS) technology is the leading RF power technology for base station applications, since it offers the best values in terms of performance over cost ratio and has a proven reliability, as well as supply chain for mass production. Over the last few years, the performance of LDMOS transistors has been continuously improving, especially in terms of power efficiency and gain. Nevertheless, Gallium Nitrate (GaN) technology is considered as the best long term replacement for the LDMOS. This is mainly due to its numerous advantages at the transistor level, such as high peak transition frequency (typically 20 GHz for GaN devices and 11 GHz for LDMOS devices) and high current density (typically 1000mA/mm for GaN devices compared to 150 mA/mm for LDMOS devices) [CHE10].

The proposed solution will be evaluated in both transistor technologies, LDMOS and GaN. Moreover, the evalution will be done over transistors with different maximum output power levels to ensure that the concept could be applied to different configurations.

This innovation corresponds to scenario 2.9.2 in [D.2.2].

5 SPECTRUM RESOURCE ALLOCATION MECHANISMS

One of the important aspects in future network deployment will be the increasing bandwidth demand, which is in confrontation with spectrum availability. In case of LTE-A networks, carrier aggregation (CA) will play an increasingly important role in providing operators the maximum flexibility for using their available spectrum. With CA, LTE-A will be able to deliver much higher throughputs than otherwise possible, by combining spectrum blocks up to 100MHz.

The scope of SHARING Task 4.4 is to study various self-organizing techniques for resource allocation, taking into account carrier aggregation features and capabilities of the future networks in the context of heterogeneous networks deployment. For example, the task will develop distributed and centralized radio resource management (RRM) strategies that ensure the stability of the network in the context of bursty traffic. Furthermore, the task will develop a learning framework in order to improve the performance of RRM strategies. A fundamental question to answer is how users can learn and improve their network utilities (rate, QoS, QoE, probability of success, etc.) in heterogeneous networks based on local information and noisy measurements?

The task will also investigate an interference management mechanism based on dynamic carrier selection in LTE-A systems with carrier aggregation. This interference management is provided by SON-flavoured algorithms for the selection of primary and secondary component carriers for a group of base stations. The component carrier selection will be based on UE measurements and statistical information about the usage of the component carriers.

Furthermore, the topic of dynamic component carrier selection will be discussed in the context of multi-flow carrier aggregation, where UEs aggregate separate flows coming from different nodes. Here issues pertaining to the flexible assignment of primary and secondary component carriers, cell range expansion techniques are at stake.

Tight time synchronization is necessary for performing advanced RRM. Therefore, the task will propose a cooperative distributed synchronization mechanism, which will improve the basic time reference provided by the macro network.

5.1 Carrier aggregation

5.1.1 Coordinated carrier aggregation in campus of home base stations

In this section we will focus on coordinated carrier aggregation for a scenario with a campus of home base stations. The aim of the proposed coordinated carrier aggregation is to improve the throughput of carrier aggregation capable UEs by the means of coordinated multi-point transmission (CoMP). The carrier aggregation capable UE is receiving in this case data and/or control from multiple base stations that coordinate their transmission such as not to use the same resources for the transmission to the UE. This will ensure transparent operation for the UE and reduce the interference arising from coordinated multipoint transmission. Description of the proposed coordinated carrier aggregation technique is given by the Figure 67.

For a large campus covered by non coordinated home base stations, the available ressources must be splitted between the campus and the neighbouring macro base stations. Roughtly speaking, using too many resources within the campus may cause resource shortage in the macro network and may complicate interference management within the campus as well.

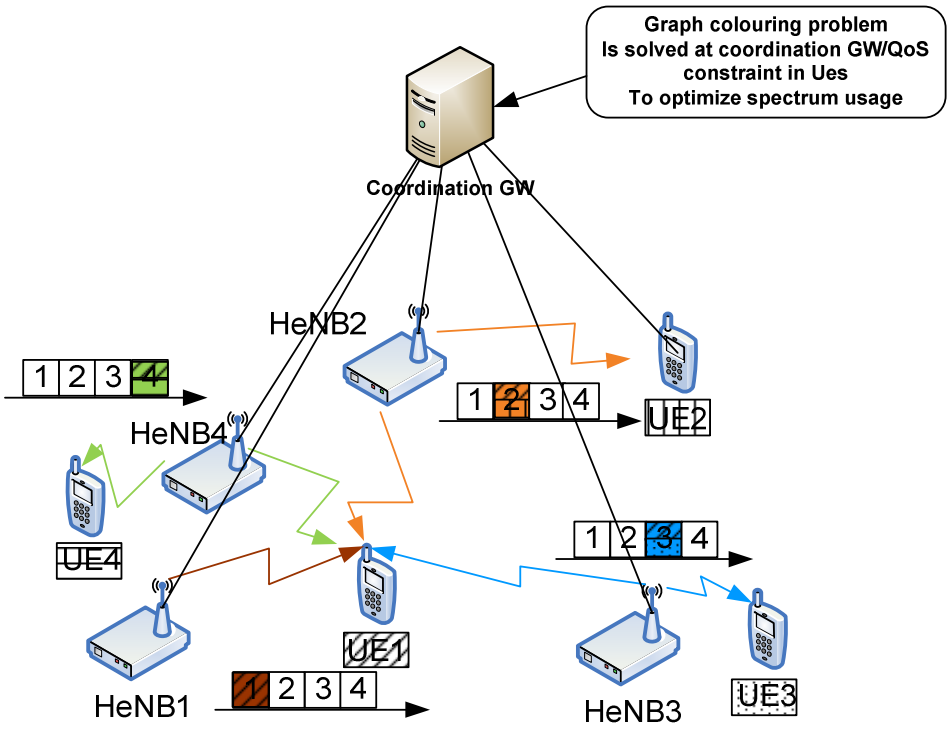


Figure 67. Coordinated carrier aggregation description

For these reasons, we propose a graph colouring based coordinated carrier aggregation with the following three basic steps:

1. Interference graph is constructed using the following principle: two base stations are considered as neighbours if a handover has occurred from one base station to the other during a predefined period of time.
2. Coordination GW (gateway) performs node colouring of the interference graph in order to determine the minimum amount of resources that are needed to perform the coordinated carrier aggregation.
3. Resource marking and allocation is performed by the coordination GW based on the current distribution of active users in the campus on the basis of the available component carriers in the network.

Node colouring problem is generally NP hard, so there is no identified solution algorithm, with a complexity increasing linearly with the number of nodes for general graph topologies. However, various heuristic methods provide coloring that is close to the optimal.

.

Node colouring problem is generally NP hard, so there is no identified solution algorithm, with linear complexity in the number of nodes for general graph topologies. However, various heuristic methods provide coloring that is close to the optimal.

In this study we will focus on the *smallest last* sequential greedy vertex colouring [AKKM], which generates recursively during the colouring process an ordering such that the node with the lowest number of neighbouring nodes is coloured as late as possible. This colouring method is shown to color the edges of planar graphs with at most 6 colors and to color *k*-degenerate graph with at most *k+1* colors, i.e. the worst case performance. We will consider also other sequential greedy colouring algorithms such as the *largest first* ordering, that generates recursively an ordering based on the node with the largest neighbouring nodes in the graph.

The interference graph is constructed from the measurements of the active user terminals as follows: Each node in the interference graph is associated to one home base station in the

campus. Two nodes are connected by an edge if they are neighbours in the neighbour relation tables, stored in each base station or if the user terminal associated to one node reports the signal to intefrence ratio (SINR) measured with respect to the other node such as it is above a given SINR threshold (SINR0). Another construction method will use interference measurements rather than SINR measurements for the interference graph construction.

Once the number of resources is found through vertex colouring, resource allocation is performed by the coordination GW, on more short term basis, based on the actual distribution of the active user terminals in the campus.

In future work we will provide a system level evaluation of the proposed coordinated carrier aggregation with graph coloring using spectrum efficiency CDF and spectrum usage as main performance indicators. . We will also discuss the signalling that may be introduced by the technique and compare it to the LTE/LTE-A networks baseline

5.1.2 Multiflow carrier aggregation

The single-flow CA technique is a basic CA approach as introduced by 3GPP LTE Release-10. In single-flow CA, the macrocell is the aggressor cell and the picocell is the victim cell. Taking a toy example consisting of 2 Component Carriers (CCs), while a picocell performs CRE to offload the macrocellular network, and to transmit UE data and UE control signaling on CC1 the macrocell causes intercell interference if no e-ICIC techniques are applied. On CC2, the expanded range pico UE (ER PUE) does not receive any data while regular PUEs are served on CC2. A recent feature in 3GPP LTE Release-12, referred to as multiflow CA, enables a better use of resources and improves system capacity, where multiple base stations (from different tiers) simultaneously transmit data to a UE on different CCs. Dual-connectivity or multiflow CA, in which users are served by different layers on different CCs, has recently emerged as a key research challenge in 3GPP LTE Release-12. Besides the multiflow CA technique, another related approach to provide an efficient and flexible network performance improvement is to split the control- and user-plane (C-and U-plane). In the introduced concept, the C-plane is always provided at low frequency band to maintain good connectivity and mobility. On the other hand, the U-plane is provided by both the macrocells and the small cells (deployed at higher frequency bands) for data transfer. Since small cells are not configured with cell-specific signals and channels, they are named as Phantom Cells. The Phantom Cell Concept is introduced, because of the lack of capacity solutions for high-traffic outdoor environments that can also support feasible mobility and connectivity. Thus, the concept of macrocell assisted picocells has been introduced as a capacity solution that offers mobility support while capitalizing on the existing macrocellular network. Hereby, the C-plane of a picocell UE is provided by a macrocell in a lower frequency band and the U-plane is provided by the picocell at higher frequency bands. For a macrocell UE, both the C-plane and U-plane are provided by the serving macrocell in the same way as in a conventional LTE system. However, a network architecture that supports the C/U-plane split, and interworking between the macrocell and the Phantom Cell is still required.

In contrast to the single-flow CA in which the macro base station is always the aggressor cell, in multiflow CA either the macro base station or the pico base station is the aggressor cell. This is because both the macro base station and the pico base station perform CRE on their primary CCs, so that a UE can be served on different CCs by different base stations based on its biased received power. Similar to the single-flow CA learning approach, the multiflow CA based e-ICIC learning algorithm considers the pico base stations and the macro base stations as players. Both the pico base stations and the macro base stations learn their optimal primary CC, CRE bias values and power levels. Algorithm 1 summarizes the steps of the Q-learning based e-ICIC for multiflow transmission. The main difference of the multiflow CA and the single-flow CA based e-ICIC learning algorithm, lies in the action definition. It is redefined for multiflow CA based e-ICIC as follows:

Action: For player pico base station p the action set is defined as $A_p = \{C_p, \beta_p, a_{pr}\}\ r \in \{1, \dots, R\}$, and for player macro base station m the action set is defined as $A_m = \{C_{mi}, \beta_m, a_{mr}\}\ r \in \{1, \dots, R\}$,

where C_i $i \in \{1,2\}$ is the component carrier index that can be selected in order to perform CRE on the selected CC and $\beta \in \{0,6,12\}$ dB is the bias value for CRE on selected CC C_i. Furthermore, a_r is the transmit power level over a set of resource blocks $\{1, \dots, R\}$. Hence, the pico base stations and the macro base stations will independently learn the CCs on which they perform range expansion, with which bias value, and how to optimally perform the power allocation. Since both the pico base stations and the macro base stations can be aggressor cells, different power levels are considered for CCs on which the base stations perform CRE, and the regular CCs which do not have CRE. In addition, the one player formulation case is considered, in which the pico base station is the player. In this case, pico base station carries out the multiflow CA based Q-learning procedure and informs the macro base station about its primary CC and the macro base station uses reduced power levels on this CC and higher power levels on secondary CC of the pico base station. A uniform power allocation is considered in this case. However, even if no CRE is performed by the macro base station, a UE can be served by both the pico base station and the macro base station on different CCs at the same time based on its biased received power. This learning algorithm will be coined as MF static QL, while the two player algorithm is named MF dynamic QL.

Algorithm 1 Dynamic Q-learning based e-ICIC algorithm for single- and multiflow CA.

loop
  for player p **do**
    Select primary CC $C^p \in \{1, 2\}$
    Select bias value b^p for primary CC C^p
    Select power level a_r^p according to $\arg\min_{a \in \mathcal{A}^p} Q^p(s, a)$ on both CCs
  end for
  Inform player m about primary CC C^p
  for player m **do**
    Select player p's secondary CC as primary CC C^m
    if multiflow CA **then**
      Select bias value b^m for primary CC C^p
    end if
    Select power level $a_r^m \in \mathcal{A}^m$ according to $\arg\min_{a \in \mathcal{A}^m} Q^m(s, a)$
  end for
  Receive an immediate cost c
  Observe the next state s'
  Update the table entry according to equation (1.3)
  $s = s'$
end loop

Let us consider a network deployment with multiple picocells overlaying a macrocellular network consisting of 3 sectors per macrocell. A network consisting of a set of $\boldsymbol{M} = \{1, \dots, M\}$ macrocells and a set of $\boldsymbol{P} = \{1, \dots, P\}$ uniformly randomly distributed co-channel picocells per macro sector is considered. We consider that the total bandwidth (BW) is divided into subchannels with bandwidth $\Delta f = 180$ kHz. OFDM symbols are grouped into resource blocks (RBs). Both macro- and picocells operate in the same frequency band and have the same number of available RBs, denoted by R. We consider that all transmitters and receivers have single-antennas. A set of UEs $\boldsymbol{U} = \{1, \dots, U\}$ is defined, whereby the UEs are dropped according to scenario #4b [3GPP10a]. We denote by $u(m)$ a macrocell UE, while $u(p)$ refers to a picocell UE. We denote by p_{mr} and p_{pr} the downlink transmit power of macro base station m and pico base station p in RB r, respectively. The SINR at macro UE (MUE) u allocated in RB r of macrocell m is:

$$\gamma_r^{u(m)} = \frac{p_r^{m(u),M} g_{m,u,r}^{MM}}{\underbrace{\sum_{j=1,j\neq m}^{M} p_r^{j(u),M} g_{j,u,r}^{MM}}_{I^M} + \underbrace{\sum_{p=1}^{P} p_r^{m(u),P} g_{p,u,r}^{PM}}_{I^P} + \sigma^2} \quad (94)$$

In eq. (94), $g_{m,u,r}^{MM}$ indicates the channel gain between the transmitting macro base station m and MUE u, $g_{j,u,r}^{MM}$ indicates the link gain between the transmitting macro base station j and MUE u in the macrocell at base station m, $g_{l,u,r}^{PM}$ indicates the link gain between the transmitting pico base station l and MUE u of macrocell m, and σ^2 is the noise power. Furthermore, I^M and I^P are the interference terms caused by the macro base stations and the pico base stations, respectively.

The SINR at PUE u allocated in RB r of picocell p is:

$$\gamma_r^{u(p)} = \frac{p_r^{p(u),P} g_{p,u,r}^{PP}}{\underbrace{\sum_{j=1,j\neq p}^{P} p_r^{j(u),P} g_{j,u,r}^{PP}}_{I^P} + \underbrace{\sum_{k=1}^{K} p_r^{p(u),M} g_{m,u,r}^{MP}}_{I^M} + \sigma^2} \quad (95)$$

Here, $g_{p,u,r}^{PP}$ indicates the link gain between the transmitting pico base station p and its PUE u, $g_{j,u,r}^{PP}$ indicates the link gain between the transmitting pico base station j and PUE u in the picocell at pico base station p, and $g_{m,u,r}^{MP}$ indicates the link gain between the transmitting macro base station m and PUE u of pico base station p. The scenario used in the system-level simulations is based on configuration #4b [3GPP10a], and is also defined as scenario 2.7.4 in [D.2.2]. We consider a macrocell consisting of three sectors and $P = \{2,4,8\}$ pico base stations per macro sector, uniformly randomly distributed within the macrocellular environment. $N_{UE} = 30$ mobile users are generated within each macro sector from which $N_{hotspot} = \lceil 2/3 \cdot N_{UE}/P \rceil$ are randomly and uniformly dropped within a 40 m radius of each pico base station. The remaining UEs are uniformly distributed within the macrocellular area. All UEs have an average speed of 3 km/h. A full buffer model is assumed for the traffic of users. As an evaluation metric, the average UE throughput is considered which is defined as the ratio of number of information bits that the user successfully receives divided by the total simulation time T. If UE u has V downlink packet calls with $W_{v,u}$ packets for the v-th downlink packet call and $b_{w,v,u}$ bits for the w-th packet, the average UE throughput for UE u is:

$$R_u = \frac{\sum_{v=1}^{V} \sum_{w}^{W_{v,u}} b_{w,v,u}}{T} \quad (96)$$

For the proposed frequency domain e-ICIC algorithms an analysis of the tradeoffs for single-flow CA (SF QL) and multiflow CA (MF static QL and MF dynamic QL) is performed. Figure 68 plots the UE throughput for two active picocells per macrocells. While the SF QL and MF static QL algorithms are in average very close to each other, the MF dynamic QL algorithm shows a performance improvement of 47% on average. A close-up view of the cell-edge UE throughput shows that the multiflow CA algorithms outperform the single-flow case. This is because in multiflow CA, cell-edge UEs are served by macro- and picocell at the same time. The behavior of the learning based frequency domain e-ICIC algorithms when increasing the number of picocells per macrocell is depicted in Figure 69. Here, the solid curves belong to the left ordinate showing the total throughput and the dashed curves refer to the right ordinate reflecting the cell-edge UE throughput. It can be observed that the MF dynamic QL algorithm outperforms the other algorithms in terms of total throughput while the SF QL algorithm is slightly better than the MF static QL algorithm for less number of picocells (and vice versa for large numbers). The SF QL algorithm shows the lowest performance for cell-edge UE throughput. It can be concluded that cell-edge UEs benefit more from multiflow CA than from single-flow CA. Interestingly, it can be observed that the MF static QL algorithm outperforms the MF dynamic QL for larger number of picocells. This is because in the two-player case, the macro base station cannot fully adapt to the ICIC strategies of all pico base

stations in the system, when the number of pico base stations is large. In Figure 70 the sum-rate of the proposed frequency domain e-ICIC algorithms for different number of UEs in case of two picocells per macrocell is depicted.

While the MF dynamic QL algorithm outperforms the other two algorithms, a trade-off between the SF QL and MF static QL algorithm can be seen. For more than 40 UEs per macrocell the MF static QL algorithm outperforms the single flow case. Hence, the multiflow CA technique does not only protect cell-edge UEs, but also improves the total sum-rate at high loads.

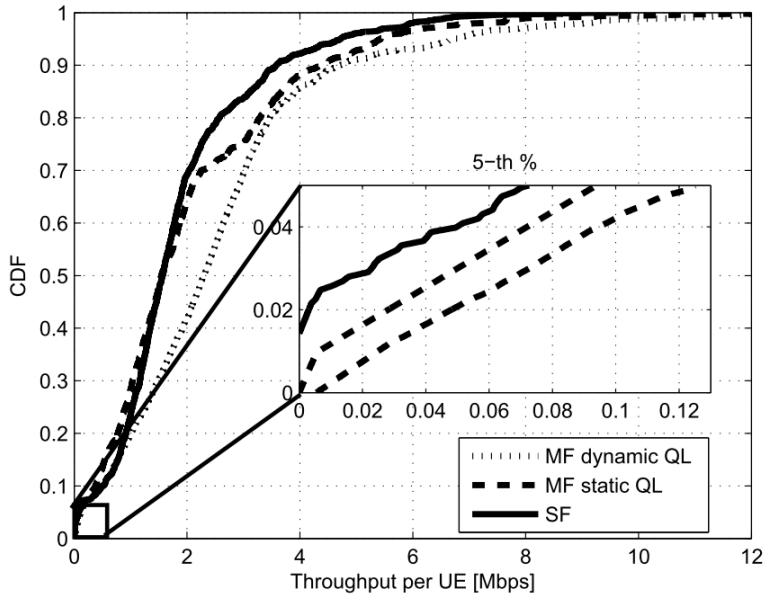


Figure 68. CDF of the UE throughput for the single and multiflow e-ICIC learning algorithm for 2 picocells per macrocell.

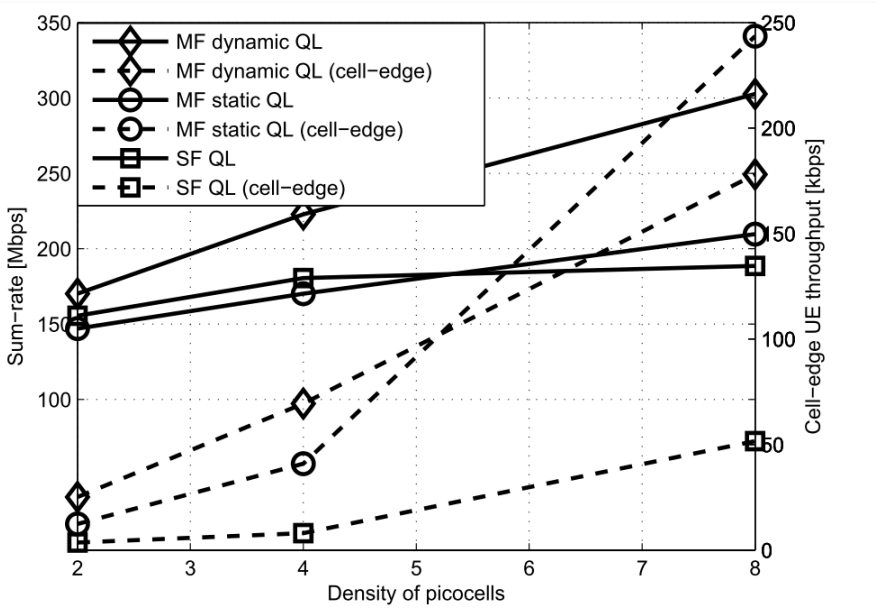


Figure 69. Total-throughput and cell edge throughput versus the number of picocells in frequency domain e-ICIC.

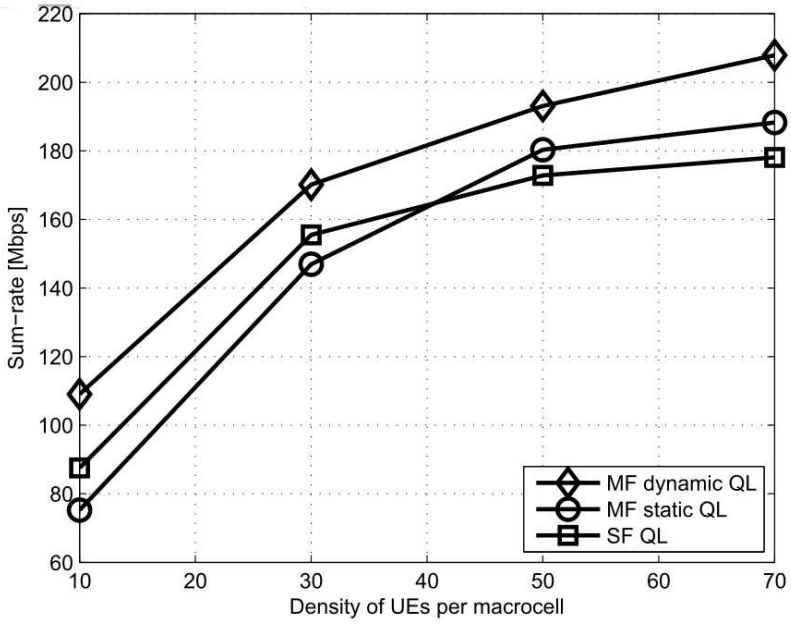


Figure 70. Sum-rate versus the number of UEs per macrocell in the frequency domain.

5.2 Distributed RRM and syncronization

5.2.1 Distributed RRM (power control) with unpredictable traffic arrival and mobility

We consider a wireless network model consisting of multiple transmitters and multiple receivers over an interference channel. All the transmitters and receivers are operating in the same frequency band thereby incurring interference on each other. Let $a_{k,j}(t)$ denote the traffic arriving in the queue of transmitter j at time t and intended to receiver k. Let $x_{k,j}(t)$ denote the data transmitted to receiver k at time t from the queue of transmitter j. The queue update equation for receiver k connected to transmitter j can be written as

$$Q_{k,j}(t+1) = \max\big(Q_{k,j}(t) - x_{k,j}(t), 0\big) + a_{k,j}(t), \ \ \forall k,j \tag{97}$$

The service delay of each user (and queue backlog) depends on the deviation between the arrival and the departure rates i.e. $\Delta_{k,j} = \big\|a_{k,j}(t) - x_{k,j}(t)\big\|$. Our objective here is therefore to minimize the quantity $\Delta_{k,j}$ over a finite time horizon T (the connection duration is finite). Minimizing $\Delta_{k,j}, \forall k,j$ over finite time horizon T (especially for small T) is stronger than ensuring the queue stability when $T \to \infty$. Minimizing $\Delta_{k,j}$ over finite T is challenging since the probability distribution functions of arrival and departure rates are unpredictable (due to for example mobility and interference).

In order to achieve the aforementioned objectives, and make the power control as simple as possible, we model in [MAL13] the evolution of the SINR (in dB scale) as well as the power variation from one timeslot to another as linear dynamic equations. We then formulate an H infinity (H^∞) control problem with linear state space equations and quadratic cost. The H infinity framework is a robust solution in the sense that it minimizes the growth of the

aforementioned quadratic cost with respect to the norm of the noise whatever the noise. Here, we explain briefly the solution developed in [MAL13].

We artificially introduce a new auxiliary variable denoted by $\gamma^*_{k,n}(t)$. This variable is controlled by our proposed controller. Our objective is then to develop a control strategy such that,

- instantaneous SINR $\gamma_{k,j}(t)$ tracks $\gamma^*_{k,n}(t)$
- $\gamma^*_{k,n}(t)$ tracks the traffic arrival rates
- $\gamma^*_{k,n}(t)$ is feasible at time t.

As a first step in our modeling and to track the feasible values of the target $\gamma^*_{k,n}(t)$ we update it as,

$$\bar{\gamma}^*_{k,j}(t+1) = \bar{\gamma}^*_{k,j}(t) + \beta_{k,j}\left(\bar{\gamma}_{k,j}(t) - \bar{\gamma}^*_{k,j}(t)\right) \tag{98}$$

where $\bar{\gamma}^*_{k,n}(t)$ denotes the dB value of $\gamma^*_{k,n}(t)$ and $\beta_{k,j} > 0$ is a very small step size.

The second step in our modeling is to let $\gamma_{k,j}(t)$ to track $\gamma^*_{k,n}(t)$ through power control. The power control should be as simple as possible in order to be implemented in practice. This can be done as follows [SAY05],

$$\overline{P_{k,j}}(t+1) = \varepsilon_{k,j}\left(\bar{\gamma}^*_{k,j}(t) - \bar{\gamma}_{k,j}(t)\right) + \overline{P_{k,j}}(t) \tag{99}$$

Where $\overline{P_{k,j}}$ is the power in dB scale and $\varepsilon_{k,j} > 0$ is a small step size between 0 and 1. In [SAY05], the aforementioned power control is used with slight modifications and analyzed in presence of Gaussian disturbances (one can refer to [SAY05] for more details).

The aforementioned power update equation is unstable in stochastic environment with unknown disturbances and does not deal with the average delay requirements of dynamic traffic. Let $\omega 1_{k,j}(t)$ be the disturbance due to SINR temporal variations, after some algebraic manipulations we obtain the following SINR variation equation:

$$\bar{\gamma}_{k,j}(t+1) = \{1 - \varepsilon_{k,j}\}\bar{\gamma}_{k,j}(t) + \varepsilon_{k,j}\bar{\gamma}^*_{k,j}(t) + \omega 1_{k,j}(t) \tag{100}$$

where $\varepsilon_{k,j} > \beta_{k,j}$ is a small step size. A similar update equation of traffic arrivals can be obtained (see [MAL13] for more details). It is worth mentioning that the update equations mentioned above may make the system unstable and will not fullfil our objective. We therefore introduce a control sequence $\mathbf{u}_{k,j}(t) = \left[u^1_{k,j}(t) \quad u^2_{k,j}(t) \quad u^3_{k,j}(t)\right]$. Let $\bar{\gamma}^a_{k,j}(t)$ be the traffic arrival(transformed into a quantity in dB scale) and $\omega 2_{k,j}(t)$ be the traffic variation between times t and $t+1$. We obtain the following update equations in a vectorial form,

$$\mathbf{y}_{k,j}(t+1) = \mathbf{A}_{k,j}\mathbf{y}_{k,j}(t) + \mathbf{B}\mathbf{u}_{k,j}(t)\mathbf{D}\boldsymbol{\omega}_{k,j}(t) \tag{101}$$

where $\mathbf{y}_{k,j}(t) = [\bar{\gamma}_{k,j}(t) \quad \bar{\gamma}^*_{k,j}(t) \quad \bar{\gamma}^a_{k,j}(t)]^T$ and $\boldsymbol{\omega}_{k,j}(t) = [\omega 1_{k,j}(t) \quad 0 \quad \omega 2_{k,j}(t)]^T$, $\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $\mathbf{A}_{k,j} = \begin{bmatrix} 1-\varepsilon_{k,j} & \varepsilon_{k,j} & 0 \\ \beta_{k,j} & 1-\beta_{k,j} & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

We then seek a control sequence $\{\mathbf{u}_{k,j}\}$ which minimizes the following cost function subject to the above state equation.

$$\min_{u_{k,j}} \max_{\omega_{j,k}} E\left\{\sum_{t=1}^{T}\left(\mathbf{L}_{k,j}^{t}\left(\mathbf{y}_{k,j},\mathbf{u}_{k,j}\right)-\pi^2 P\boldsymbol{\omega}_{k,j}(t)P^2\right)\right\} \quad (102)$$

where, $\boldsymbol{L}_{k,j}^{t}\left(\mathbf{y}_{k,j},\mathbf{u}_{k,j}\right)=\left\|\mathbf{y}_{k,j}(t)\right\|_{Q}^{2}+\left\|\mathbf{u}_{k,j}(t)\right\|^{2}$ and variable π^2 is the level of attenuation. For more details on the above cost function, one can refer to [MAL13][Hinf91].

Using H^{∞} control theory [Hinf91], we then develop to the following robust and distributed power control algorithm,

- Each transmitter allocates its own power using $\overline{P_{k,j}}(t+1)=\varepsilon_{k,j}\left(\bar{\gamma}_{k,j}^{*}(t)-\bar{\gamma}_{k,j}(t)\right)+\overline{P_{k,j}}(t)+u_{k,j}^{1}(t)$
- Each transmitter updates $\bar{\gamma}_{k,j}^{*}$ using $\bar{\gamma}_{k,j}^{*}(t+1)=\bar{\gamma}_{k,j}^{*}(t)+\beta_{k,j}\left(\bar{\gamma}_{k,j}(t)-\bar{\gamma}_{k,j}^{*}(t)\right)+u_{k,j}^{2}(t)$
- Vector $\mathbf{u}_{k,j}(t)$ is updated using $\mathbf{u}_{k,j}(t)=-\mathbf{B}^{T}\mathbf{M}_{k,j}(t+1)\left(\bar{\mathbf{M}}_{k,j}(t+1)\right)^{-1}\mathbf{A}_{k,j}\mathbf{y}_{k,j}(t)$

Where $\bar{\mathbf{M}}_{k,j}(t+1)=\mathbf{I}+(\mathbf{B}\mathbf{B}^{T}-\pi^{-2}\mathbf{D}\mathbf{D}^{T})\mathbf{M}_{k,j}(t+1)$, $\mathbf{I}$ is the identity matrix and $\mathbf{M}_{k,j}(t)=\mathbf{Q}+\mathbf{A}_{k,j}^{T}\mathbf{M}_{k,j}(t+1)\left(\bar{\mathbf{M}}_{k,j}(t+1)\right)^{-1}\mathbf{A}_{k,j}$ subject to $\pi^{2}\mathbf{I}-\mathbf{D}^{T}\mathbf{M}_{k,j}(t+1)\mathbf{D}>0\ \ \forall t$ and $\mathbf{M}_{k,j}(T+1)=\mathbf{Q}$. For more details one can refer to [MAL13][Hinf91].

In order to assess the performance of our approach, the mean power consumed by both algorithms is depicted in Figure 71, and the evolved mean queue length is depicted in Figure 72 after 5000 time slots. These results are obtained for 25 transmitter-receiver pairs uniformly distributed in a circular region of radius 500 m. The channel is a frequency selective Rayleigh fading channel with exponential delay profile. A Poisson traffic arrival model is considered. The packet size is assumed to be 5 kbits in the simulations. We can see that our algorithm outperforms the opportunistic algorithm developed in [LE06]. The power consumed and the queue length are much less for our algorithm.

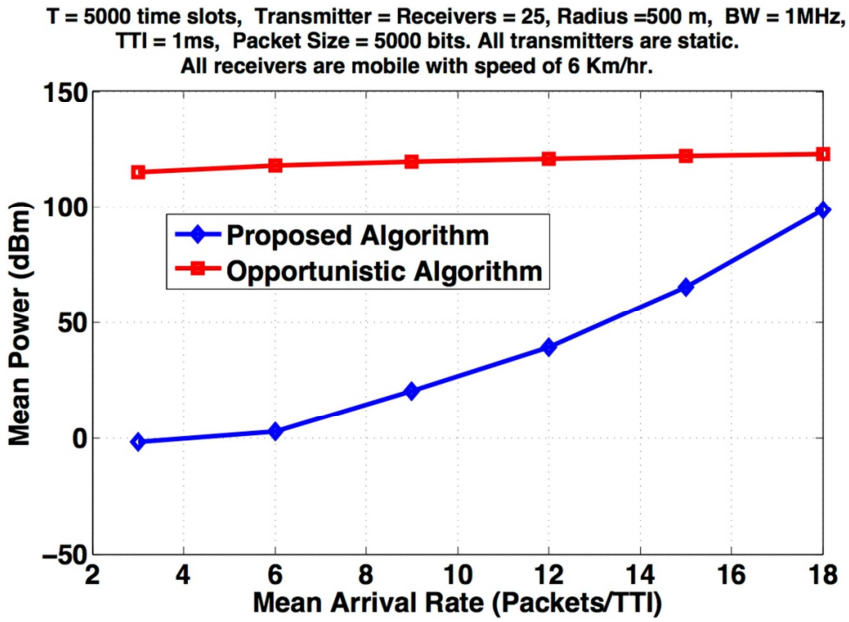


Figure 71. Mean Power (dBm)

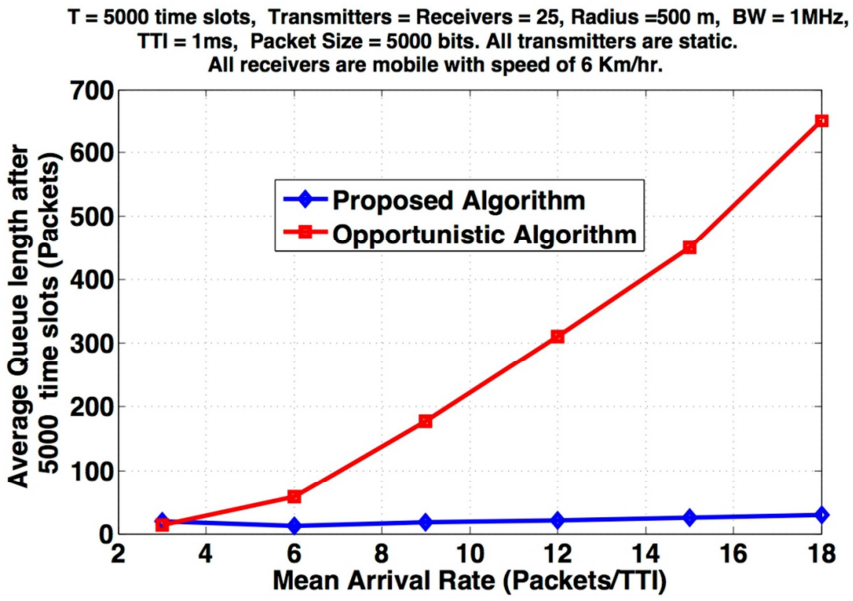


Figure 72. Average Queue Length

This innovation corresponds to scenario 2.9.10 in [D.2.2].

5.2.2 Distributed RRM with no information exchange between transmitters

In this section, we provide a brief description of the work in [MAC13] on distributed RRM with the assumption that the transmitters do not exchange information between each other. Although the framework is general and can be used in many applications, we consider here the context of resource allocation in heterogeneous networks where femto and macro cells co-exist and interfer with each others. Each femto has the objective to minimize its users' SINR outage and in parallel keep its impact on the macro users' outage as small as possible. Notice that here we consider the realistic assumption that, due to mobility and interference, the transmitters do not have a closed form expression of the SINR density functions. We provide here a fully distributed power control strategy that achieves a Nash equilibrium.

We consider a wireless network comprising two femto cells and one macro cell. We consider a simple setting of one user per cell. The macro access point and the macro user are represented by subscript M and m, respectively, and the femto access point and its user are represented by the subscript F and f, respectively. h_{Mm} represents the channel between the macro access point and the macro user, while $h_{F_j f_i}$ represents the channel between the femto access point j and the femto user i. $h_{F_j m}$ represents the interference channel between the femto access point j and the macro user. Finally, $h_{M f_i}$ represents the channel between the macro access point and the femto user i.

Let $g_{Xx} = |h_{Xx}|^2$ be the channel gain. The downlink SINR of macro user Γ_m is defined as

$$\Gamma_m = \frac{p_M g_{Mm}}{\sigma_m^2 + \sum_{i=1}^{2} p_{F_i} g_{F_i m}} \tag{103}$$

where σ_m^2 is the noise variance and $\sum_{i=1}^{2} p_{F_i} g_{F_i m}$ represents the interference caused by the femtocells 1 and 2.

The downlink SINR of femto user i, Γ_{f_i}, is defined as

$$\Gamma_{f_i} = \frac{p_{F_i} g_{F_i f_i}}{\sigma_{f_i}^2 + \sum_{j \neq i} p_{F_j} g_{F_j f_i} + p_M g_{M f_i}} \tag{104}$$

where $\sigma_{f_i}^2$ is the noise variance, $\sum_{j \neq i} p_{F_j} g_{F_j f_i}$ is the inter-femto interference and $p_M g_{M f_i}$ represents the interference coming from the macro access point. In addition, we assume that the macro and femto access points have imperfect knowledge of their own users' SINRs.

We then consider the following problem where the macro user is given priority while maintaining a reasonable quality of service (QoS) for the femto users. Let R_m^k and $R_{f_i}^k$ be the rewards of the macro and the femto respectively. We formulate the following game where macro and femto have the objective to maximize their own reward independently.

$$\begin{aligned} &\max R_m^k = \alpha \boldsymbol{E}_{\gamma_m}[\boldsymbol{P}(\Gamma_m^k \geq \gamma_m)] \\ &\max R_{f_i}^k = \beta \boldsymbol{E}_{\gamma_{f_i}}\left[\boldsymbol{P}\left(\Gamma_{f_i}^k \geq \gamma_{f_i}\right)\right] - \alpha\left(1 - E_{\gamma_m}\left[\boldsymbol{P}\left(\Gamma_{f_i}^k \geq \gamma_{f_i}\right)\right]\right) \end{aligned} \tag{105}$$

where $\alpha > 0$ and $\beta > 0$ are design parameters which can be adjusted to find the desired balance in performance between the femto and the macros. Where $\mathrm{P}(\Gamma_m^k \geq \gamma_m)$ and $\mathrm{P}\left(\Gamma_{f_i}^k \geq \gamma_{f_i}\right)$ are given by,

$$\mathrm{P}\left(\Gamma_{f_i}^k \geq \gamma_{f_i} \middle| D_x^{k-1}\right) = \int_{\gamma_{f_i}}^{\infty} \Phi\left(\Gamma_{f_i}^k \middle| D_x^{k-1}\right) d\Gamma_{f_i}^k \tag{106}$$

$$\mathrm{P}(\Gamma_m^k \geq \gamma_m | D_x^{k-1}) = \int_{\gamma_m}^{\infty} \Phi(\Gamma_m^k | D_x^{k-1})\, d\Gamma_m^k \tag{107}$$

To evaluate the above probabilities we need to track the densities of $\Phi\left(\Gamma_{f_i}^k \middle| D_x^{k-1}\right)$. In [MAC13], we provide a distributed bayesian density tracking that allows computing a numerical value of the aforementioned rewards. One can refer to [MAC13][SAL93] for more details.

A solution to the above problem is a state-independent equilibrium in the sense that no node has incentive to change its action when the other nodes keep their choice and the equilibrium strategy does not depend on the state. It may depend on the entire distribution of states. In [MAC13], we develop a distributed algorithm that converges to the Nash equilibrium. Notice that the game here is based only on numerical observation of the rewards (non-model based game). The algorithm is as follows.

At each time instant k, each transmitter updates its power p_X^k, by adding the sinus perturbation to the intermediary variable $\hat{p}_X^k$ (using the first equation below), and makes the transmission using p_X^k. Then, each transmitter gets a realization of the reward R_x^{k+1} from its corresponding receiver at time $k+1$ which is used to compute $\hat{p}_X^{k+1}$ using the second equation below.

$$p_X^k = \hat{p}_X^k + a_X \sin\left(\Omega_X \hat{k} + \phi_X\right) \tag{108}$$

$$\hat{p}_X^{k+1} = \hat{p}_X^k + \lambda_k l_X a_X \sin\left(\Omega_X \hat{k} + \varphi_X\right) R_x \tag{109}$$

where $\hat{k} := \sum_{k'=1}^{k} \lambda_{k'}$. $\Omega_X \neq \Omega_{X''}, \Omega_{X''} + \Omega_X \neq \Omega_{X'}$ and $\Omega_{X''}\ \forall i$ are small (e.g. $\lambda_k = \frac{1}{k+1}$). $\phi_X \in [0,2\pi]\ \forall X \in \{M, F_1, F_2\}$, $k \in Z$.

In [MAC13], we provide numerical results to assess the performance of our framework. The macro cell radius is equal to 500 m whereas the radius of each femto cell is 25 m. We consider a frequency selective Rayleigh fading channel with exponential delay profile. The power spectral density of noise is -174 dBm/Hz. We compare our framework to the standard power control scheme [YAT95]. Figure 73 shows that with our proposed scheme, femto and macro cells are able to maintain a success probability of 65 – 75% which is much better than the standard power control method [YAT95]. Figure 74 shows the Pareto boundary for our setting generated by using all possible transmit power scenarios by all the access points. The circle in this figure represents our operating region which is quite close to the boundary.

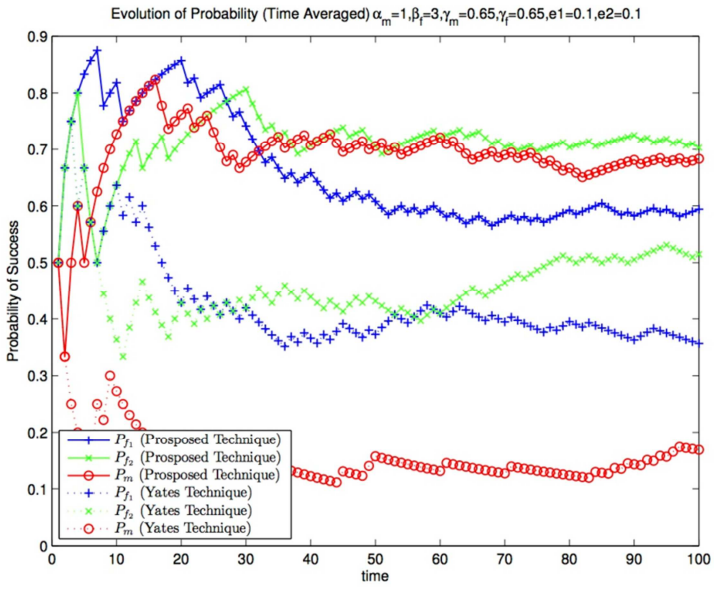


Figure 73. Evolution of success probability.

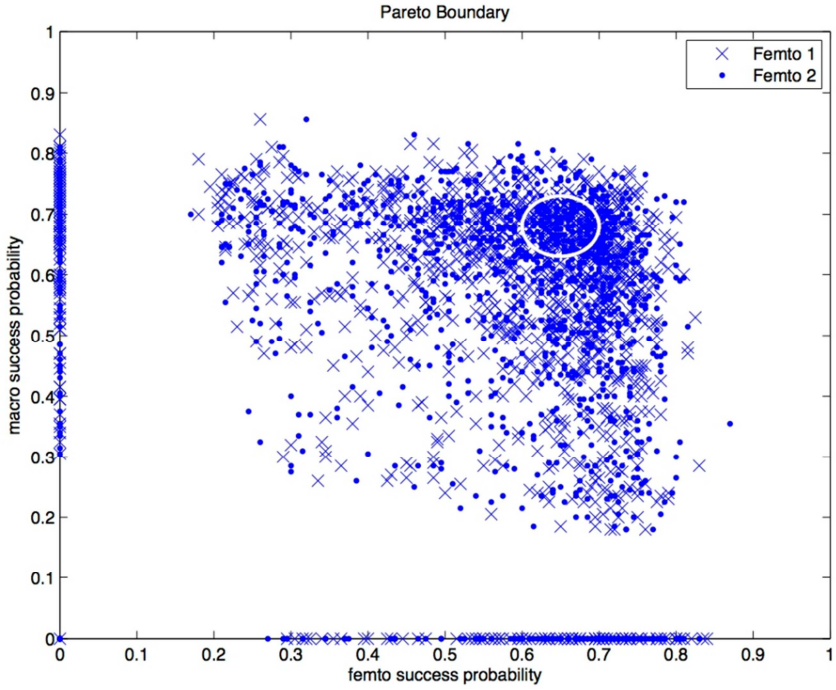


Figure 74. Pareto boundary.

This innovation corresponds to scenario 2.9.10 in [D.2.2].

5.2.3 Distributed synchronization algorithm based on over-the-air signalling transmission for heterogeneous network deployments

In this study we propose to discuss and evaluate the performance of distributed synchronization algorithms based on over-the–air signalling for heterogeneous networks. The basic scenario considered in the study is a scenario where a group of home base stations (HeNBs) is deployed within the coverage area of a macro base station network. The group of base stations may be coordinated and deployed in campus, i.e. the deployment is coordinated through a coordination GW or randomly deployed in the macro coverage region. Each home base station has a limited number of neighboring home base stations as shown in Figure 75.

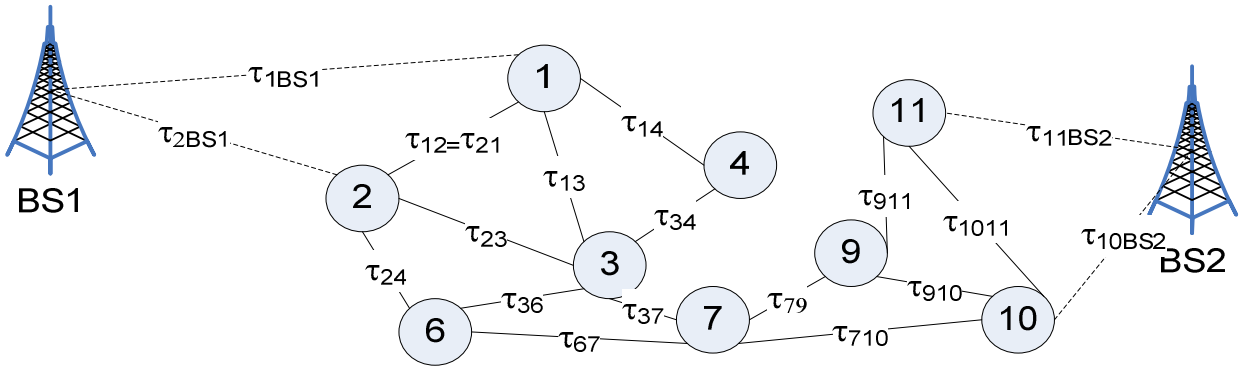


Figure 75. Description of distributed synchronization.

The objective of the study is to study algorithms for providing good time synchronization, i.e. synchronization of the subframe starting times for the home base stations of the group of the HeNBs (numbered in Figure 75).

What is intended in this study is to provide the distributed synchronization algorithm that is robust to unknown propagation delays between the home base stations and between the macro base stations and home base stations without considering specific scheduling of the transmission/reception times between the nodes as in gossip based algorithms [Boy06] The proposed distributed synchronization is based on over-the-air signalling between the home base stations. The algorithm starts with coarse synchronization of around 1 ms that is provided by neighboring macro base stations. Then, the algorithm refines the synchronization until a synchronization precision of 1 micro second is achieved.

The proposed algorithm is based on combination of consensus averaging and random broadcast were the clock drifts are corrected by the reference clock of the neighbouring macro base station. The consensus algorithm will allow the synchronization to be achieved very fast (around 100 iterations) and the random broadcast feature is essentially used to avoid specific HeNB-to-HeNB synchronization protocols.

In future studies we will provide the details of the synchronization algorithm and provide initial performance evaluation both in terms of system level simulations, complexity and robustness to time varying campus topology.

6 HETEROGENEOUS NETWORK DEPLOYMENTS – PERFORMANCE AND STRATEGIES

6.1 Fundamental performance limits of heterogeneous networks

The objective of this work is to investigate and analyze the performance of heterogeneous cellular networks as a means to quantify the improvement in terms of coverage and rate by using such topologies. This will provide concrete measures on how much (and using which design/technologies) HetNets can give an answer to the data traffic "tsunami". The main challenges in performing this study are the following:

- The spatial modeling of the base station locations in a HetNet remains an open topic, since little real-world information is yet available about picocell or femtocell deployments. It seems that the standard grid-based model that has been used in the past is not scalable to an accurate model of a multitier HetNet. A first model, which is very tractable and allows us to derive closed-form expressions for key performance metrics, is a statistical one in which the base stations are located according to a Poisson point process in a two-dimensional plane. Such a spatial distribution for base station locations corresponds to complete randomness, and a large class of powerful results and analytical tools are available from the field of stochastic geometry. The Poisson model will be the starting point, however our performance analysis should consider other spatial models and random distributions that introduce a certain level of correlation and dependence among the base station placements.
- The issue of cell association should be taken into account and revisited as compared to traditional cellular networks. Although in existing macrocell networks, the max-power coverage regions for each base station are designed to have roughly the same amount of traffic, i.e. more base stations are deployed in areas that generate more traffic; a meaningful cell association policy should assign users to base stations that offer them the best user-perceived rate. However, optimizing the rate for all users is very complex as the rates of the users are coupled. Several relaxations and considerations should be made to be able to analyze the performance of HetNets with efficient cell association and realistic load balancing.
- A very important issue that is often neglected in the analyses is that of backhaul connection. It is often assumed that the main design challenge is the base station-mobile link and that the base station has a high-speed backhaul connection that easily handles the data flowing to and from it. In this task, the effect of backhaul connectivity and quality/robustness will be taken into account in the performance analysis of HetNets. Issues related to mobility, type of connection (wired vs. wireless) and propagation environments will play a critical role in the analysis.

This task is related to scenario 2.9.6 in [D.2.2].

6.2 Asymptotic performance limits under heavy traffic

Here, the objective is to analyze the performance of heterogeneous networks in terms of throughput and queueing stability taking into account the traffic pattern. Most of existing work in heterogeneous networks (e.g. resource allocation, performance analysis, physical layer design) focuses on througphut and power consumption and tends to ignore the traffic patterns. More precisely, we will study the performance of some advanced MIMO schemes as well as advanced resource allocation policies in heterogeneous networks and characterize the asymptotic performance limits of the network when the traffic arrival is dynamic. We will use heavy traffic asymptotic modelling of the network to study the underlying control problem. It turns out that the models in this asymptotic case become more tractable to handle and their study can reveal some fundamental trade-offs and useful information for the system's behaviour even when the system is in moderate or light traffic regime. By heavy traffic, we

mean that the arrival rate is very close to the boundary of the capacity region. It is worth noting that delay grows to infinity if the arrivals are pushed toward the boundary of the rate region. The objective is then to characterize the performance of the network up to a certain growth coefficient. The challenge here is to include the buffer occupancy dynamics in the analysis and show the existence of limit processes that characterize asymptotically the behaviour of the network. The stability (and even delay bounds) can be then investigated in this case. Based on the performance analysis, new traffic-aware resource allocation strategies and MIMO precoding schemes can be developed as well.

This section is related to scenario 2.9.8 in [D.2.2].

6.3 Ellispoid techniques for coverage control in HetNet campus scenario

In this study we will discuss coverage control techniques for the campus of small cells as defined in our previous contributions as a group of heterogeneous base stations that are deployed in uncoordinated manner in a given coverage region under a GW coordination.

The coverage in networks is in general dependent on the overall system parameters, such as available PRBs, scheduling algorithms used by the nodes, service type and transmission powers. Coverage control optimization problem can be viewed as the problem of optimizing the worst possible coverage when considering enough PRBs and minimum possible service in each cell of the network. In this case, the coverage control problem is viewed as the optimization of the cell-edge SINRs in the network so that the minimum service is possible in each cell. In this work we will discuss the application of the ellipsoid method, i.e. multidimensional bisection search methods. The studies presented in this section are initial thoughts about the application of ellipsoid techniques for the solution of coverage control problem and will be further refined within future work of SHARING WP4.

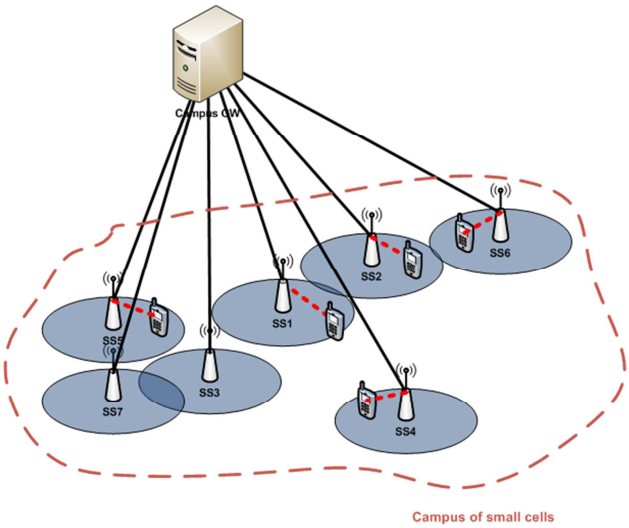


Figure 76. Campus deployment of small cells

Consider a campus of N small cells coordinated through campus GW as shown in the Figure 76. The normalized cell-edge signal to interference ratio of the small cell i is given by the following relation:

$$SINR(i) = \frac{\bar{\alpha}_{ii} P_i}{\sum_{j \neq i} \bar{\alpha}_{ij} P_j + 1} \tag{110}$$

where the parameter $\bar{\alpha}_{ij}$ is the normalized path gain with respect to the noise power N_0 evaluated in the cell-edge of the cell i. P_i is the downlink radiated power of the base station i. The coverage control problem may be formulated as follows: Find the vector of the radiated powers $[P_1, \dots, P_N]$ such that the cell-edge SINR for all the base stations is above a predefined threshold γ. So, the coverage control optimization problem may be formulated in the following way:

$$\begin{aligned} &\max \gamma \\ &st \;\; \bar{\alpha}_{ii}P_i - \gamma \sum_{j \neq i} \bar{\alpha}_{ij}P_j \geq \gamma, \;\; i = 1, \dots, N \end{aligned} \tag{111}$$

The constraints of the optimization problem can be further grouped into a matrix equation that is dependent from the normalized path gains and cell edge SINR threshold, given as:

$$A(\gamma)[P_1, \dots, P_N]^T \succ \gamma[1, \dots, 1]^T \tag{112}$$

The powers are the optimization variables of the problem and are defined to be in the interval (P_{min}, P_{max}). The solution of the optimization problem is to find a feasible vector of powers such that the SINR threshold is maximized. The bisection search is a very simple method to find the optimum SINR threshold γ^* as well as the optimum power vector that are solution to the coverage control optimization.

The steps of the bisection search are given by the following:

1) Define initial search interval for the SINR threshold $\gamma \in (U, L)$.
2) Calculate the bisection of the interval as $\gamma_0 = \frac{U-L}{2}$.
3) Test for the feasibility of the constraints in γ_0.
   a) If the constraints are feasible, then $L = \gamma_0$.
   b) If the constraints are not feasible, then $U = \gamma_0$.
4) If the interval length $|U - L|$ is below a precision ε, then break, otherwise the step 1 is repeated.

Key ingredient of the bisection search is to test for the feasibility of the constraints linear matrix inequality and this feasibility test is performed by the means of the ellipsoid technique.

The ellispod technique starts by defining a hyperball around the multidimensional search domain $(P_{min}, P_{max})^N$. This hyperball is defined by the matrix G such that all the vectors of the ellipsoid are fulfilling the relation $\underline{P}^T G^{-1} \underline{P} \leq 1$, where the notation $\underline{P}$ stands for the vector of the powers in the campus. The initial input vector of the ellipsoid method is the center of this ball, denoted as $\underline{P} = \underline{P}^0$.

The steps of the ellipsoid method are the following:

1) The inequality constaints are tested for the power vector $\underline{P}$ and a cutting plane is found, i.e. a row of the matrix $A(\gamma)$ such that the constraints are not fullfield, i.e. $a(\gamma)^T \underline{P} < \gamma \underline{1}$.
2) This cutting plane defines two regions of the ball and the most likely region to achieve the constraints is selected and the search ellipsoid is updated in this region as the following

$$\underline{P} = \underline{P} - \frac{1}{N+1} G \underline{a}(\gamma) \tag{113}$$

$$G = \frac{N^2}{N^2 - 1}\left(G - \frac{2}{N+1}\right) G \underline{a}(\gamma) \underline{a}(\gamma)^T G \tag{114}$$

3) The step (1) is iterated until the volume of the ellipsoid is below a predefined threshold.

The output of the described ellipsoid algorithm consists of the power vectors that fulfill the constraints of the optimization problem for a given value of the SINR threshold γ that is fixed by bisection search.

In order to illustrate the performance of the ellipsoid technique in terms of the optimization of the cell-edge SINR, we have simulated the coverage of a campus of 16 open HeNBs deployed indoor and taken the CDF of the cell-edge SINRs as the main performance measure. The obtained results are shown in Figure 77.

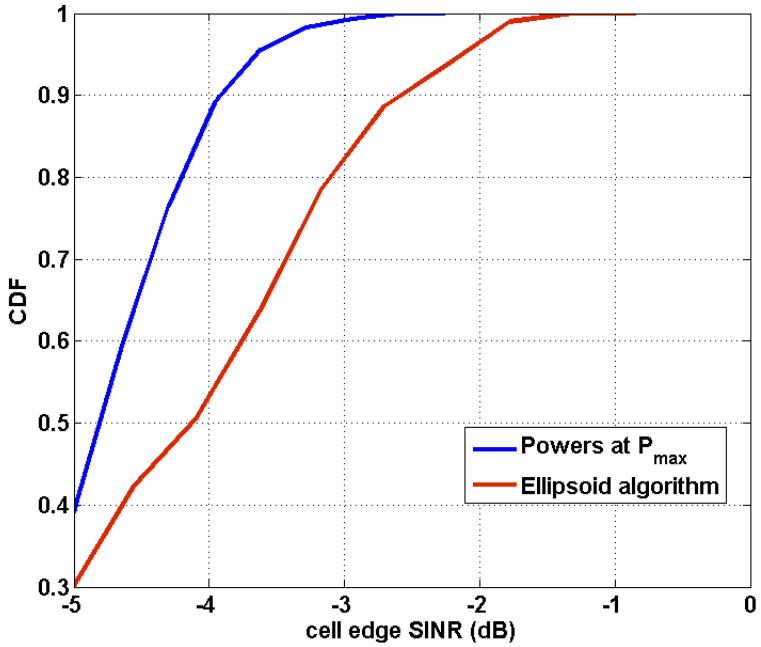


Figure 77. Initial evaluation of the ellipsoid algorithm.

It is seen that ellipsoid technique improves the overall cell-edge SINR performance of around 13%. However, the convergence of the ellipsoid method is very low which limits the applicability of the technique to low mobile user terminal.

In future studies we will investigate ways to improve the ellipsoid technique both from performance and convergence point of view.

This innovation is related to scenario 2.2.2 described in [D.2.2].

6.4 Distributed techniques for coverage control in HetNet campus scenario

In this study we will propose a distributed solution to the coverage control problem described in Section 6.3. The proposed solution will improve the overall robustness of the coverage control technique by the means of over-the-air signalling between the available base stations in the campus.

The basic idea is to use consensus averaging to provide distribution of the coverage control and to investigate convergence and stability of the distributed technique.

This innovation is related to scenario 2.2.2 described in [D.2.2].

6.5 Offloading strategies for WiFi & small cells integration

6.5.1 Traffic offloading status

Traffic offloading can be defined as the process that enables traffic deriving from cellular mobile devices such as smartphones or tablets to be handled by non-cellular technologies such as (and essentially) WiFi with the aim of both releasing capacity on the cellular network and improving quality of service through better indoor coverage. As a result of this definition, a WiFi only tablet that would be connected to a carrier WiFi network would not be considered under our definition as offloaded traffic.

Offloading is usually done on WiFi hotspots or on femtocells connected on users' fixed broadband connection. In many ways, it presages heterogeneous networks with small cells deployed in densely populated areas and is acting as an offload mechanism for the macro network. The advantage of WiFi as of today is that it is operating on frequencies that do not interfere with the macro network.

According to the WIK, 74% of the traffic from smartphones and tablet was transported over WiFi in 2013 but this impressive figure probably includes tablets with no cellular connectivity as well as the fact that this offloaded traffic is the result of a choice of the users and not a seamless process where traffic would be automatically derived to WiFi in a congested situation. In other words, we are still far from the ideal vision where operators could dynamically manage this process. Also because the operator is (still) not managing the offload process, users often connect to WiFi networks which the operator has no control on.

In other words, a large share of mobile traffic is carried over WiFi, but with little control from the operators.

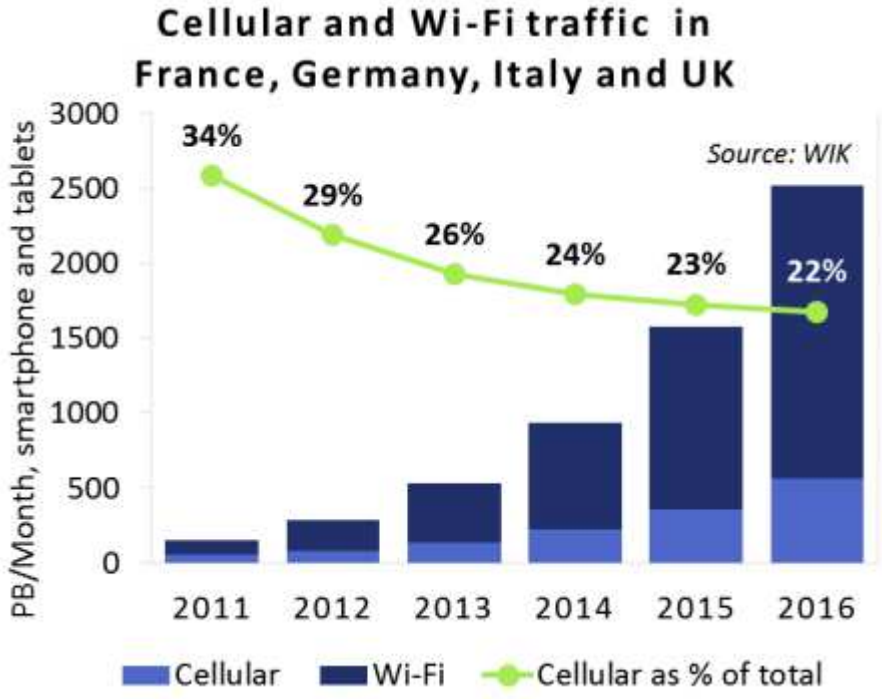


Figure 78. Cellular and WiFi traffic in France, Germany, Italy and UK.

Several degrees of offloading can be thus determined, depending on how much control the carrier has on both the quality of service on the offloaded network and on the decision itself to offload a user from the cellular network to a WiFi network:

- No control: At this stage, it is the user that decides to connect to a WiFi network to benefit from better throughput, coverage or simply save on its data plan. The destination network may be a homespot or hotspot. The operator has no control,

neither on the decision, nor on the quality of the connection on the destination network.

- Limited control: At this stage, the operator has WiFi hotspots which the subscribers can join with their own credentials. Users have to enter their credentials manually to connect to the hotspot. The decision to connect to a WiFi network comes from the user and is not initiated by the operator but the later has a relative control on the experience on its carrier WiFi hotspot.
- Some control: Here, the connection to a WiFi network is seamless for the user. Thanks to features such as Hotspot 2.0, the SIM card is used as an authentication tool and connection to the carrier WiFi network is seamless. The carrier has a control on the quality of service since it is operating the carrier WiFi. Because the connection is seamless, the user's device is more likely to offload its traffic.
- Total control: At this ultimate stage, connection to a carrier WiFi network is seamless to the users, as at previous stage but now the operator can devise policies to dynamically offload or onload the users depending on the available capacity on the cellular network. Thanks to ANDSF feature, the carrier WiFi is completely integrated to the core network and rules can be defined to jump off or on from a network to another.

Ultimately, devices will be able to switch WiFi on and off depending on the availability of WiFi signals and on the traffic conditions on the cellular network. This feature would be controlled by both the device and the network. For the users, it will bring a longer battery lifetime since WiFi radio won't be always on and for the network operator, it will ensure that WiFi offloading can be used when required by the network. There has been some work to enable automatic on and off switching of WiFi depending on geolocation and availability of known WiFi access point but those solutions do not take traffic condition on the cellular network in consideration.

6.5.2 Operators' strategies

What are the criterions for offloading someone to WiFi?

- Device capability: Which generation of WiFi does it support, which bands and how many antennas?
- What are the traffic conditions on both cellular and WiFi network?
- Is it a trusted network or not? Does the operator have control on the offloaded network?
- What is the status of the subscriber? Is it a proprietary subscriber? How would proprietary be defined? Depending on its premium status or not (monetization), depending on the kind of usage and its requirements (e.g. very small latency required for gaming or video conferencing for instance, or very high throughputs for Ultra HD streaming).
- What are the usages and what kind of QoS does it require?
- How could/should the offloading be monetized?
- How far are the device and its user from the small cell? In other words, which radio technology is more power efficient for a given distance to the small cell?

In terms of WiFi offloading strategy, several elements are to be taken into account by the operator:

- Do I deploy my own WiFi network or do I rely on roaming agreements? In the future, operators with small cell strategy will increasingly co-locate WiFi and cellular

connectivity, partly eliminating the question. Outside the country of operation, however, the question will remain, with lessening of roaming fees as an argument for developing its own network.

- How do I define network selection policy? In a small cell environment, WiFi is more suited for devices that are the closest to the small cell while LTE will be better suited for people farther away from the small cell.

With LTE offering similar throughputs to WiFi network, the rules where a device automatically connects to a WiFi network when available are no longer desirable.

6.5.2.1 AT&T

AT&T sees Hotspot 2.0 as an enabler for new WiFi services and as an important part of its small cell strategy. Its multimode small cells should as a result embed WiFi in addition to 3G/4G. Currently, AT&T has deployed Distributed Antenna Systems (DAS) in public venues as well as carrier WiFi. Both systems have separate architecture. The small cells that have been deployed are currently UMTS and HSPA+ but will later on support 4G and WiFi together.

In terms of strategy, AT&T deploys small cells in venues which are lower in capacity and DAS will remain in larger venues. This is because current small cells only support a single technology, single frequency allocation and a single sector. As a result those current small cells are not currently designed for high capacity environments.

First small cells were deployed by AT&T during the first quarter of 2013 and first 4G LTE small cells will be deployed by the end of 2014. In the case of venues such as enterprises where WiFi has already been deployed in the 2.4 GHz band, AT&T would probably install small cells because the 2.4 GHz band only has 3 non-overlapping channels so deploying another WiFi system would deteriorate the performance of the initial system.

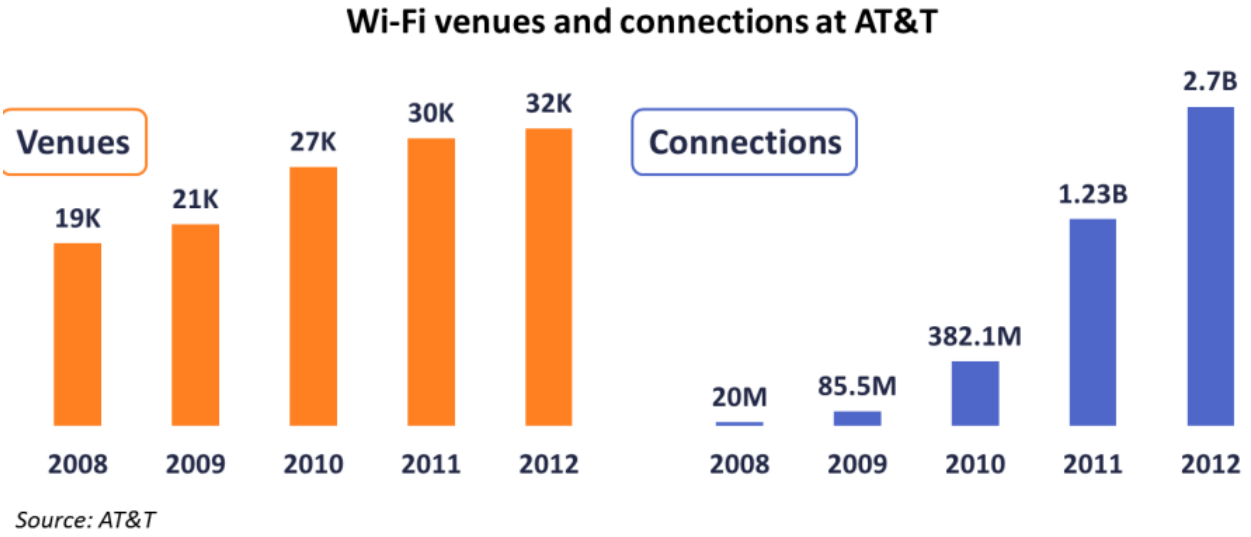


Figure 79. WiFi venues and connections at AT&T.

It is important to highlight that AT&T has a specific division called Antenna Solution Group which is focused on deploying equipment and antennas on the field. They work with venues whereby they propose the installation of equipment to support more traffic, something that they later on sell to operators. As a result it is critical for them to support multiple radio access technologies and frequency bands in the future because venues do not want to be able to propose such services to a single wireless operator. When the group started to deploy DAS

on the field, it was critical to support multi-operator because of the upfront cost of such solutions.

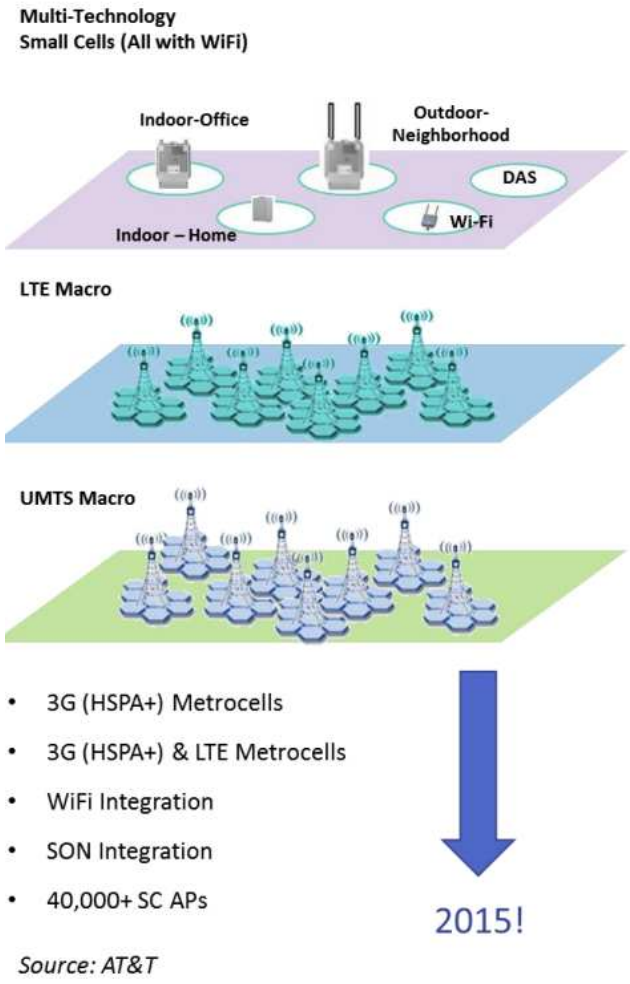


Figure 80. AT&T's vision of multi-technology small cells.

6.5.2.2 China Mobile

In China, WiFi is not suitable for outdoor environment because of the regulation that imposes a power limit at 500 mW. Small cells from China Mobile will support both 3G and 4G and will use different frequency bands than the macro network. For the 4G small cells, the 2.3 GHz band (band 40) will be used.

In order to make the WiFi offloading more seamless and effective, China Mobile will deploy ANDSF when it will be ready, which will require updates both on the infrastructure and on the devices. ANDSF can be implemented either in the application layer or in the OS at a lower layer.

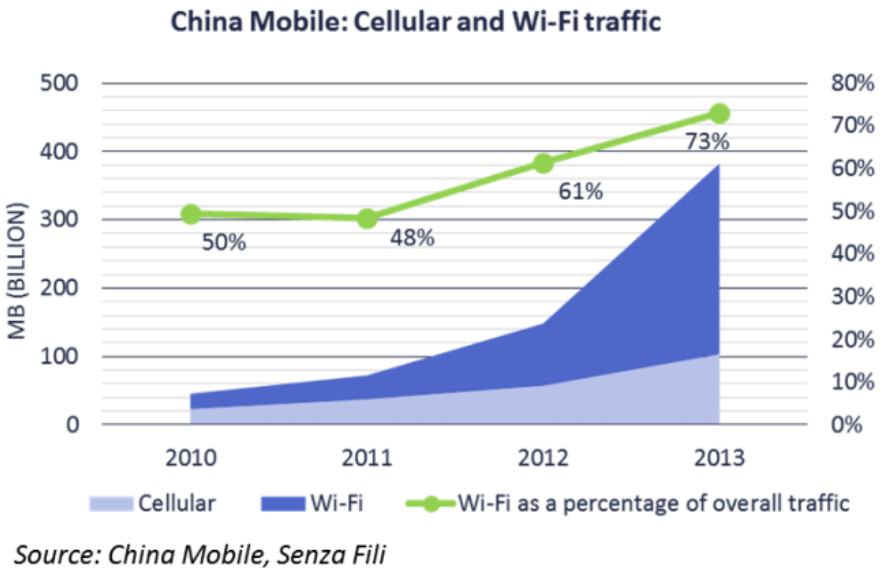


Figure 81. China Mobile: Cellular and WiFi traffic.

Figure 82. NGH. Implementation by China Mobile.

WiFi has always been a fundamental piece of the mobile strategy of China Mobile. During the MWC 2011 in Barcelona, the CEO of China Mobile declared that mobile devices including smartphones and tablets should be set by default on WiFi, since data consumption by these handheld devices were overloading cellular networks.

While revenue from WiFi data traffic represented only 2.5% of its overall mobile data traffic revenues in 2012, the operator set an ambitious plan to considerably enlarge its WiFi network, deciding to roll out 1.41 million units in 2012.

Indications have been that the Chinese carrier was able to meet this objective reaching a total of 3.83 million WiFi hotspots across China at the end of 2012.

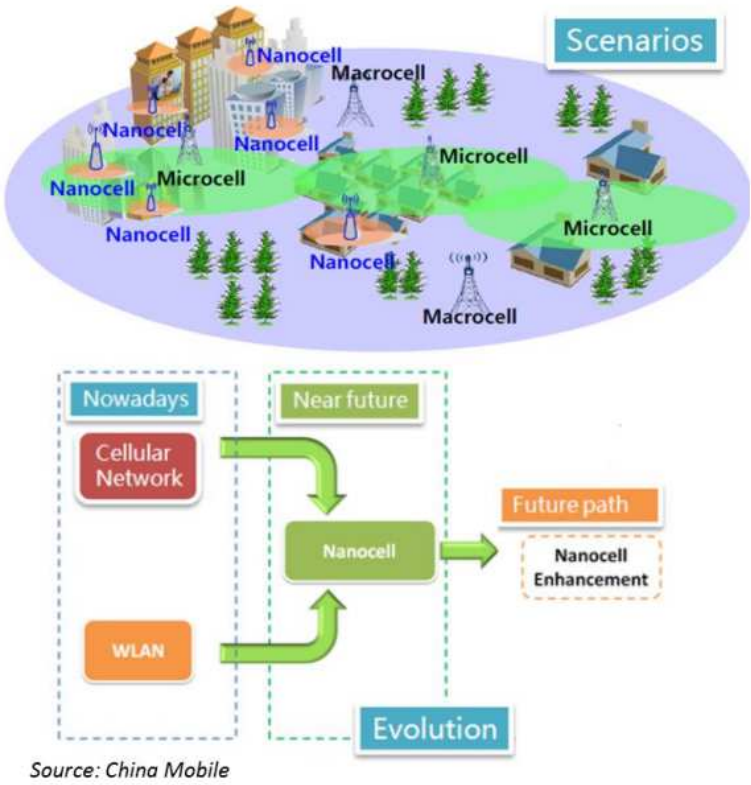


Figure 83. Towards nano cells – China Mobile.

6.5.2.3 KDDI

KDDI is the second largest operator in Japan with 39 million subscribers as of September 2013 and a network whose number of base stations is estimated to be superior to 100,000. KDDI has deployed around 220,000 hotspots complemented by 1.65 million residential WiFi units.

It is currently offloading a sizable share of its mobile data traffic with WiFi and Wimax, since 43% of mobile data traffic was offloaded to the WiFi network deployed by Ruckus Wireless during busy hours. Offloading is supported through a specific solution implemented on Android devices. It is not clear whether and how iPhones are supported in that respect.

It is interesting to note here that Wimax is also being used as a means of offloading traffic. With the launch of its Wimax 2 network, which is compatible with TD-LTE, the operator has devices capable of supporting Wimax in addition to TD-LTE and of course WiFi as well.

Figure 84. Share of mobile data traffic offloaded on WiFi (& WiMAX) (Source: Tefficient based on operator quarterly report).

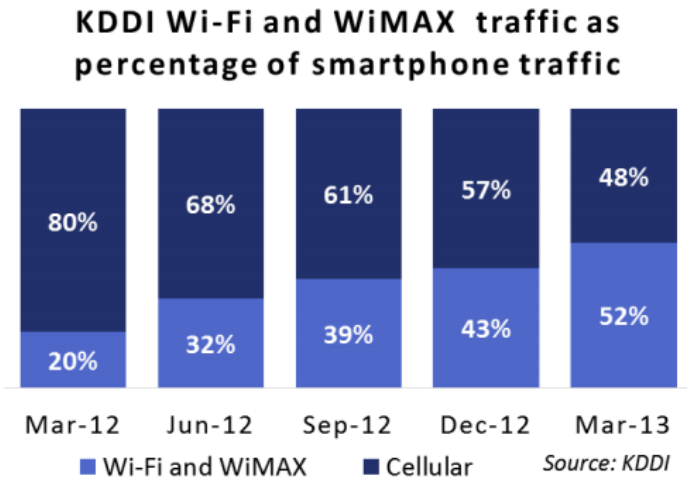


Figure 85. KDDI WiFi and WiMAX traffic as percentage of smartphone traffic.

6.5.2.4 BT (UK)

In May 2013, BT announced that its WiFi network had reached the number of five million hotspots in the UK, thus crossing the cap of eight million hotspots worldwide. The operator also gave a breakdown of the distribution of its WiFi hotspots over the UK: 764,000 in the South East, 464,000 in Scotland, 519,000 in the East, 501,000 in the South West and over 543,000 in London. The British carrier also reported that it was expanding its network by 20,000 hotspots per week. According to BT, its WiFi network supported more than 400 million connections during the year between April 2012 and March 2013. During this period more than 13 PB (PetaBytes) of data were carried over the infrastructure.

Although a fixed operator, WiFi connectivity provided by BT is clearly aimed at smartphone users. BT's WiFi homespot started to grow when the operator started to make the community membership an opt-out option rather than an opt-in option.

Figure 86. BT WiFi coverage in the UK (blue colour: areas with dense coverage) (Source: BT)

6.5.2.5 O2 (UK)

In April 2012, O2 announced that it inked a five-year agreement with McDonald's restaurants. The contract covered equipping, and management, of the 1,200 UK McDonald's restaurants with WiFi hotspots before the beginning of the Olympic Games. The whole WiFi network would have to support the traffic generated by an average of 750,000 connections per month.

Just a few weeks before, in March 2012, O2 reported that it will also equip the 1,600 restaurants belonging to the food firm Mitchells & Butlers.

6.5.2.6 (Germany)

In March 2013, the German operator stated that it had started offering its consumers the possibility to connect to eight million hotspots worldwide via WLAN TO GO service. This is based on a partnership with Fon, the largest WiFi provider in the world.

With this new service, subscribers who share their WiFi connection at home are assured access to millions of hotspots over the world for free.

Continuing their statements on WiFi strategy, the telco announced that in order to face the data surge of the years ahead, users will benefit from 2.5 million hotspots in Germany by 2016 with WLAN TO GO service.

Currently, 3000 km of German railroads are covered with WiFi hotspots and the carrier expects to increase this number to 5200 km by the end of 2014.

6.5.2.7 Orange (France)

Orange was reportedly involved in the development of Passpoint specification and is said to have roaming agreements with US operator AT&T. Orange's WiFi network consists of 40,000 hotspots covering 10,000 towns in France and the operator has a roaming agreement nationally with SFR, Bouygues, Hub Telecom in airports and Naxos. In 2011, the operator highlighted that WiFi offload was necessary to the company but that a seamless way of connecting to WiFi network was required for the offload to be effective. According to Orange, this requires;

- Strong authentication methods
- IP Sec must be implemented
- Orange would like to use the 5GHz band for offload to be more effective but most smartphones only support the 2.4 GHz band

For Orange, WiFi hotspots are both B2B (Business-to-Business) services and a way to offload the cellular network and bring improved capacity in specific areas with high density of population of users. The main challenge for Orange is to mix the use of both licensed and unlicensed spectrum together, with two very different business models.

6.6 Performance of heterogeneous network deployments in a large-scale real environment

With the aim of assessing the network performance, SIRADEL has developed a LTE simulator [Bra12] based on site-specific radio propagation predictions which are obtained with its highly efficient ray-based model Volcano [Cor09], and highly realistic multi-cell, multi-user scenarios. SIRADEL will elaborate dynamic scenarios as a basis for HetNet network evaluation, relying on a real environment (e.g. an existing urban area represented by 3D buildings), including non-uniform, multi-service and multi-environment user traffic, multi-cell

and multi-layer network deployments, site-specific path-loss models, advanced channel models and prediction of the wireless backhaul links. Figure 87 shows an example of a multi-layer deployment in a real dense urban environment. Requirements to be addressed by SIRADEL in the frame of the SHARING project for a realistic dynamic scenario are:

- Elaboration of realistic time-variant traffic maps
- Dispatching of multi-floor indoor users and outdoor users according to these traffic maps
- Assignment of various traffic demands to the users
- Assignment of adapted mobility model to the users (indoor, outdoor pedestrian, outdoor vehicular)
- Multi-cell multi-user 3D radio predictions
- Modelling and prediction of the wireless backhaul links
- Monte Carlo simulations for a fine radio resource allocation and interference modelling

These scenarios, channel predictions and traffic distributions are built within the framework of SHARING WP3. They will then feed the system-level simulators to evaluate the performance of innovative intra-RAT and inter-RAT offloading strategies and SON algorithms in terms of service coverage, network capacity, energy efficiency and fairness, thus providing a common evaluation framework for SHARING Tasks 4.1, 4.2, 4.3 and 4.4. In general, this work is related to scenarios 2.3.6, 2.3.7 and 2.4.1 defined in [D.2.2].

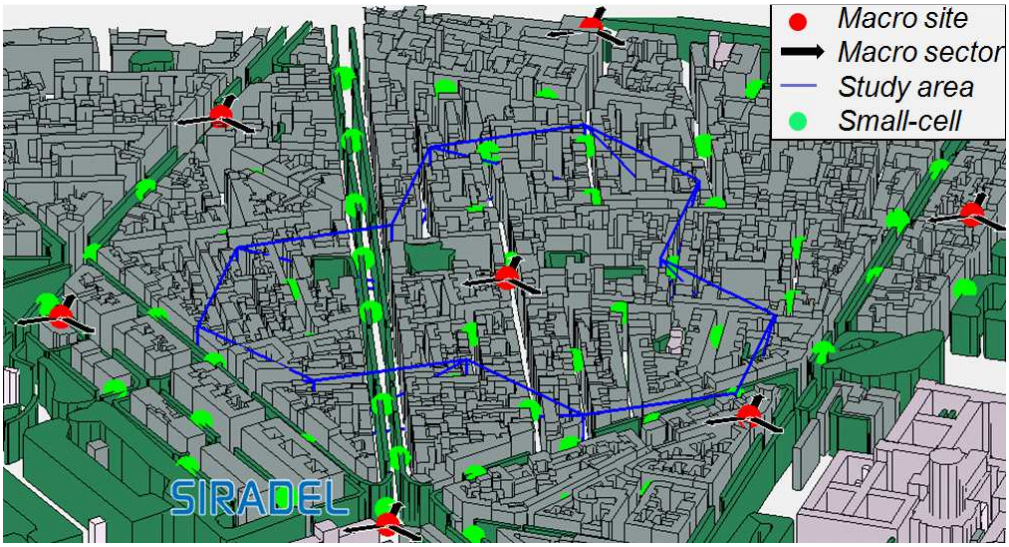


Figure 87. Example of a multi-layer deployment in a real dense urban environment.

Modelling and prediction of the wireless backhaul links:

Large-scale deployments of public-access small cells are expected in the next coming years to cope with the growing traffic demand resulting from a larger use of data-hungry smart phones. These low-height base stations (3 to 8 meters) are usually placed on urban furnitures such as light or utility poles or even on building facades. Their locations are usually defined *a priori* based on an existing macro network and traffic demand, i.e. they are added to offload the macro network and to increase capacity.

All small cells must be connected to the core network either by a wireline backhaul when available or through various wireless technologies involving Line-of-Sight (LOS) and Non-LOS at various frequencies (sub-6 GHz, microwave, millimeter wave). In case of wireless technology, network elements denoted as hubs are connected to multiple small cells and

ensure the link between the small cells and the core network. In macro networks the only technology that is used for wireless backhaul is the microwave backhaul, which requires a LOS link with a very good alignment of antennas. The way hubs are connected to the core network is very analogous to what exists for macro cells, but hubs should also be able to provide the backhaul to multiple small cells, most of the time in NLOS conditions due to the complexity of urban environment and the configurations of small cells (see example in Figure 88). Furthermore, the placement of backhaul hubs aims at providing enough signal power to their connected small cells but also to limit the interference levels on other links, both downlink and uplink. Therefore, their locations should be well dimensioned, not too dominant but still able to serve the small cell clusters.

Backhaul link must also have a high capacity and support typical throughputs of 50 to 150 Mbps in case of LTE small cells. There should be a special emphasis on capacity as these capacity demands are expected to increase especially with multi-mode small cells (e.g. HSPA, LTE and WiFi) [Col13]. This performance must be guaranteed because it directly impacts the QoS of end-users.

In this context, connecting the small cell layer to the existing network infrastructure brings new challenges to which current simulation tools are not adapted. SIRADEL LTE simulator [Bra12] is able to simulate the user data network but under ideal backhaul assumption. This can lead to incorrect small cell network deployment with overestimated performance. Assessing jointly the backhaul and the data networks is thus esential to estimate in a more realistic way heterogeneous network metrics and offloading. On the other hand, the backhaul design can be optimized to achieve the expected performance of the small cell layer.

To cope with wireless NLOS backhaul links, a site-specific 3D ray-based propagation model [Cor09] is necessary [Let14]. It relies on 3D geo-referenced models of the environment (building layout, altitude, etc.) and precise hub/small-cell locations to simulate the impact of shadow fading. The 3D ray-based model is able to calculate the lateral multipath contributions and to provide accurate signal variation and channel properties for all configurations [Let13], allowing for a fine orientation of the antenna towards the best propagation path. Figure 88 shows an example of a NLOS hub – small cell link prediction in which the direct path is not a dominant path.

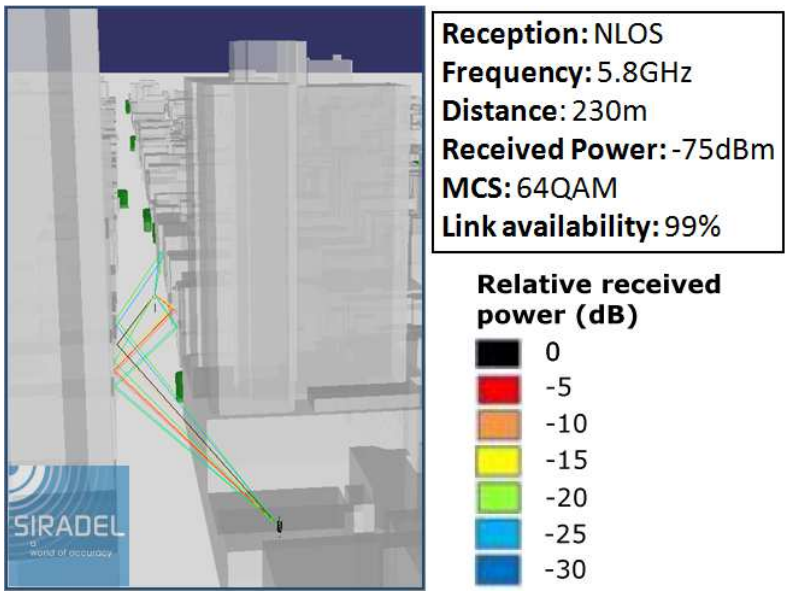


Figure 88. Typical NLOS Hub – Small cell radio link prediction in dense urban environment.

SIRADEL will develop a tool for the design and simulation of the wireless backhaul, including:

- Ray-tracing propagation model able to cope with a wide frequency range
- MIMO
- Interference modes (avg background, co-channel)
- Auto antenna orientation hub + small-cell
- Capacity estimations and traffic loads
- Margins (types, analytical/Monte-Carlo process to obtain statistics)

Figure 89 shows a possible display of this backhaul design tool. In complement with the LTE simulator, this tool will allow to simulate the wireless backhaul traffic and possible congestions in a realistic small-cell scenario, thus taking into account the backhaul capacity limitation in the offloading strategies and metrics.

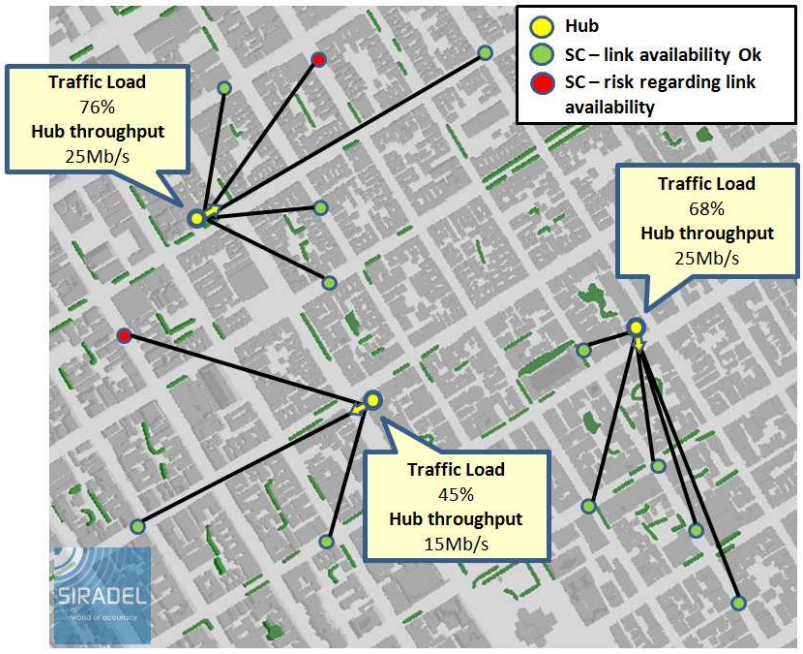


Figure 89. Sub-6 GHz backhaul network capacity analysis.

Traffic modelling:

As pointed out in the introduction of this deliverable, deploying low-power sites within traffic hotspots is considered as one way to increase the capacity of a mobile network. However, as the deployment of small cells is relative to traffic hotspot areas, it should not be conducted by the same rules as a macro cell network, i.e. an almost regular geographical scheme. Indeed, optimising the deployment requires an accurate representation of how traffic is distributed in space and time [Col12]. It is also essential at the simulation level to assess the two-tier network performance and offloading metrics, as well as energy saving mechanisms. Besides traffic estimation at a given time, data traffic forecasting also enables to evaluate the long-term viability of the network. Modelling traffic can be decomposed in two scales. On the large-scale, the traffic demand is described by its distribution in space and time. This description is useful to spot hot traffic areas at specific day time. On a smaller scale, i.e. when the large-

scale properties of the traffic are invariant, modelling the traffic aims at describing intrinsic details of data flows, which are related to the demanded wireless services. Depending on the desired level of abstraction, the simulation of heterogeneous networks will include only large-scale properties or both scales. Some examples are presented hereafter.

Network operators use radio planning tools to estimate the radio coverage, either in terms of covered area or covered population. In the latter case, traffic is modelled as a population map associated with a target signal level. Population maps are rough estimates of how traffic is distributed in space and time. They can either be purchased or reconstructed using marketing data (government open data or commercial data, e.g. Pitney Bowes) and geodata information (building locations, types and heights). When available, population maps may be refined using operator network measurements, e.g. the average number of users served by a given macro cell. Whereas such population maps are adequate for rough estimations, they are not accurate enough to be used for small cell deployment. Indeed, the space distribution is limited by the large-scale characteristics of population data, while time variations (e.g. commuting hours) are not taken into account. Moreover, coverage is simply modelled by a target signal level, which is not enough to yield QoS metrics.

On the other hand, some academic and commercial system-level simulators provide packet-based emulation of networks [Hen08], [Iku10]. Such packet-based simulators work using realistic data flows and include both large and small-scale traffic modelling. In [Iku10], a traffic type is assigned to each user amongst different traffic types (FTP, Web Browsing / HTTP, Video Streaming, VoIP, Gaming). Each traffic type models packet inter-arrival duration and packet size according to recommendations defined in [3GPP07]. Such packet-based simulations allow scheduling at the base station, leading to realistic interference between base stations in LTE. While such simulations are more accurate, they are too resource-expensive to be used in large-scale environments. Moreover, while small-scale traffic models are elaborate, packet-based simulators do not provide tools or methods to realistically distribute users and services in space.

Small-cell deployment raises two main challenges: how to obtain a more accurate user and traffic profile distributions both in space and time and how to move from basic coverage to accurate QoS estimations with fast and simple operational tools. SIRADEL will work on those two subjects to complete its simulation chain.

To obtain more accurate traffic distributions, an effort will be made by crossing data already used and innovative data to improve existing generation methods. Besides geodata, marketing data and network monitoring data, we plan to use enriched data contained in SIRADEL geo-database (e.g. building commercial characterization) as well as to exploit web-available data. Indeed, the World Wide Web is abundant with information that is commonly known under the term Big Data. Sorting and exploiting useful information is the key challenge of Big Data. In the context of traffic modelling, we will focus on points of interest provided by platforms such as Foursquare or Factual as well as social network information, e.g. locations and moves from Twitter users. Figure 90 shows an example of a user density map generated using points of interest. Such maps, in complement with indoor/outdoor maps, building height maps and global distribution of user profiles will be the basis of the traffic model that will enable a 3D distribution of the data traffic over the study area.

Traffic models will then be integrated into SIRADEL LTE system simulator to refine QoS estimations. Some large-scale scenarios (over an area covered by several macros) will be run using dynamic system simulations. Results will lead to the elaboration of abstraction models that characterize the impact of resource management and SON algorithms on cell selection, network performance and co-channel interference levels. The integration of those abstraction models into fast operational simulation processes will provide reliable and straightforward evaluation of the network implementing the considered algorithms. This will possibly make it possible to combine different algorithms within the simulation platform and to run network optimization design processess.

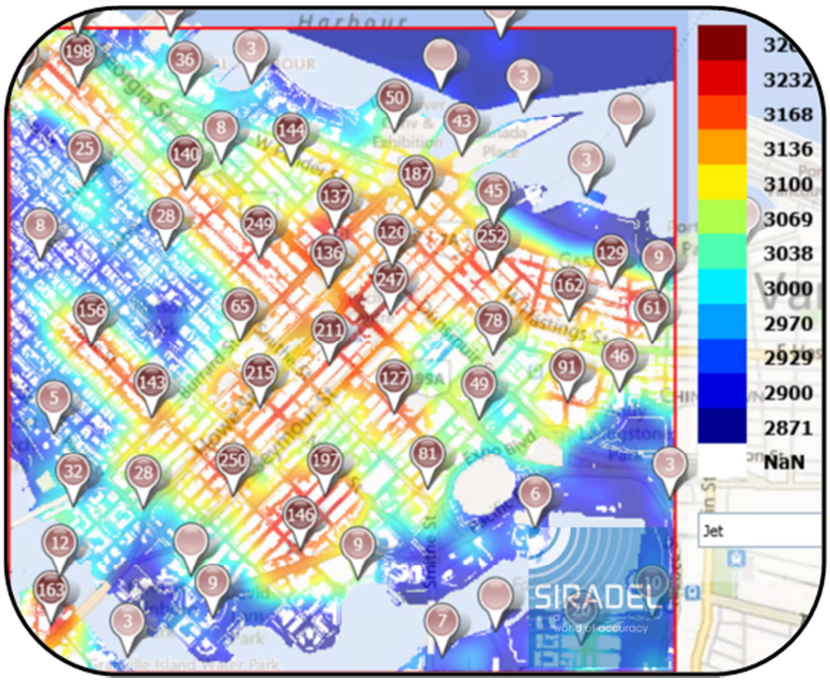


Figure 90. Outdoor-only user density (UE/km²) generated using points of interest.

7 CONCLUSION

This deliverable has provided an initial view from SHARING Work Package 4 on new opportunities, challenges and innovative concepts for SON and/or heterogeneous network deployments. In addition to the concept descriptions, some initial evaluation results, where available, have been presented.

New concepts are presented within the areas of: intra-system radio access offloading (Task 4.1), inter-system radio access offloading (Task 4.2), energy saving mechanisms (Task 4.3) and spectrum resource allocation (Task 4.4). In addition to presenting new concepts, this deliverable discusses also the performance and deployment strategies of heterogeneous network deployments within various scenarios.

Based on the obtained evaluation results the proposed concepts are indeed able to improve both the user performance and the overall system capacity. Furthermore, the proposed concepts are shown to reduce the overall network energy consumption. In general, the observed performance enhancements are expected to contribute to lower CAPEX and OPEX for the operators. However, more work is still needed to investigate the real potential of the proposed concepts.

REFERENCES

[3GPP07] 3GPP TSG-RAN1#48 (Orange; China Mobile; KPN; NTT Docomo; Sprint; T-Mobile; Vodaphone & Telecom Italia) "LTE physical layer framework for performance verification" 2007.

[3GPP10a] 3GPP TR 36.814, v9.0.0 "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects", March 2010.

[3GPP11a] 3GPP TSG RAN WG1, R1-114336, "Base Station Power Model", San Francisco, CA, USA, November 2011.

[3GPP12] 3GPP TS 36.104, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Base station (BS) radio transmission and reception. Release 12"

[3GPP887] 3GPP TR36.887, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Study on energy saving enhancement for E-UTRAN. Release 12".

[AKKM] Adrian Kosowski, Krystof Manuszewski, "Classical coloring of graphs", available at http://fileadmin.cs.lth.se/cs/Personal/Andrzej_Lingas/k-m.pd

[BAL05] Constantine A. Balanis. "Antenna Theory, Analysis and Design." Third edition, 2005,Wiley-Interscience.

[Bra12] M. Brau, Y. Corre and Y. Lostanlen, "Assessment of 3D network coverage performance from dense small-cell LTE", IEEE International Conference on Communications (ICC) 2012, Ottawa, June 2012.

[Boy06] S.Boyd, A.Ghosh, P.Prabhakar,D.Shah, "Randomized Gossip algorithms", *IEEE Transactions on Information Theory*, Special issue of *IEEE Transactions on Information Theory and IEEE ACM Transactions on Networking*, June 2006, 52(6):2508-2530.

[CHE10] N. Chevaux, M. M. De Souza, "Class-D power amplifiers using LDMOS and GaN power devices: a comparative analysis," MELECON 2010 - 2010 15th IEEE Mediterranean Electrotechnical Conference, April 2010, pp.691-694.

[Cis14] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2013-2018", February 2014.

[Col12] Coletti, C.; Hu, L.; Huan, N.; Kovács, I. Z.; Vejlgaard, B.; Irmer, R. & Scully, N. "Heterogeneous deployment to meet traffic demand in a realistic LTE urban scenario" Vehicular Technology Conference (VTC Fall), 2012 IEEE, 2012, 1-5.

[Col13] Mikael Coldrey et al., "Non-line-of-sight small-cell backhauling using microwave technology", IEEE commun. magazine, Sept. 2013.

[COM13] R. Combes et. al., "Interference coordination in wireless networks: a flow-level perspective," to appear in Proc. of IEEE INFOCOM 2013, preprint: https://sites.google.com/site/richardcombesresearch/research

[Cor09] Y. Corre and Y. Lostanlen, "Three-dimensional urban EM wave propagation model for radio network planning and optimization over large areas", IEEE Transactions on Vehicular Technology, Vol. 58, No. 7, pp. 3112-3123, Sept. 2009.

[D.2.2] Celtic-Plus SHARING, Deliverable D2.2, "Scenarios, KPIs and Evaluation Methodology for Advanced Cellular Systems", 2014.

[EAR12] INFSO-ICT-247733 EARTH, Deliverable D4.3, "Final Report on Green Radio Technologies", 2012.

[EAR12a] EU FP7 EARTH project, Deliverable D2.3: "Energy efficiency analysis of the reference systems, areas of improvements and target breakdown" January 2012.

[Eri13] Ericsson, "Ericsson Mobility Report – On the Pulse of the Networked Society", November 2013.

[Fal12] Laetitia Falconetti, Pål Frenger, Harald Kallin, Thomas Rimhagen, "Energy Efficiency in Heterogeneous Networks", in Proc. IEEE Online Conference on Green Communications (GreenCom) 2012.

[FOR01] A. I. J. Forrester, A. Sóbester, A. J. Keane, Engineering Design via Surrogate Modelling: A Practical Guide, John Wiley & Sons, Chichester, ISBN 978-0-470-06068-1.

[GAR04] M. Garca-Lozano et al., "UMTS optimum cell load balancing for inhomogeneous traffic patterns", Proceedings of IEEE Vehicular Technology Conference (VTC), 2004 Fall.

[GOW1] WiFi Report 2013. GOWEX Spain February 2014. http://es.scribd.com/doc/206648696/GOWEX-Informe-WiFi-2013

[GRT13] Green touch, Mobile Communications WG Architecture "Reference scenarios", May 2013.

[Hen08] Henderson, T.; Lacage, M.; Riley, G.; Dowell, C. & Kopena, J. "Network simulations with the ns-3 simulator" SIGCOMM demonstration, 2008.

[Hinf91] T. Basar and P. Bernhard, H∞-Optimal *Control and Relaxed Minimax Design Problems: A Dynamic Game Approach.* Birkhauser, Boston, MA, 1991.

[Hil13] Kimmo Hiltunen, "Utilizing eNodeB Sleep Mode to Improve the Energy-Efficiency of Dense LTE Networks", in Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) 2013, London, The United Kingdom, September 2013.

[Iku10] Ikuno, J.; Wrulich, M. & Rupp, M. "System level simulation of LTE networks" Proc. of the 71st IEEE Vehicular Technology Conference (VTC 2010-Spring), 2010, 1-5.

[ISOM04] I. Siomina et al., "Optimization of pilot power for load balancing in WCDMA networks," Proc. of IEEE GLOBECOM, 2004, Vol. 6, pp. 3872-3876.

[JLI05] J. Li et al., "Improved pilot power adjustment for load balancing in the CDMA system," Proc. of IEEE WiCOM 2005, Vol. 2, pp. 1010-1023.

[KG11] M.Khanfouci,N.Gresset, "Consensus based decentralization of interior point methods for heterogeneous network energy saving" Proc. of PIMRC, 2011, 300-304.

[LE06] K. K. Leung and C. W. Sung, "An opportunistic power control algorithm for cellular network," *IEEE/ACM Transactions on Networking,* vol. 14, pp. 470-478, Jun. 2006

[Let13] F. Letourneux, S. Guivarch and Y. Lostanlen, "Propagation Models For Heterogeneous Networks", in Proc EuCAP'13, Gothenburg, April 2013.

[Let14] F. Letourneux, S. Guivarch and Y. Lostanlen, "3D Propagation and Environment Modeling for NLOS Wireless Small-cell Backhaul", in Proc EuCAP'14, The Hague, April 2014.

[MAC13] A. F. Hanif, T. Hamidou, M. Assaad, D. Zeghlache, "Distributed Power Control in Femto Cells using Bayesian Density Tracking", IEEE Allerton conference 2013.

[MAL13] N. Ul Hassan, M. Assaad and H. Tembine, "Robust Power Control in Arbitrary Wireless Networks", IEEE Communication Letters, vol. 17, no 6, pp. 1124-1127, 2013.

[MAR10] T. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas", *Wireless Communications, IEEE Transactions on, vol. 9, pp. 3590 -3600, November 2010.*

[MIN14] W. Min, E. Ramos, Y.P. Wang, N. Lidian, S. Nammi, M. Curran, *Evaluation of Mobility Performance and Deployment Scenarios in UMTS Heterogeneous Networks*. IEEE VTC 2014 Spring; February 18, 2014

[MOH12] L. S. Mohjazi et al., "Self-Optimization of Pilot Power in Enterprise Femtocells Using Multi objective Heuristic," Journal of Computer Networks and Communications, Vol. 2012, Article ID 303465.

[ORF08] Sophocles J. Orfanidis. "Electromagnetic Waves and Antennas." 2008. ch 15 pp 693-728.

[PED12] K.I. Pedersen et al., "eICIC Functionality and Performance for LTE HetNet Co-Channel Deployments," Proc. of IEEE VTC Fall 2012, September 2012.

[PICH13] V. Picheny et al., "A benchmark of kriging-based infill criteria for noisy optimization", Structural and Multidisciplinary Optimization, on press, 2013, preprint: http://hal.archives-ouvertes.fr/hal-00658212/

[RAM11] J. Ramiro, K. Hamied, "Self-Organizing Networks (SON): Self-Planning, Self-Optimization and Self-Healing for GSM, UMTS and LTE", Wiley, 2011.

[Rec_2001] Rec, ITU-T (2001), P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and

speech codecs".

[Rec_2003] Rec, ITU-T (2003), G.107, "The E-model, a computational model for use in transmission planning".

[SAL93] N. Gordon, D. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," *Radar and Signal Processing, IEE Proceedings F,* vol. 140, no. 2, pp. 107–113, 1993.

[SAY05] A. Subramanian and A. H. Sayed, "Joint rate and power control algorithms for wireless networks," *IEEE Transactions on Signal Processing,* vol. 53, no. 11, pp. 4204-4214, Nov. 2005.

[Ski12] Per Skillermark, Pål Frenger, "Enhancing Energy in LTE with Antenna Muting", in Proc. IEEE Vehicular Technology Conference (VTC) 2012 Spring, Yokohama, Japan, May 2012.

[SPEN04] Spencer, Q.H.; Swindlehurst, A.L.; Haardt, M., "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *Signal Processing, IEEE Transactions on* , vol.52, no.2, pp.461,471, Feb. 2004.

[Tell99] Tellado, "Peak to average power reduction for multicarrier modulation", University, September 1999.

[VLA13] Vlad-Ioan Bratu et al., "Antenna Tilt Load Balancing in Self-Organizing Networks," International Journal of Research in Wireless Systems (IJRWS), Vol. 2, No. 1, pp. 21-26, March, 2013.

[WAN12] Y. Wang et al., "Performance Analysis of Enhanced Inter-Cell Interference Coordination in LTE-Advanced Heterogeneous Networks", Proc. of IEEE VTC Spring 2012, September 2012.

[Yat95] R. Yates, "A framework for uplink power control in cellular radio systems," *Selected Areas in Communications, IEEE Journal on,* vol. 13, no. 7, pp. 1341–1347, 1995.

[YOO06] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming", Selected Areas in communications, IEEE Journal on, vol. 24, pp. 528 - 541, March 2006.

GLOSSARY

ACRONYM	DEFINITION
3G	Third Generation cellular system
3GPP	Third Generation Partnership Project
AAS	Active Antenna Systems
ABR	Average Bit Rate
ABS	Almost Blank Subframes
AC	Alternating Current
ACPR	Adjacent Channel Power Ratio
AM	Amplitude Modulation
ANDSF	Access Network Discovery and Selection Function
AP	Access Point
B2B	Business-to-Business
BB	Baseband
BCR	Block Call Rate
BER	Bit Error Rate
BLER	Block Error Rate
BLUP	Best Linear Unbiased Prediction
BS	Base Station
BW	Bandwidth
CA	Carrier Aggregation
CAPEX	Capital Expenditure
CC	Component Carrier
CCO	Capacity and Coverage Optimization
CDF	Cumulative Distribution Function
CDS	Connected Dominant Set
CN	Core Network
CoMP	Coordinated Multi-Point Transmission and Reception

CQI	Channel Quality Indicator
CRE	Cell Range Extension
CREB	Cell Range Expansion Bias
CRS	Cell-specific Reference Signal
CURE	Cell Uplink Range Expansion
DAS	Distributed Antenna System
DC	Direct Current
DCR	Drop Call rate
DFT	Discrete Fourier Transform
DL	Downlink
DTX	Discontinuous Transmission
EAP-SIM	Extensible Authentication Protocol – Subscriber Identity Module
EARTH	Energy Aware Radio and Network Technologies
EDF	Earliest Deadline First
eICIC	Enhanced Inter-Cell Interference Coordination
eNB	Evolved NodeB
eNodeB	Evolved NodeB
EPC	Evolved Packet Core
EVM	Error Vector Magnitude
FTP	File Transfer Protocol
FTT	File Transfer Time
GaN	Gallium Nitrate
GERAN	GSM EDGE Radio Access Network
GW	Gateway
HeNB	Home eNodeB
HetNets	Heterogenous Networks
HIP	Host Identity Protocol
HM	Handover Margin

HS	High Speed
HSPA	High Speed Packet Access
HTTP	Hypertext Transfer Protocol
HWN	Heterogeneous Wireless Networks
IBO	Input Back Off
ID	Identity
IP	Internet Protocol
ISD	Inter-Site Distance
ITU-T	International Telecommunication Union – Telecommunication Standardization Sector
JPM	Joint Performance Metric
KPI	Key Perforamance Indicator
L2	Layer 2
L3	Layer 3
LB	Load Balancing
LDMOS	Laterrally Diffused MOS
LHS	Latin Hypercube Sampling
LPN	Low-Power Node
LTE	3GPP Long Term Evolution
LTE-A	LTE-Advanced
LOS	Line Of Sight
LSAS	Large Scale Antenna System
MADM	Multiple Attribute Decision Making
MBS	Macro Base Station
MIH	Media Independent Handover
MIHF	Media Independent Handover Function
MIMO	Multiple-Input Multiple-Output
MIPv4	Mobile-IP version 4
MIPv6	Mobile-IP version 6

MIS	Maximum Independent Set
MLB	Mobility Load Balancing
MLE	Maximum Likelihood Estimation
MNO	Mobile Network Operator
MOS	Mean Opinion Score
MSC	Message Sequence Chart
MUE	Macro UE
NGH	Next Generation Hotspot
NLOS	Non-Line Of Sight
NP	Network Parameter
OFDM	Orthogonal Frequency Division Multiplexing
OMC	Operations and Maintenance Center
OOB	Out Of Band
OPEX	Operational Expenditure
OS	Operating System
OTT	Over The Top
PA	Power Amplifier
PAE	Power Added Efficiency
PAPR	Peak to Average Power Ratio
PBCH	Physical Broadcast Channel
P-CPICH	Primary Common Pilot Channel
PDCP	Packet Data Convergence Protocol
PESQ	Perceptual Evaluation of Speech Quality
PF	Proportional Fair
PM	Phase Modulation
PRB	Physical Resource Block
PS	Proactive Scheduling
PSD	Power Spectral Density
PSNR	Peak Signal to Noise Ratio

PSS	Primary Synchronization Signal
PTS	Partial Transmit Sequence
PU	Public
PUE	Pico UE
QAM	Quadrature Amplitude Modulation
QL	Q-Learning
QoE	Quality of Experience
QoS	Quality of Service
RAT	Radio Access Technology
RB	Resource Block
RE	Range Extension
RET	Remote Electrical Tilt
RF	Radio Frequency
RL	Reinforcement Learning
RLC	Radio Link Control
RMS	Root Mean Square
RNC	Radio Network Controller
RRC	Radio Resource Control
RRM	Radio Resource Management
RRU	Remote Radio Unit
RSRP	Reference Signal Received Power
RSS	Received Signal Strength
RTN	Real-Time Network
SAP	Service Access Point
SBS	Small cell Base Station
SCBS	Small Cell Base Station
SFN	Single Frequency Network
SHARING	Self-Organized Heterogeneous Advanced Radio Networks Generation

SIM	Subscriber Identity Module
SINR	Signal to Interference and Noise Ratio
SIP	Session Initiation Protocol
SIR	Signal to Interference Ratio
SLA	Service-Level Agreement
SLM	Selective Mapping
SNR	Signal to Noise Ratio
SON	Self Optimizing/Organizing Network
SPPP	Spatial Poisson Point Process
SSPA	Solid State Power Amplifier
SSS	Secondary Synchronization Signal
TD-LTE	Time Division Long-Term Evolution
TDD	Time Division Duplex
TDM	Time Division Multiplexing
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
TR	Tone Reservation
TTI	Transmission Time Interval
UE	User Equipment
UL	Uplink
UMTS	Universal Mobile Telecommunications System
UTRAN	Universal Terrestial Radio Access Network
VBR	Variable Bit Rate
VoIP	Voice over IP
VSC	Virtual Small Cell
WCDMA	Wideband Code Division Multiple Access
WEP	Wired Equivalent Privacy
WiFi	Wireless Fidelity
WiMAX	Worldwide Interoperability for Microwave Access

WLAN	Wireless Local Area Network
WP	Work Package
WPA	WiFi Protected Access
WPA2	WiFi Protected Access II
X2	Interface between eNodeBs