Fast Muting Adaptation for LTE-A HetNets with Remote Radio Heads

Beatriz Soret*, Klaus I. Pedersen†*, Troels Kolding†, Hans Kroener†, Ioannis Maniatis†
* Aalborg University
† Nokia Siemens Networks
bsa@es.aau.dk

Abstract—Enhanced Intercell Interference Coordination (eICIC) techniques boost the performance of co-channel deployments of macro and small cells. However, features like the Almost Blank Subframes (ABS) are intended to be semistatically configured, due to the limited amount of information and the intrinsic delay over the X2 inter-cell interface. In this paper we investigate dynamic solutions for the allocation of the muted resources with small cells in the form of Remote Radio Heads (RRH), connected to the macro cell through a low latency interface. A simple eICIC framework for intercell Radio Resource Management in macro-RRH cases is developed. Based on minimal information collection, we propose an algorithm that adjusts the ABS ratio on a fast basis, aiming at balancing the instantaneous load between the macro and the RRH layer. Performance results with bursty traffic and low and high load conditions show that the dynamic algorithm provides gains of 40-50% both in 5%-ile and 50%-ile user throughput compared to the semi-static case.

I. INTRODUCTION

Migration from macro-only networks to heterogeneous networks (HetNet) is a promising option for increasing the capacity of cellular systems such as LTE-Advanced networks [1]. A HetNet consists of a mixture of macro cells and smaller low power nodes such as pico and femto cells. HetNets have been extensively studied in academia, industry, and standardization bodies such as 3GPP [2]. In this study we focus on the downlink performance of LTE-Advanced co-channel HetNet deployments [3], [4], meaning that macro and small cells are using the same carrier frequency. One of the main challenges in this scenario is the macro interference to the users in the small cells [5] - [7], that can be alleviated by the use of enhanced inter-cell interference coordination (eICIC) techniques [8]. The eICIC scheme relies on time-domain interference coordination between the macro and small cell layer, where some subframes are partially muted at the macro layer to lower the experienced interference level by the users served at the small cell layer. With eICIC the offload of traffic to the small cell can be improved, resulting in a more balanced load between different cell types, yielding better overall system level performance and a higher end-user throughput.

The majority of the published eICIC studies have focused on cases where the small cells are realized with pico cells [6] [7] or closed subscriber group femto cells [9]. For the cases with macro and pico, it is typically assumed that each cell has its own radio resource management (RRM) algorithms including packet scheduling, link adaptation and hybrid ARQ [10], while the coordination of the eICIC muting patterns is supported via the backhaul X2 interface between the different base stations (called eNBs in LTE). Due to the X2 signaling delays, and for the sake of the overall system stability, it has typically been assumed that muting patterns are only updated on a slow time scale of hundreds of milliseconds or several seconds [7]. This implies that muting patterns cannot track fast traffic fluctuations, but rather aim at capturing the envelope of the average traffic variations.

An alternative implementation consists of deploying Remote Radio Heads (RRH) instead of pico cells. Thus, each RRH is connected via a high-speed low latency fronthaul to a macro cell, enabling centralization of the major RRM algorithms in the macro eNB. This essentially means that implementation of intercell RRM algorithms is possible for clusters of connected macro and RRHs, offering opportunities for further performance optimization. Thus, applying the eICIC concept to the macro-RRH case would in principle allow for fast intercell decision making on whether a subframe shall be muted or not. The objective of this study is to further study fast intercell RRM adaptation of eICIC muting for macro-RRH cases, as compared to scenarios with semi-static adaptation of muting patterns for macro-pico cases. A simple eICIC framework for macro-RRH cases is developed, where so-called optional subframes are selected to be either muted or not on a fast basis to maximize the overall system performance, allowing tracking of fast traffic fluctuations. An intercell RRH algorithm is proposed and evaluated in a complex LTE-Advanced HetNet scenario with bursty traffic models. We opt to evaluate the performance through statistically realiable simulations, with a system-level simulator enabling realistic assumptions for scenarios with multiple users and cells.

The rest of the paper is organized as follows. In Section II the network model and the eICIC framework for macro-RRH cases are detailed. Section III describes the fast intercell RRM adaptation, including the muting adaptation and the scheduling of the available resources. In Section IV we analyze the performance of the proposed algorithm. Finally, concluding remarks are given in Section V.

II. EICIC FRAMEWORK FOR MACRO-RRH NETWORKS

A. Network Model

We consider an LTE-Advanced network with macro cells and small cells (RRH). The intercell fast adaptation is conducted at a cluster level, consisting of one macro eNB and its underlying RRHs (only one in the example in Figure 1). The cluster is hence composed by a macro cell $\text{C}^m$ and the
set of small cells in its coverage area \( C = \{C^1, \ldots, C^s, \ldots, C^S\} \), with \( 1 \leq s \leq S \) and \( \{s, S\} \in \mathbb{N} \). Moreover, the set of User Equipments (UEs) is denoted by \( U = \{U^1, \ldots, U_u, \ldots, U_U\} \), with \( 1 \leq u \leq U \) and \( \{u, U\} \in \mathbb{N} \), where the superscript \( l = m \) indicates a macro cell UE, while \( l = s \) indicates an RRH user at cell \( C^s \). The macro eNBs are inter-connected through X2 interface, but no extra coordination among macro cells is considered.

In order to select the best serving cell, UEs measure the reference signals of nearby cells, transmitted at a constant power. Thus, the eNB offering the highest received reference signal is selected as the serving cell for the UE. However, the large difference in downlink transmit power among macro and small cell eNBs in a HetNet significantly reduces the coverage area of the small cell, and UEs will tend to connect to distant high-power macro cells rather than to low-power small cells. This imbalance is corrected by expanding the range of the small cell. Thus, a positive bias denoted as Range Extension (RE) offset is added to the reference signal measured from small eNBs, pushing more UEs to this layer [6].

With the application of the RE bias, we can distinguish between three kinds of users with very different interference conditions (Figure 1). First of all, we have the macro users connected to the macro eNB. Secondly, center RRH users are users in the default coverage area of the small cell. Finally, RE users are in the extended area of the small cell and connected to it due to the application of the RE offset.

### B. Subframe Classification

During the muted eICIC subframes, the macro cell still transmits essential system information and reference signals in order to provide support to the legacy UEs. The muted (or protected) subframes are therefore named Almost Blank Subframes (ABS) [7]. Macro users are not scheduled during those subframes, leading to a much lower macro interference that allows the small cells to serve UEs that are located in the extended area outside the default small cell coverage. To enable a fast muting allocation, the muting ratio shall be adjusted on a subframe basis. For this purpose we define three kinds of subframes in the macro layer (see Figure 2):

- Normal subframes: the data channel is transmitted
- Mandatory ABS: the data channel is not transmitted
- Optional ABS: they can be used either as ABS or normal subframes

Shortly before the begining of an optional ABS, the macro cell decides if it should be used as normal or ABS. It is worth mentioning that the notation optional ABS is just internal eNB notation, and not known to the UEs. Moreover, all clusters apply the same periodically repeated subframe structure, so that the mandatory ABS and the normal subframes are synchronized among eNBs, whereas the use of the optional ABS is decided at the cluster level. If the number of optional ABS per period is low, then most of the resources are semi-statically configured and the adaptation to the traffic fluctuations will be rougher. As the number of optional ABS increases, the accuracy of the adaptation increases. In the limit, no mandatory ABS or normal subframes would be defined so that all the resources can be used as protected or not depending on the traffic variations. However, it is convenient to define at least one normal and one mandatory ABS subframe for UE measurement and feedback purposes, as discussed later (II-D).

As shown in Figure 2, macro users are normally scheduled during normal subframes, but they are not scheduled during mandatory ABS. In the case of the small cell, center users are not affected so much by macro interference, whereas RE users suffer from strong inter-layer interference during normal subframes. Therefore, resource allocation should aim at protecting RE users. Thus, subframes overlapping mandatory ABS or optional ABS used as ABS should be used as much as possible for users in the worst SINR conditions (RE users) since in those subframes the macro interference is minimized. Only if those cannot be filled by the UEs in the extended area, the centre UEs might compete for the remaining resources.

### C. Notation

For the sake of clarity, the following notation within a cluster is defined:

- \( u_{macro} \) is the number of active users connected to the macro cell.
- \( u_{center} \) is the number of active RRH center users connected to small cells within the cluster.
- \( u_{RE} \) is the number of active RRH users in the cell-extended area within the cluster (RE users).
- \( U \) is the total number of active users in the cluster, so \( U = u_{macro} + u_{center} + u_{RE} \).
* $P_c(i)$ is the transmit power of cell $C_c$ at subframe $i$.  
* $P_{c_{\text{max}}}$ is the maximum transmit power of cell $C_c$, in this study 46dBm for macro cells and 30dBm for RRH.  
* $P_{c_{\text{base}}}$ is the transmission power of eNB at cell $C_c$ when it is muted. The data channel is muted but other common channels and Common Reference Signals need to be transmitted at constant power all the time, and thus $P_{c_{\text{base}}}$ is not zero and it is perceived as a residual interference during the protected subframes. User equipments with advanced receivers are able to cancel this residual interference.

Furthermore, we define the next elements in coherence with the subframe structure:

* $T_{\text{ABS}} \in \mathbb{N}$ is the ABS period (Figure 2).\footnote{The ABS period used in this paper is $T_{\text{ABS}} = 8$, to be a multiple of the ABS muting pattern defined in the LTE-A standard which is periodical with 40 subframes for FDD mode.}
* $z_{\text{ABS}}$ and $n_{\text{ABS}}$ are the number of subframes used as ABS and normal subframe in the current ABS period, respectively. $1 \leq z_{\text{ABS}} < T_{\text{ABS}}$ and $1 \leq n_{\text{ABS}} < T_{\text{ABS}}$, with $z_{\text{ABS}} \in \mathbb{N}, n_{\text{ABS}} \in \mathbb{N}$, i.e. at least one subframe is configured as normal subframe and another one as mandatory ABS. Both counters are increased everytime a subframe is used as protected resource and reset at the begining of the ABS period.

$s_f(i)$ refers to the kind of subframe at time $i$, with $s_f(i) \in \{\text{NORMAL}, \text{MANDATORY, OPTIONAL}\}$

### D. Measurement Reports

In an LTE system a UE $U_{c}^{u}$ feeds back periodically to its serving cell $C_c$ measurement reports to assist the cell selection procedure and channel quality indicators $CQI(U_{c}^{u})$ to assess radio channel conditions. When using ABS subframes, the reported channel quality indicators do not capture the diverse channel conditions between normal and protected subframes [7]. This mismatch is solved in LTE-Advanced specifications by defining restrictions in the measurements, so that the UE can be configured to perform separate measures at the small cell for normal and protected subframes. With the subframe structure in Figure 2, it is necessary to keep a minimum number of normal and mandatory ABS subframes to ensure that the restriction mechanism in LTE-Advanced is applicable here with no extra changes in the specification. Thus, RRH UEs are configured to perform and report separate channel measurements for mandatory ABS and normal subframes, and no measurements during optional ABS:

$$CQI(U_{c}^{u}), s_f(i) = \text{MANDATORY or NORMAL}$$

(1)

while macro UEs will take the channel condition measurements during normal subframes:

$$CQI(U_{c}^{m}), s_f(i) = \text{NORMAL}$$

(2)

In the eNB, the packet scheduler applies the proper measure during optional ABS depending on whether the subframe is going to be used as ABS or normal.

### III. INTERCELL FAST RRM ADAPTATION

The goal in the design of the intercell fast RRM algorithm is to improve the throughput of the users under the worst conditions, i.e. the cell-edge users within the cluster. Balancing the resource usage between the macro and the small cell layer provides the highest cell-edge user throughputs in co-channel deployments [7]. Therefore, the algorithm aims at balancing the available resources between both layers, and then allocating them to users. The intercell RRM functionality is hence divided into two parts: First of all, the muting ratio in the macro layer is adjusted based on the load of each layer. Secondly, the scheduling algorithm will assign the available resources to the active users with the goal of optimizing the overall system performance while ensuring fairness among users.

#### A. Muting Adaptation

The pseudo-code of the muting adaptation is illustrated in Algorithm 1. The main principle consists of checking the load at the macro and RRH layer at each optional ABS, and based on those measures decide whether the optional ABS shall be used as normal subframe or protected subframe.

The load in the RRH layer is defined as the percentage of users in the range extension area as compared to the total number of users in the cluster, and analogously in the macro layer with the percentage of macro users. Notice that the load measure in the RRH layer refers only to RE users, since those are the ones benefiting more from ABS resources. It has been verified through simulations that this option (not using center users) is giving the best performance results. The load measure includes the total number of RE users in all the RRHs of the cluster.

The dynamics of the algorithm is as follows: The total transmission power at the macro eNB depends on the kind of subframe. If it is an optional ABS, the algorithm ensures first of all that the percentage of macro users is served with an appropriate percentage of full power subframes:

$$u_{\text{macro}} / U < n_{\text{ABS}} / T_{\text{ABS}}$$

(3)

If (3) is fulfilled, then it is checked if the percentage of ABS resources assigned so far is lower than the percentage of high-interfered users:

$$u_{\text{RE}} / U > z_{\text{ABS}} / T_{\text{ABS}}$$

(4)

If (4) is true, the optional ABS is protected, and the RE users will have an opportunity to be scheduled in the next subframe. Therefore the algorithm ensures first of all the service of macro users, since the coverage area of the macro eNB is much larger, and the macro cell-edge users do not have the option of being scheduled with reduced interference conditions. In the case of the RRH eNB, it is always transmitting at full power. For dense deployment of small cells, the low-power eNB interference might be significant and solutions coordinating also the resource allocation among small cells are for further study.

It is worth noting that the application of the algorithm within a cluster may likely lead to the use of different muting patterns in neighbouring macros. Having a tight coordination...
Algorithm 1 Fast Muting Adaptation

1: iter = 0
2: while (iter < iter_max) do
3:   if (iter mod T_ABS == 0) then
4:     n_ABS = 0
5:     z_ABS = 0
6:   end if
7:   switch (s(iter))
8:     case NORMAL:
9:       P^m(iter) = P^m_max
10:    n_ABS = n_ABS + 1
11:   case MANDATORY:
12:    P^m(iter) = R^max
13:    z_ABS = z_ABS + 1
14:   default:
15:     (Optional Subframe)
16:     if (u_{UE} / U > z_ABS / T_ABS) and (u_{macro} / U < n_ABS / T_ABS) then
17:       P^m(iter) = R^max
18:       {Use it as ABS subframe}
19:     else
20:       P^m(iter) = P^m_max
21:       {Use it as normal subframe}
22:   end if
23:   end switch
24:   for s = 1 : S do
25:     P^s(iter) = P^max
26:   end for
27:   iter = iter + 1
28: end while

between macro cells to have exactly the same ABS patterns gives the optimal performance with semi-static muting [8]. However, as it will be shown in the performance results, the gain from dynamically adjusting the ABS is still very significant in spite of the non-explicit inter-macro coordination.

B. Resource Scheduling Algorithm

Once the transmission power is set, the scheduling algorithm will assign the available resources to the corresponding users. In this paper, we select a commonly used scheduler, Proportional Fair (PF). PF is applied separately at each cell (macro cell and each small cell), since users are connected to only one cell. Thus, the resource element at cell C^u is assigned to the user that maximizes the following scheduling metric:

\[
\arg \max_u \left\{ M_{u,k} \right\}
\]

where \( u \) is the user index and \( k \) is the Physical Resource Blocks (PRB) group index. According to the LTE-Advanced physical layer structure, one PRB is the minimum resource element, constituted of 12 consecutive subcarriers with subcarrier spacing of 15kHz, for one transmission time interval.

The scheduling metric of PF is calculated by dividing the instantaneous throughput by the average throughput:

\[
M_{u,k} = \frac{r_{u,k}}{R_u}
\]

where \( r_{u,k} \) is the estimated throughput for user \( u \) at the \( k^{th} \) PRB group, and \( R_u \) is the average throughput for that user. \( r_{u,k} \) is based on the channel quality indicator reported by the user either on the last normal subframe or on the last mandatory ABS. PF aims at maximizing the utility function \( \sum_u \log(R_u) \) by exploiting multi-user diversity.

IV. PERFORMANCE RESULTS

A. Simulation Assumptions

The scenario is in coherence with the definition given in [12] for hotzone deployment of RRH eNBs in macro cells. The network topology consists of a standard hexagonal grid of three-sector macro eNBs complemented with a set of outdoor RRHs. Macros and RRHs share the same 10MHz of bandwidth at a carrier frequency of 2GHz. There are a total of 7 macro sites (21 macro cells) with wrap around to simulate the interference effect of a larger network. The macro intersite distance is 500m, and the minimum distance among RRHs is 40m. One cluster is composed of one macro cell and 4 underlying RRHs. The system-level simulator follows the LTE specifications, including detailed modeling of major radio resource management functionalities such as packet scheduling, hybrid ARQ and link adaptation [10]. We use closed loop 2x2 MIMO with pre-coding and rank adaptation and UEs using an Interference Rejection Combining (IRC) receiver with cancellation of Common Reference Signals interference during protected subframes (advanced users). Macros and RRHs are transmitting at 46dBm and 30dBm, respectively, with macro antenna pattern as defined in [12] and RRHs with omni-directional antennas.

We consider a dynamic traffic model with Poisson call arrival, assuming a finite payload \( B = 1 \text{Mb} \) for each call. Once the payload has been successfully delivered to the UE, the call is terminated. We define both the total arrival rate \( \lambda \cdot \text{UE} / s \) and the average offered traffic load \( L = \lambda \cdot B \). With this small value of \( B \), the users transmit their load very fast leading to significant traffic fluctuations in the system. Even though it is not shown in this paper, higher values of payload (up to 10Mb) have also been simulated. Not only the trends and conclusions discussed here remain invariant but also the relative gains are in the same range.

The 5%-ile and 50%-ile user throughput obtained from the Cumulative Distribution Function curves are used as performance indicators. We compare dynamic muting adaptation in the centralized architecture with standard eICIC cases where radio resource management functionalities are independently working at each eNB and the macro ABS is semi-statically configured.

The number of normal and mandatory ABS is set to the minimum (1 subframe), so that the muting ratio can vary from 1/8 to 7/8. It has been checked by simulations that giving the maximum flexibility to the muting adaptation provides the maximum benefit in terms of user throughput.

B. User Throughput

In Figures 3 and IV-C we show the 5%-ile and 50%-ile user throughput as a function of the average offered load. There are three curves, corresponding to no eICIC, eICIC with picos and eICIC with RRHs and fast muting allocation. As expected, both the 5%-ile and the 50%-ile user throughput decrease as the average offered load increases for the three cases.

In the eICIC curve with semi-static muting adaptation, the optimal eICIC parameter settings for the different values of offered load are indicated (aiming at maximizing the cell-edge throughput) [7]. Both the ABS muting ratio and the
RE are adjusted to track the average envelope of the traffic. It is illustrated how the optimal eICIC configuration varies versus the offered traffic load by displaying the best settings of ABS muting ratio and RE. At low offered load, there is little, or marginal, gain from applying eICIC. This is due to the fact that there is only marginal other-cell interference, and the gain in this low loaded cases comes from the application of a small RE offset at the pico. As the offered load increases, both macros and picos start having higher probability of transmitting (and thus causing interference to other cells), and the system converges to using more ABS at the macros and higher RE at the picos.

With RRHs and fast muting decisions, the only parameter to be optimized is the RE (with the best value shown in the Figures), since the ABS is dynamically adjusted by the algorithm. The behaviour is analogous to the semi-static ABS muting ratio: For low load, small values of RE are recommended because of the few number of users in the system. As the load increases, it is convenient to offload users to the RRH layer by application of higher values of the RE offset. However, the optimal RE when using dynamic ABS is lower than the semi-static case, due to the better load balancing by using the appropriate ABS at each time.

The gain of eICIC compared to no eICIC is well-known [6] - [8], but in Figures 3 and IV-C it is observed that there is still a significant gain both in 5%-ile and 50%-ile when moving from semi-static muting ratio to dynamic solutions: For a target cell-edge throughput of 1Mbps, the semi-static adaptation supports 48Mbps of offered load, while the fast adaptation allows up to 67Mbps, leading to a relative gain of 40%. Similarly, the relative gains of 3Mbps and 5Mbps are 54% and 43%, respectively. Next, we analyze other indicators like the number of users and the resource utilization to explain in more detail where this throughput gain comes from.

C. Number of Users and ABS Muting Ratio

Figure 5 shows the CDF of the number of active users in the macro (solid line) and RRH (dashed line) layer, for different values of the average arrival rate varying from 10 users/s to 70 users/s (corresponding to an average offered load varying from 10Mbps to 70Mbps). For the RRH layer the total amount of RRH users in the cluster is shown. For low values of load (10 users/s), both layers are empty most of the time. The probability of having at most one user goes up to 95%, being the probabilities for both layers very close. As the user arrival rate increases, the differences between layers become noticeable. In both cases the probability of empty cell decreases and the number of active users increases, but the growth is much more significant in the macro layer. With 70Mbps of load, the RRH layer still has more than 90% of probability of having less than 10 active users while the macro layer is clearly saturated.

In view of Figure 5 we expect the muting algorithm to use a small muting ratio most of the time: for low load, because both layers are empty most of the time, and for high load due to the higher amount of users in the macro layer. This expected behaviour is plotted in Figure 6 where the CDF of the muting ratio is plotted for different values of the cluster arrival rate λ. For comparison, also the optimal semi-static muting ratio in scenarios with macros and picos is indicated with circles in the x-axis for the different arrival rates (the optimal obtained as explained in IV-B). Most of the time the intercell algorithm is using the minimum muting ratio, 1/8. For low load, both layers are empty most of the time, so the algorithm tends to use the minimum muting ratio. As the load increases, the number of users in both layers grows at a similar rate, and the average muting ratio slightly increases. For high values of load (70 users per second) the number of macro users significantly increases, and the algorithm will try to reduce the muting ratio as much as possible to serve the macro layer.

D. Average Resource Usage

In Table I we show the average PRB usage with dynamic and semi-static muting ratio for different values of the average offered load. It has previously been observed [7] that the optimal value of RE for each offered load corresponds to the
Fig. 5. Number of active users in the macro and RRH layer for different arrival rates.

TABLE I. AVERAGE RESOURCE USAGE

<table>
<thead>
<tr>
<th>Average offered load</th>
<th>Dynamic ABS ratio</th>
<th>Static ABS ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>macro</td>
<td>RRH</td>
</tr>
<tr>
<td>10Mbps</td>
<td>21.5%</td>
<td>22.7%</td>
</tr>
<tr>
<td>30Mbps</td>
<td>51.3%</td>
<td>48.6%</td>
</tr>
<tr>
<td>50Mbps</td>
<td>77.2%</td>
<td>72.6%</td>
</tr>
<tr>
<td>70Mbps</td>
<td>91.3%</td>
<td>89.3%</td>
</tr>
<tr>
<td>80Mbps</td>
<td>95.5%</td>
<td>95.0%</td>
</tr>
</tbody>
</table>

value that balances the average PRB usage in both layers. However, as it can be checked in the table, when the muting ratio is semi-statically configured with the RE offsets in Figure 3, the average PRB usage is only well balanced for small values of offered load. As the load increases, the PRB usage at the macro layer increases faster than at the RRH layer. The maximum achievable load is hence lower for the semi-static case than for the dynamic solutions, where up to 80Mbps can be served with a resource utilization of 95% in both layers.

V. CONCLUSIONS

A simple eICIC framework for HetNets with macro cells and RRHs is developed, aiming at boosting the network performance by means of dynamic muting decisions. The proposed intercell fast muting adaptation algorithm dynamically adjusts the ABS muting according to the instantaneous load conditions at the macro and RRH layer. The input required by the algorithm reduces to the number of macro users and the number of RRH users in the worst interference conditions. Performance results show relative gains of 40-50% for fast adaptation in scenarios with bursty traffic (both low and high load) as compared to cases with semi-static ABS patterns that can only be adjusted on time-scales of hundreds of milliseconds. As future work, the investigated techniques can be adapted to distributed HetNets with macro and pico cells by fully exploiting the information exchange over X2.

Fig. 6. CDF of the muting ratio for semi-static muting ratio and dynamic muting ratio solutions.

ACKNOWLEDGMENT

The authors would like to thank David Ruiz and Juan M. Gómez, from Aalborg University, and Jens Steiner and Mads Brix, from Nokia Siemens Networks, for their valuable help in carrying out this study.

REFERENCES