When CoMP is beneficial – and when it is not. Selective coordination from a spectral efficiency and a users’ throughput perspective

Christian M. Mueller
University of Stuttgart, Institute of Communication Networks & Computer Engineering
Pfaffenwaldring 47, 70569 Stuttgart, Germany
Email: christian.mueller@ikr.uni-stuttgart.de

Abstract—Coordinated Multi-Point (CoMP) transmission schemes can provide large spectral efficiency gains in cellular networks. Because of the coordination among neighbor basestations, the planning interval for resource allocation with CoMP is longer than for uncoordinated local transmission modes. Especially for bursty data traffic, this planning interval might be too long to keep up with changes in offered load at small timescales. In this case, coordination gains might turn into losses when the coordinated resource allocation doesn’t fit the system state anymore. To avoid these problems and to leverage the CoMP schemes’ potential at higher layers, previous work has investigated traffic aware transmit mode selection schemes with the objective to maximize spectral efficiency. In this work, we do not focus on spectral efficiency, but on per-user throughput. We derive a suitable coordination threshold to improve throughput and evaluate its impact on system performance.

I. INTRODUCTION

We consider Coordinated Multi-Point (CoMP) transmission schemes in the downlink of an LTE-like system. Although standardization for LTE-Advanced currently focuses on the application of CoMP in a single eNodeB, i.e. an eNodeB with several remote radio heads, we consider a distributed CoMP implementation where several macro eNodeBs cooperate in a decentralized fashion. We focus on the CoMP schemes Coordinated Scheduling (CS) and Coordinated Precoding (CP), which are less demanding in terms of backhaul bandwidth and latency than Joint Transmission (JT). For an overview of CoMP and its role in LTE-Advanced, we refer to [1], [2].

Previous research on CoMP schemes focused on physical layer issues and on algorithms to coordinate resource allocation and antenna parameters in a way that improves spectral efficiency [3], [4]. While most authors assume full-buffer traffic, some authors investigated the impact of bursty data traffic on CoMP schemes [5]–[7]: For bursty traffic, the amount of data in a user’s buffer at the basestation is variable. It might happen that a user’s buffer at the basestation runs empty while a coordination process with neighbor basestations is ongoing. The constraints on resource allocation the basestations agreed upon might then not fit the new system state anymore and actually degrade the system’s performance. If the transmitted objects are large, buffers run empty very rarely and we can neglect this effect. The system state varies slowly and the spectral efficiency gains from coordinated transmission of large objects outweigh any potentially negative impact when buffers run empty. However, if the transmitted objects are small, buffers run empty regularly and the system state changes quickly. In this case, the coordination process among the basestations might not be able to keep up. Unfortunately, for Internet data traffic, the object size distribution is heavy-tailed, with many small objects and only few very large objects [8]. This also holds for data traffic in wireless networks [9].

In our previous work, we proposed a traffic-aware transmit mode selection algorithm which uses coordinated transmission for large objects and uncoordinated transmission for small objects [7], [10]. The objective of the transmit mode selection was to maximize the spectral efficiency of the system. An improvement in spectral efficiency however does not necessarily lead to an improvement in user experience. If we trade spectral efficiency against transfer delay, user experience might actually degrade despite the increase in spectral efficiency.

In this paper, we do not focus on spectral efficiency, but on an improvement of a user’s throughput. We propose a so-called user-centric transmit mode selection, which selects the transmit mode that maximizes the user’s throughput. This throughput maximization is at the expense of spectral efficiency. We analyze this trade-off using a lightweight analytical model and present numerical results for a sample scenario. We show that network operators can tune the transmit mode selection algorithm such that a balance between spectral efficiency and user throughput is achieved.

Section II describes our high-level view on coordinated transmission. Section III presents the different transmission modes and introduces our notation. Section IV proposes our user-centric transmit mode selection scheme. Section V describes our evaluation scenario and presents numerical results.

II. AN ABSTRACT VIEW ON COORDINATED TRANSMISSION

In general, a coordinated data transmission (i.e. a transmission using a CS/CP CoMP scheme) is preceded by a setup or preparation phase. This setup phase consists of a measurement step and a coordination step. In the measurement step, user terminals measure and report the channel characteristics towards their serving cell and possibly also towards neighbor cells. In the coordination step, the basestations use these measurements to coordinate resource allocation for their
users. The challenge here lies in finding suitable resource allocations and transmission parameters which reduce inter-cell interference and maximize system throughput. See [3] for an overview of different objective functions and solutions.

In this paper, we abstract from the particular coordination algorithm and characterize a CoMP scheme solely by the following three parameters:

**Coordination setup time** $\tau$. Channel measurements and signaling for inter-basestation coordination take a certain time. This time is variable and depends on the channel measurement type, the propagation delay on backhaul links and the coordination algorithm. We denote this setup time as $\tau$.

**Coordination gain** $G$. The coordination of the resource allocation yields a better SINR for coordinated transmissions. We model this gain by adding a constant offset $G$ to a user’s current SINR. Because the Shannon capacity is a logarithmic function, users with a bad SINR benefit more from coordinated transmission than users that already experience a good SINR. This corresponds to the behavior in a real system, where cell-edge users benefit more from coordinated transmission than cell-center users.

**Coordination overhead factor** $\eta$. The use of coordinated transmission comes at a certain cost. First, there is a need for uplink resources to transmit channel measurement results. Second, signaling messages between basestations consume resources on backhaul links. Third, there is cost in terms of constraints imposed on resource allocation in serving and neighbor cells. These constraints are the outcome of the coordination algorithm and normally lead to an increase in system throughput. However, in some cases, the system state changes too quickly and the coordinated resources cannot be used as intended. As indicated before, this happens if a user’s buffer at the basestation ran empty while coordination among the basestations was ongoing.

Because it is unclear how to weight backhaul or uplink resources against downlink resources, we neglect uplink and backhaul overhead and measure coordination overhead in terms of downlink resources only. As in [7], we assume that a basestation coordinates resources with its neighbors as long as the user’s buffer is not empty. When the buffer runs empty, no more resources are allocated to this user and the coordination stops. Due to the non-zero setup time $\tau$, there is a certain time between the instant the buffer runs idle (and hence the end of the data transmission) and the release of all constraints on resource allocation. During this time, resource allocation in serving and neighbor cells is still constrained, which might lead to a loss in spectral efficiency compared to an unconstrained resource allocation.

To what extent periods with constrained resource allocation degrade spectral efficiency depends on many factors. If there are many users, there might be no degradation at all because the basestations are able to reuse the constrained resources for other users. If there are few users and the constraints largely restrict the number of allowed precodings, the constrained resources might not be used at all because the basestation cannot serve any user with the remaining set of precodings. To model the variable influence of these constraints, we introduce the overhead factor $\eta \in [0;1]$. An overhead factor $\eta = 0$ means that all resources can be fully reused and periods with constrained resource allocation have no negative impact. An overhead factor $\eta = 1$ means that all constrained resources cannot be used at all.

### III. Transmission modes and spectral efficiency

We distinguish between different **uncoordinated** and **coordinated** transmission modes. We characterize each mode by the following characteristic functions:

- $\chi(s, c)$ Number of resource units (symbols) required to transfer an object of size $s$ at SINR $c$, without any overhead such a transmission might cause.
- $\omega(s, c)$ Overhead in number of resource units that is caused in serving or neighbor cells (see II).
- $\psi(s, c)$ Transfer time of an object of size $s$, including potential waiting times.

We detail these functions for the different modes further below. Another important parameter is the (constant) bandwidth $r$ a basestation allocates to a user. This bandwidth can also be interpreted as a rate of resource units or symbols.

### A. Reference transmission modes

We consider two reference transmission modes. Our baseline is the **uncoordinated** transmission, which could be a transmit diversity scheme and does not require any coordination with neighbor basestations. The number of resource units required to transfer an object of size $s$ at SINR $c$ is:

$$\chi^{(\text{uncoord})}(s, c) = s/\gamma(c)$$

The function $\gamma(c)$ yields the spectral efficiency at SINR $c$. For simplification, we assume that the SINR is constant over the whole duration of the object transfer. We further treat interference as white noise and use the Shannon equation to determine the spectral efficiency, i.e. $\gamma(c) = \log_2(1 + c)$.

The perfectly coordinated transmission denotes the upper bound on spectral efficiency where coordination among basestations is ideal and happens in zero time:

$$\chi^{(\text{perfect})}(s, c) = s/\gamma(c + G)$$
Please note that overhead $\omega(s, c)$ is zero for both reference modes. The transfer time for both reference modes is the ratio of the number of resource units and the allocated rate:

$$\psi^{\text{(coord/perfect)}}(s, c) = \frac{\chi^{\text{(coord/perfect)}}(s, c)}{r}$$

(3)

B. Delayed transmission with precedent coordination

In [7], [10], we proposed two different ways to govern a coordinated transmission, denoted as delayed transmit mode and immediate transmit mode. We have shown that delayed mode provides better spectral efficiency than the other scheme at the cost of increased transfer delay. For lack of space, we restrict our analysis in this paper to the delayed transmission mode. An application of this analysis to immediate mode however is straightforward.

In delayed mode, the base-station buffers incoming data until the coordination process with neighboring base-stations has been completed. The data is then transmitted on the coordinated resources at a better SINR compared to a transmission on uncoordinated resources.

For small objects, the overhead caused by coordination might outweigh the spectral efficiency gains from the coordinated transmission of this object. To maximize spectral efficiency, we propose to use coordinated transmission only for objects larger than a certain selective coordination threshold $S_{\psi}$ [10]. If an object is smaller than $S_{\psi}$, the base-station does not trigger a coordinated transmission, but transfers the object in uncoordinated mode\(^1\). The characteristic functions of the delayed transmit mode and threshold $S_{\psi}$ are:

$$\chi^{\text{(del)}}(s, c) = \begin{cases} s/\gamma(c) & \text{for } s \leq S_{\psi} \\ s/\gamma(c+G) & \text{otherwise} \end{cases}$$

(4)

$$\omega^{\text{(del)}}(s, c) = \begin{cases} \frac{s}{\gamma(c)} & \text{for } s \leq S_{\psi} \\ 0 & \text{otherwise} \end{cases}$$

(5)

$$\psi^{\text{(del)}}(s, c) = \begin{cases} \frac{s}{\gamma(c)} & \text{for } s \leq S_{\psi} \\ \frac{s}{\tau + \frac{s}{\gamma(c+G)}} & \text{otherwise} \end{cases}$$

(6)

$$S_{\psi} = \frac{r\tau}{\frac{1}{\gamma(c)} - \frac{1}{\gamma(c+G)}}$$

(7)

\(^1\)Transmit mode selection might depend on several additional factors, such as user location or terminal capabilities. The influence of these other factors however is out of scope of this paper.

IV. USER-CENTRIC SELECTIVE COORDINATION

The delayed transmit mode improves spectral efficiency at the expense of an additional transfer delay $\tau$. This delay reduces the throughput experienced by a user. We calculate a user’s throughput as the ratio between the object size $s$ and the transfer duration for this object. The transfer duration is measured from the arrival of the object at the base-station until the last byte has arrived at the user’s terminal:

$$\Gamma^{\text{(mode)}}(s, c) = \frac{s}{\psi^{\text{(mode)}}(s, c)}$$

(8)

Figure 2 illustrates this effect for a constant SINR, an arbitrary selective coordination threshold $S_{\psi} = 100$ bytes and different bandwidths $r$. For $r = 10^5$ Hz, throughput is decreased only slightly and only for a small range of object sizes around 100 to 200 bytes. However, for larger bandwidths around $r = 10^6$ Hz, throughput remains significantly below the level of uncoordinated transmission for a wide range of object sizes. Unfortunately, 99% of all objects of Internet data traffic fall in this range [8]. From a user perspective, throughput with coordinated transmission would actually be worse than without CoMP, although the system’s spectral efficiency might be improved. Please note that this effect is independent of overhead factor $\eta$.

To ensure that coordinated transmission schemes do not degrade user throughput, we propose to use coordinated transmission only for objects larger than a threshold object size $S_{\Gamma}$. To determine $S_{\Gamma}$, we set

$$\Gamma^{\text{(del)}}(s, c) > \Gamma^{\text{(uncoord)}}(s, c)$$

(9)

and solve for $s$. Using Eq. (8), Eq. (1) and Eq. (6), we get

$$S_{\Gamma} = r\tau \left( \frac{1}{\gamma(c)} - \frac{1}{\gamma(c+G)} \right)^{-1} = \frac{1}{\eta} \cdot S_{\psi}$$

(10)

Figure 3 compares the thresholds $S_{\psi}$ and $S_{\Gamma}$ over SINR $c$ and the product $r\tau$. Figure 3 and Eq. (10) show that $S_{\Gamma}$ is always larger or equal to $S_{\psi}$. The smaller the overhead factor $\eta$, the larger is the difference between $S_{\Gamma}$ and $S_{\psi}$.

V. PERFORMANCE EVALUATION

Network operators can trade off spectral efficiency against user throughput by choosing a selective coordination threshold.
in the interval \([S_0; S_T]\). We illustrate this trade-off in a numerical example. For our evaluation, we use the analytic model which we initially proposed in [10]. For the sake of completeness, we reproduce our model here.

A. Wireless system model

We consider a single basestation which sequentially transfers objects to its users (Figure 5). The basestation serves only one user at a time. After the basestation has completed the transfer of an object to one user, it immediately starts the transfer of a new object to a random user. We assume that system bandwidth \(r\) is constant. The basestation always allocates the whole bandwidth. Please note that this model describes a wireless network with rate-resource allocation.

The size of an object \(s\) is random, with probability density function (pdf) \(f_s(x)\). The channel quality \(c\) is the long-term SINR of a user, which is assumed constant over the whole duration of an object transfer. We treat interference as noise and use the Shannon equation for the mapping of SINR to spectral efficiency. We assume an infinite number of users and model the channel quality as a random variable with pdf \(f_c(x)\).

Using the characteristic functions introduced in section III, Eq. (8) determines the user throughput of an object transfer. The spectral efficiency \(\varrho\) of an object transfer is

\[
\varrho(s, c) = s / \left( \chi(s, c) + \omega(s, c) \right)
\]

Because the spectral efficiency is a function of \(s\) and \(c\), we can calculate the expected value of \(\varrho\) by the law of the unconscious statistician [11]:

\[
E[\varrho] = \int d(s, c) f_{SC}(s, c) ds dc
\]

where \(f_{SC}(s, c)\) is the joint pdf of \(s\) and \(c\). Because the random variables \(s\) and \(c\) are independent of each other, \(E[\varrho]\) becomes

\[
E[\varrho] = \int d(s, c) f_s(s) f_c(c) ds dc
\]

Equation (13) gives the average spectral efficiency of an object transfer. If we are interested in the time-average of the spectral efficiency of our system, we have to derive an expression for the spectral efficiency at an arbitrary instant \(t_0\). If we pick an arbitrary \(t_0\), we have a higher chance of observing object transfers with a long duration, e.g., the transfer of a large object at low SINR. This bias with the transfer duration is a result from renewal theory, known as the waiting time paradoxon [12]. Our system model constitutes a renewal process, with the recurrent events being the time instances at which the transmission of an object ends. The time an object transfer occupies the wireless channel is \(d(s, c) = \chi(s, c) / r\). The probability that an arbitrary time instant \(t_0\) falls in an object transfer of duration \(d\) is [12]:

\[
f_D[t](d) = K \cdot d \cdot f_D(d)
\]

where \(K = 1 / E[d(s, c)]\), and pdf \(f_D(d)\) is the unbiased pdf of \(d\). In the following, we use the notation \([t]\) to distinguish between biased and unbiased values.

By application of equation (14) and the law of the unconscious statistician, we get for the expected value of \(d[t]\):

\[
E[d[t]] = \int d(s, c) f_{ST}[c|t](s, c) ds dc
\]

With this result, we get the following expression for the time-average of the spectral efficiency of our system:

\[
E[\varrho[t]] = \int \varrho(s, c) f_{ST}[c|t](s, c) ds dc
\]

For a more in-depth discussion of this model and a validation with simulation results we refer to [10].

B. Scenario and parameters

Wireless system and coordinated transmission. We use an empirical distribution of SINR \(c\), gathered from wideband SINR measurements in a system-level simulation of a 3GPP Case 1 scenario with 3D antenna patterns, 19 sites, pathloss and shadowing (no fast fading) according to [1]. The cdf of our SINR shows a good fit with Fig. A.2.2-2 in the system simulator calibration section of [1]. Because we use the Shannon equation instead of LTE transport formats, we clipped the SINR at 22 dB. To model the improvement in SINR from coordinated transmissions using a CS/CP CoMP scheme, we add a constant offset of \(G = 6\) dB to a user’s current SINR. This offset roughly corresponds to a cancellation of the first up to second most significant interferers in our 3GPP Case 1 3D scenario. For the coordination setup time \(\tau\), we consider \(\tau = 10\) ms as our minimum value. 3GPP documents specify an average backhaul delay of 10 ms between LTE macro cells as a realistic value [13].

Data traffic. Most applications on mobile devices use HTTP and HTTPS for data transfer [9], [14]. This comprises web surfing using a browser, so-called smartphone apps, the AppStore or Android Market and even mobile video applications. In total, HTTP and HTTPS based applications account for more than 80% of the traffic volume on mobile devices [9], [14]. We therefore restrict our model to objects sent over HTTP(S). The distribution of HTTP(S) object sizes is heavy-tailed [8], [9]. We use the HTTP(S) response size distribution published in [8]. It provides an empirical object...
size distribution function which is a mixture of three log-normal distribution functions. We set the range of valid object sizes to 10 up to 10^3 bytes.

C. Results on user throughput

Figure 4 plots the cumulative distribution function (cdf) of the user throughput \( \Gamma(s,c) \) for a bandwidth \( r = 10^6 \) Hz and a coordination setup time \( \tau = 10 \) ms. The three curves depict the delayed transmit mode without any selective coordination threshold (dotted line), with selective coordination threshold \( S_\rho \) (dashed line) and with user-centric selective coordination threshold \( S_T \) (solid line). The throughput cdf for coordination threshold \( S_\rho \) depends on the overhead factor \( \eta \). Figure 4 shows the result for \( S_\rho \) at \( \eta = 0.1 \). For \( \eta = 0 \), \( S_\rho \) becomes 0 and the throughput cdf of \( S_\rho \) converges to the dotted line. For \( \eta = 1 \), \( S_\rho \) converges to the solid line.

Especially for small overhead factors \( \eta \), user throughput differs significantly, depending on whether \( S_\rho \) or \( S_T \) is used as selective coordination threshold. Without coordination threshold, the throughput for more than 60% of the object transmissions is below 1 Mbps. With \( S_T \), less than 10% of the transmissions experience a throughput which is that low. The throughput cdf with \( S_T \) also is more steep then the other cdf curves, which indicates better fairness. Expressed in terms of Jain’s fairness index \([15]\), \( S_T \) yields a fairness index of 0.80, \( S_\rho \) with \( \eta = 0.1 \) yields 0.73 and without selective coordination threshold, the fairness index only is 0.43.

Figure 6 compares the average user throughput with coordinated transmissions to an uncoordinated reference system (i.e. a system with only uncoordinated transmissions). The solid line is the ratio of the average user throughput with user-centric selective coordination threshold \( S_T \) over the average user throughput in an uncoordinated system. The dashed lines depict the same ratio for selective coordination threshold \( S_\rho \) for different coordination overhead parameters \( \eta \). Because the average throughput depends on the product of coordination setup time \( \tau \) and bandwidth \( r \), we depict the gains over the product \( r\tau \). We cover a wide range of values: The lower end of the x-axis in Figure 6 could be a system with a small setup time \( \tau = 10 \) ms and a user allocated bandwidth of 10 kHz \((r\tau = 10^2 \) Hz\)). The upper end could be a system with a large setup time \( \tau = 1 \) s and a large user allocated bandwidth of 100 MHz \((r\tau = 10^3 \) Hz\)).

From Figure 6, we can make the following observations: With \( S_T \), the average throughput is always larger or equal to the average throughput in an uncoordinated system, which is the expected behavior. While for small \( r\tau \), the average throughput with \( S_T \) and \( S_\rho \) is larger than the average throughput with uncoordinated transmission, as soon as \( r\tau > 10^4 \) Hz\)), the average throughput with \( S_T \) and uncoordinated transmission are equal. This is because the waiting time for coordination setup becomes large compared to the object transmission time as \( r\tau \) increases. In other words, the higher bitrate of a coordinated data transmission cannot compensate the waiting time of the coordination setup phase, at least not for our current heavy-tailed object size distribution.

Another observation from Figure 6 is, that if an operator optimizes for spectral efficiency and selects coordination threshold \( S_\rho \) instead of \( S_T \), the average user throughput can become significantly worse than in an uncoordinated system. There is only a small degradation of the average throughput as long as overhead factor \( \eta \) is close to 1, but it gets worse the smaller \( \eta \) becomes.

D. Results on spectral efficiency

Figure 7 depicts the gain respective loss in average spectral efficiency \( E[\varphi[t]] \) of coordinated transmission using the delayed transmit mode \([10]\). The horizontal lines \( E[\varphi[t]](\text{perfect}) \) and \( E[\varphi[t]](\text{uncoord}) \) indicate the reference values for uncoordinated and perfectly coordinated transmission. The dotted lines in Figure 7 depict the spectral efficiency without selective coordination threshold for different overhead factors \( \eta \). The solid lines depict the spectral efficiency if \( S_\rho \) is used as selective coordination threshold. For \( \eta = 0 \), spectral efficiency is constant and equal to the upper bound. For \( \eta > 0 \), spectral efficiency degrades as \( r\tau \) becomes larger. The reason for this degradation is the coordination overhead \( \omega(s,c) \), which increases with \( r\tau \), while the number of resources required for the object transfer, \( \chi(s,c) \), is independent of \( r\tau \). By application of selective coordination threshold \( S_\rho \), we cannot eliminate this degradation, but we can avoid that the spectral efficiency becomes worse than our uncoordinated reference.
system. We can thus observe that, in contrast to our previous observation on average user throughput, spectral efficiency gets worse the larger $\eta$ becomes.

Regarding the trade-off between user throughput and spectral efficiency, Figure 8 shows the ‘cost’ in spectral efficiency an operator has to pay to improve user throughput. It plots the ratio of $E_{[\eta]}[S_T]$ over $E_{[\eta]}[S_e]$, over a range of $r\tau$ values and for different overhead factors $\eta$. If $S_T$ is used instead of $S_e$, the loss in spectral efficiency is negligible as long as $r\tau$ is small ($< 10^4$ Hz s), but user throughput is significantly improved (compare Figure 6). For very large $r\tau$ values ($> 10^6$ Hz s), the loss in spectral efficiency when using $S_T$ very much depends on the overhead factor $\eta$. The smaller $\eta$, the larger is the difference between $S_T$ and $S_e$ and the more ‘expensive’ (in terms of spectral efficiency) a user-centric selective coordination becomes. The loss is largest for $\eta = 0$ and almost attains 50% at large $r\tau$ values. However, already for $\eta = 0.1$, the loss is in the range between 0 and 15% and it gets smaller the larger $\eta$ becomes. Hence, if we do not consider $\eta$ as a measure for the coordination overhead, but as a scaling factor for $S_T$ (compare Eq. (10)), we can conclude that a selective coordination threshold of $\approx 0.1 \cdot S_T$ already significantly improves user throughput and fairness at the cost of only a small loss in spectral efficiency.

VI. CONCLUSION

The use of coordinated multi-point transmission schemes in the downlink of an LTE-like system comes at a certain cost. There is overhead generated on backhaul links, the uplink, and, in case of bursty data traffic, possibly also in the downlink. In addition, pending transmissions might be deferred until resource allocation for this transmission has been coordinated with neighbor basestations. This waiting time can lead to a degradation in the user-perceived throughput, despite the higher bitrate of the CoMP-enabled transmission.

We proposed a user-centric transmit mode selection scheme which avoids this throughput degradation at the expense of spectral efficiency. We evaluated the effectiveness of our transmit mode selection and quantified the loss in spectral efficiency by comparison to our previous results [10]. We have shown that operators can trade-off spectral efficiency (increased capacity) against user throughput (service quality) by an appropriate choice of our selective coordination threshold.

We currently work on a transmit mode selection scheme which considers both, user throughput and object transfer time. For highly interactive applications, object transfer time might be the more significant metric to evaluate user experience. When using TCP, the object transfer time might be more constrained by the round-trip-time than the bitrate, which is not yet reflected in our model.

ACKNOWLEDGEMENT

The author wants to thank the anonymous reviewers for their many helpful comments and suggestions.

REFERENCES