Downlink transmission in multi-carrier systems with reduced feedback

Wang, Yuanye; Pedersen, Klaus; Sørensen, Troels Bundgaard; Mogensen, Preben

Published in:
IEEE Vehicular Technology Conference. Proceedings

DOI (link to publication from Publisher):
10.1109/VETECS.2010.5493844

Publication date:
2010

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Abstract— in this paper we address the problem of reducing the feedback for the downlink transmission in multi-carrier systems. In these systems multiple Component Carriers (CCs) are aggregated together to form a wide spectrum. Consequently, a large feedback overhead is required to report the channel quality information over such a wide bandwidth. We first generalize two existing feedback reduction techniques, and then propose a new one. These techniques use different feedback schemes across the CCs, or allow some CCs to be un-reported, for the purpose of reducing the amount of feedback. Performance for the investigated techniques is evaluated in a realistic system setting with different traffic conditions and terminal categories. Based on the obtained results, the selection of proper feedback reduction technique under different load conditions can be made. The findings are useful for reducing the feedback overhead in future generation wireless communication systems that operate over multiple CCs.

I. INTRODUCTION

Channel-aware packet scheduling and link adaptation have been used in many systems, which efficiently improve the system performance as compared to the case with no knowledge of the channel quality [1]-[3]. A feedback link and transmission resources are required to provide the transmitter with the Channel Quality Indicator (CQI). In multi-carrier systems, multiple Component Carriers (CCs) are aggregated to form a wide bandwidth, with independent physical layer transmission carried out within each CC [4][5]. If full CQI is reported on each CC, it may lead to very high overhead, and may be infeasible for portable devices. Certain feedback reduction techniques can limit the feedback with only limited reduction of the system performance.

We categorize the feedback reduction techniques into two groups. The first group is based on compression of the feedback information, e.g. CQI quantization, which reports a quantized discrete value instead of the full channel state [6]-[8], or a discrete cosine transform based compression [9]. In Orthogonal Frequency Division Multiple Access (OFDMA) based systems, because a user is most likely to be scheduled on the sub-carriers with good channel quality, a scheme that reports only the best resource blocks is shown to give good performance [10]. This scheme is known as best-M. It can be implemented in several ways, e.g. individual best-M, which feeds back the CQI value for each of the selected blocks, or average best-M, which feeds back the average CQI for the selected ones. The latter has been adopted by the 3rd Generation Partnership Project (3GPP) for the Long Term Evolution (LTE) systems [11][12]. The feedback can also be reduced by increasing the feedback interval when mobility is low [13]. The other group of feedback reduction techniques works through Layer-3 CC selection, which selects the proper CC(s) for the frequency selective CQI feedback of each user. It can therefore only be used in multi-carrier systems. In our previous work [14], we have shown that by assigning each user with a subset of the available CCs, the feedback can be significantly reduced while still maintaining good performance. Users feed back CQI information pertaining only to the selected CC, whereas the remaining CCs are either not reported, or reported indirectly through a wideband CQI measure. In this study, we extend the investigations in [14] by studying more advanced CQI reduction techniques under different load conditions. The backward compatible mode with a mixture of terminal categories, i.e. the legacy users which operate on a single CC and the advanced users which support multi-CC transmission, is also studied.

In this paper, we adopt the average best-M method to compress the feedback within each CC, and devote our effort to the cross-CC feedback reduction. We first generalize the two techniques developed in [14], and then propose a new technique that achieves better performance when cell load is high. As a case study, the performance for these techniques is evaluated in a downlink LTE-Advanced system with feedback in the uplink. Based on the obtained results, a recommendation of the proper feedback reduction technique under different load conditions is given in the end.

The paper is organized as follows: Section II describes the feedback reduction techniques under investigation; Section III describes the simulation methodology and assumptions; In Section IV, the performance for the investigated feedback reduction techniques is evaluated and compared against each other with various load conditions and terminal categories. Finally, Section V concludes the paper.

II. SYSTEM MODEL AND FEEDBACK REDUCTION TECHNIQUES

A. System model of LTE-Advanced downlink transmission

A typical wireless communication system with feedback link is depicted in Fig. 1, where the receiver estimates the channel quality and then sends it back to the transmitter to facilitate the channel-aware packet scheduling and Link Adaptation (LA).

Fig. 1. A downlink wireless communication system with channel information feedback and channel-aware packet scheduling, link adaptation.

According to the 3GPP standard [15], there are 50 Physical Resource Blocks (PRBs) in a 10 MHz bandwidth, each PRB equals 12 consecutive sub-carriers (each of 15 kHz bandwidth). Because average best-M is used for CQI reporting, following [11], three neighboring PRBs are grouped together and associated with one CQI value. $Q_{CQI}$ bits are used to quantize each CQI value. With $Q_{CQI} = 4$ bits quantization and 25 dB dynamic range, we obtain a quantization step size of 1.6 dB [15][16]. Various techniques for reducing the amount of feedback information have been developed. The ones used in this study are described hereafter.
B. CQI compression

The compression of CQI feedback has been extensively studied in traditional systems without carrier aggregation [6]-[10], [12]-[15]. From the various compression techniques, we chose average best-M for our study because of its superior performance. The best-M scheme used in our study works as follows: The user first estimates the channel quality, and then calculates the geometrically averaged CQI value of the best-M CQI groups. This average CQI is encoded differentially using $Q_{best-M} = 2$ bits relative to the wideband CQI $W$ by the mapping:

$$\Delta_{best-M} = f(CQI_{best-M} - W)$$

$W$ is obtained by averaging the CQIs across the whole CC bandwidth. The one-sided and non-linear mapping $f$ from the offset level to the differential CQI value $\Delta_{best-M}$ is defined in Table-I.

<table>
<thead>
<tr>
<th>Differential CQI value</th>
<th>Index offset level</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>≤ 1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>≥ 4</td>
</tr>
</tbody>
</table>

The feedback should contain the indices corresponding to the offset level, the wideband CQI, and the position of the selected CQI groups. The total number of bits required for best-M within each CC is:

$$Q_{best-M} = Q_A + Q_{CQI} + \left\lfloor \log_2 \left( N_{CQI} M \right) \right\rfloor$$

(1)

where $\left\lfloor \cdot \right\rfloor$ is the ceil function and $N_{CQI}$ is the total number of PRB groups.

Because each user selects the CQI groups based on its own channel condition, there is a risk that all users experience good channels on the same CQI groups, which will lead to a potential conflict in packet scheduling. To overcome this problem we use a simple approach to estimate the average CQI value $\Delta$ for the unreported PRB group(s) based on the best-M feedback information:

$$\Delta = W - M\Delta_{best-M} Q_A / (N_{CQI} - M)$$

(2)

With information about $\Delta$ the packet scheduler has full freedom to schedule all PRB groups.

C. Layer-3 component carrier selection at base station or user side

For a user transmitting over $U$ CCs, the total number of feedback bits calculated from (1) has to be raised by a factor of $U$, if the same feedback scheme is used in each CC. This creates very high feedback overhead and may require too much transmit power, especially for cell-edge users. There is therefore good incentive to limit the number of scheduled CCs, for which feedback is required, provided this can be done without too much system and user performance degradation. Consequently, Layer 3 CC selection plays a major role in reducing the amount of feedback overhead.

1) Balanced CC load at base station

The Balanced CC Load (BCL) algorithm is based on the CC selection technique in [14] (Round Robin CC selection). While each user is assigned with only 1 CC in [14], here we generalize it to allow $U$ CCs being selected for each user. The target is to balance the load (in terms of number of active users) with respect to the bandwidth for each CC. In the case when all CCs have the same size, it leads to an even distribution of users over the available CCs. For the purpose of load balancing, BCL should be carried out at the base station, which knows the selected CCs for each user.

Because the CC assignment for the users is fixed, only the CQI values for the assigned CCs are needed. With equal CC bandwidth this gives

$$Q_{BCL} = UQ$$

(3)

where $Q$ is the number of bits required within each CC, e.g., $Q_{best-M}$, when average best-M is used.

2) Best-CC & Wideband CQI at user side

The Best-CC & Wideband CQI (B&W) algorithm can be considered as a generalized scheme of the CQI compression developed in [14]. It dynamically selects the best $U$ CCs and feeds back the frequency selective CQI values for these CCs. The remaining CCs are indicated with only the wideband CQI measure. The B&W algorithm is simple to implement because the selection is performed at the user side in a distributed manner, with no additional feedback or coordination on the selections made by other users.

Among various selection metrics, we have found that the wideband CQI gives good performance. In case when the CCs have the same interference condition, wideband CQI based selection eventually leads to same number of users on each CC. However, if the CCs are of different sizes or have different interference conditions, a scaling of the wideband CQI may be required to distribute users efficiently across the CCs.

The feedback with B&W is constituted with three parts: the bits indicating the selected CCs, the frequency selective CQI feedback for the best CCs and the wideband CQI for each remaining CCs. Assume equal bandwidth for the CCs, the CQI word-size can be expressed as

$$Q_{B&W} = \left\lfloor \log_2 \left( \frac{N}{U} \right) \right\rfloor + UQ + Q_{CQI}(N-U)$$

(4)

where $N$ is the total number of aggregated CCs in the system.

3) Best-CC at user side

The proposed best-CC (BCC) is a simplification of B&W. It allows a user to operate on the best CCs only, and consequently report CQI information for these specific CCs. The motivation for this simplification is:

1. To remove the overhead for sending the wideband CQI.

2. The wideband CQI appears at the transmitter as a degenerated version of the frequency selective CQI values. It may be beneficial to schedule resources to a user with slightly lower but more accurate CQI values, than to a user with an inaccurate wideband CQI value. If this is the case, the additional feedback of wideband CQI on the worst CCs brings no gain on system performance.

The BCC algorithm requires only the first two components of the feedback information in (4), i.e.

$$Q_{Best-CC} = \left\lfloor \log_2 \left( \frac{N}{U} \right) \right\rfloor + UQ$$

(5)
Normally, the more number of CCs that have frequency selective CQI feedback, the better is the system performance. Meanwhile, the feedback overhead also becomes higher. The number of feedback bits required for the different schemes as a function of \( U \) can be seen in Fig. 4.

III. SIMULATION METHODOLOGY AND ASSUMPTIONS

The performance of the feedback reduction techniques is evaluated in a quasi static downlink multi-cell system level simulator that follows the LTE specifications defined in [17], including detailed implementations of Layer-3 CC selection, Layer-2 packet scheduling, Hybrid Automatic Repeat Request (HARQ) and LA functionalities. The simulation scenario is Macro-cell case #1 as defined in [15]. The simulation parameters are summarized in Table II. The link to system mapping is based on the exponential effective metric model [18].

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting / description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test scenario</td>
<td>3GPP Macro-cell case #1 (19 sites, 3 cells per site)</td>
</tr>
<tr>
<td>Aggregation pattern</td>
<td>4 CCs at 2.0 GHz, 10MHz per CC</td>
</tr>
<tr>
<td>Number of PRBs</td>
<td>50 per 10 MHz bandwidth (12 subcarriers per PRB)</td>
</tr>
<tr>
<td>Sub-frame duration</td>
<td>1 ms (11 OFDM data symbols plus 3 control symbols)</td>
</tr>
<tr>
<td>Modulation and coding schemes</td>
<td>QPSK (1/3 to 3/4)</td>
</tr>
<tr>
<td></td>
<td>16-QAM (2/5 to 5/6)</td>
</tr>
<tr>
<td></td>
<td>64-QAM (3/5 to 9/10)</td>
</tr>
<tr>
<td>User receiver</td>
<td>2-Rx Interference Rejection Combining</td>
</tr>
<tr>
<td>HARQ modeling</td>
<td>Ideal chase combining with maximum 4 retransmissions.</td>
</tr>
<tr>
<td>Feedback interval &amp; delay</td>
<td>Feedback every 5 ms, with 6 ms delay.</td>
</tr>
<tr>
<td>Channel Quality Indicator (CQI) reporting</td>
<td>1 CQI per 3 PRBs; 1.6 dB quantization step; log normal error with 1 dB standard deviation; default reporting mode: frequency selective CQI feedback using Best-M, M=5 per 10 MHz bandwidth.</td>
</tr>
<tr>
<td>Layer-2 packet scheduler</td>
<td>Time domain: Round robin (maximum 20 users selected)</td>
</tr>
<tr>
<td></td>
<td>Frequency domain: Cross-CC proportional fair (with user throughput collected from all assigned CCs)</td>
</tr>
<tr>
<td>1st transmission BLER target</td>
<td>10%</td>
</tr>
<tr>
<td>Traffic type</td>
<td>Finite buffer with Poisson arrival and various payload sizes. Maximum 50 users per cell.</td>
</tr>
</tbody>
</table>

For the three CC selection techniques, BCL operates based only on the load condition. It assigns \( U \) CC(s) to each user when it arrives. The CC assignment remains unchanged until the user finishes the transmission and departs from the system. B&W and BCC require the knowledge of wideband CQI for the dynamic CC selection. The selection is assumed to be updated at the same rate as the CQI estimation, i.e. every 5 ms. Simulation campaigns are conducted as one long simulation run (up to 200 seconds) with Poisson traffic arrival in a dynamic birth-death process. When a user arrives (birth), a fixed payload is being transmitted, and when finished the user departs from the system (death). The offered load is given by the product of the payload size and the number of user arrivals per second.

As to performance indicators, we mainly look into the cell-edge user throughput, which is defined as the 5th percentile worst user throughput over the simulated users. The average user throughput, according to [20], is mainly determined by the offered load and the cell capacity constraint, but insensitive to the feedback reduction techniques.

IV. PERFORMANCE WITH FEEDBACK REDUCTION

In this section, we evaluate the performance for each of the Layer-3 CC selection techniques in Section II.C when combined with the best-M CQI compression scheme in Section II.B.

A. Performance with maximum one CC using best-M

The cell-edge user throughput for \( U=1 \) with different arrival rates is shown in Fig. 2: maximum one CC is reported using the best-M scheme. The corresponding average number of active users with the different techniques is shown in Fig. 3. The case with best-M in all CCs \((U=4)\) is taken as a reference. Note that each user is assumed to have a fixed payload size of 2 Mbits.

![Fig. 2. Cell-edge user throughput with various Poisson arrival rates. Payload size is 2 Mbits. For cross-CC CQI reduction, maximum one CC is reported using the best-M CQI compression scheme.](image)

From Fig. 2 we can see that:

1. At low cell load, the performance of B&W is close to the reference using best-M in all CCs (~10% lower). The reason is that with less number of users, each user tends to transmit with a wide bandwidth. Wideband CQI is sufficient in this situation. In fact, when the load is very low, wideband CQI for all CCs (B&W with \( U=0 \)) is able to provide good performance. BCC has the worst performance because each user is scheduled on only one CC, and hence there is a high possibility that at least one CC is unused. In this case, the system will suffer from poor trunking efficiency. If \( K \) denotes the number of users in the system, a simple estimate of this probability is

\[
P = \left( 1 - (1 - (1 - U / N)^K)^V \right)^{KU} \geq N, \quad 1, \quad \text{elsewhere}
\]

When \( N=4, U=1, K=5 \), the calculated empty probability is 66%. With a higher number of users, or more CCs assigned to each user, the probability reduces rapidly and the performance of BCC will be significantly improved.

2. At high cell load, B&W cannot fully exploit the frequency domain diversity, and it offers the worst performance among the investigated techniques. BCC, on the other hand, outperforms the others. Its performance is even better (13%~14% higher) than using best-M in all CCs, because BCC results in only the best users on a certain CC. In this sense, BCC can be considered as a time domain channel-aware packet scheduler, which improves the scheduled user SINR condition, while
maintaining allocation fairness among the users (because each user selects the best $U$ CCs).

From the average number of active users that is shown in Fig. 3, we can clearly see that as the arrival rate increases, more and more users will remain active in the system, offering higher multi-user diversity. However, when the arrival rate is so high that the system cannot support, the admission control will put a hard limit on the number of active users, which is 50 in our simulation. The switching point for using B&W or BCC happens at 18 arrivals per second (Fig. 2), which corresponds to an offered load of 36 Mbps and 12~13 active users in a cell.

**B. Performance with higher number of CCs using best-M**

As discussed in Section II.C, it is possible to improve the downlink throughput by allowing more CCs to be reported with best-M ($U>1$). Meanwhile, the feedback overhead will also increase. For B&W, its performance at low load is already close to using best-M in all CCs ($U=4$), therefore it is not worth increasing the overhead to further improve throughput. However, BCL and BCC are expected to see a large performance improvement with increased $U$.

The feedback overhead for different techniques is shown in Fig. 4, from which we can see, when $U=2$, BCL and BCC have already higher overhead than B&W with $U=1$. As to cell-edge users, Fig. 5 shows that B&W with $U=1$ still offers the highest throughput at low load. Using BCC with $U=2$ is only beneficial when the number of arrivals is between 15 and 23 users per second. For even higher arrival rate, BCC with $U=1$ offers same performance at a lower feedback rate. For the sake of simplicity, in the following analysis we consider only the adaptation between B&W and BCC based on the load condition. For both schemes, only one CC needs to be reported with frequency selective CQIs ($U=1$). The rest are either not reported, or indicated with a wideband CQI. At medium load, it may be beneficial to have more than one CCs reported with frequency selective CQIs, at the cost of higher feedback overhead. This adaptation of $U$ based on load condition is left for future work.

The results shown before are obtained with a fixed payload size of 2 Mbits. When reducing the payload size, the same offered load can be achieved with a higher user arrival rate. Fig. 6 shows the performance with a smaller payload size of 1 Mbits. In this case, B&W is beneficial when the arrival rate is less than 36 users per second per cell. Beyond this level, BCC can be used. The average number of active users at the switching point is 13 to 15 for BCC and B&W, respectively. With both payload sizes, the switching point corresponds to an offered load of around 36 Mbps, when 13+ active users simultaneously exist in the system.

**C. Performance with a small payload size**

The feedback overhead for different techniques is shown in Fig. 4, from which we can see, when $U=2$, BCL and BCC have already higher overhead than B&W with $U=1$. As to cell-edge users, Fig. 5 shows that B&W with $U=1$ still offers the highest throughput at low load. Using BCC with $U=2$ is only beneficial when the number of arrivals is between 15 and 23 users per second. For even higher arrival rate, BCC with $U=1$ offers same performance at a lower feedback rate. For the sake of simplicity, in the following analysis we consider only the adaptation between B&W and BCC based on the load condition. For both schemes, only one CC needs to be reported with frequency selective CQIs ($U=1$). The rest are either not reported, or indicated with a wideband CQI. At medium load, it may be beneficial to have more than one CCs reported with frequency selective CQIs, at the cost of higher feedback overhead. This adaptation of $U$ based on load condition is left for future work.

The results shown before are obtained with a fixed payload size of 2 Mbits. When reducing the payload size, the same offered load can be achieved with a higher user arrival rate. Fig. 6 shows the performance with a smaller payload size of 1 Mbits. In this case, B&W is beneficial when the arrival rate is less than 36 users per second per cell. Beyond this level, BCC can be used. The average number of active users at the switching point is 13 to 15 for BCC and B&W, respectively. With both payload sizes, the switching point corresponds to an offered load of around 36 Mbps, when 13+ active users simultaneously exist in the system.

**D. Performance with mixed LTE-Advanced users and LTE-Rel’8 users, with CC selection only for cell-edge users**

In the results shown before, we assume all users are LTE-Advanced. However, an LTE-Advanced system is expected to operate in a backward compatible manner, which means both LTE-Advanced user and LTE-Rel’8 user may exist. Due to the hardware constraint, an LTE-Rel’8 user should only be scheduled on a single CC, whereas an LTE-Advanced user can potentially be scheduled on all CCs.

BCL offers worse performance than the other two techniques, therefore should only be used for the LTE-Rel’8
users who do not support multi-CC transmission. The feedback for the LTE-Rel’8 users cannot be further reduced by using any of the CC selection technique, because they are already using the scheme with the least amount of feedback bits. For LTE-Advanced users operating on several CCs, feedback reduction is required on the cell-edge to avoid running out of power. Using the same threshold value as in [14], we enable the feedback reduction (B&W or BCC) for LTE-Advanced users with path loss larger than 107 dB. LTE-Advanced users with smaller path loss values report using best-M in all CCs. The simulation result when half of the users are LTE-Rel’8 is summarized in Fig. 7.

![Fig. 7. Cell-edge user throughput with 50% LTE-Advanced and 50% LTE-Rel’8 users. Payload size is 2 Mbits for all users.](image)

From Fig. 7 we can see that, the trend of the different techniques is similar to before. However, the curves have come closer because the different feedback reduction techniques apply only for a small portion of the users (LTE-Advanced users at the cell-edge). For arrival rates below 22, B&W achieves more than 95% of the performance of full CQI reporting. For arrival rates beyond this level, BCC offers 4% more cell-edge user throughput than full CQI reporting.

V. CONCLUSION

In this paper, we have studied several techniques for feedback reduction across multiple Component Carriers (CCs). Our study is made with best-M for frequency selective CQI reporting, but the investigated feedback reduction techniques can be combined with any CQI compression method. We first generalize two existing algorithms to allow more flexibility in the trade-off between downlink performance and feedback overhead. A new algorithm called Best-CC (BCC) is proposed, which reports the best CCs only.

Based on the simulation results, we recommend Best-CC together with wideband CQI reporting (B&W algorithm) to be used when cell load is low, since it achieves approx. 90% of the cell-edge user throughput of best-M CQI reporting in all CCs (no CQI reduction), and at a CQI feedback reduction of 57%. BCC is suitable for high cell load, in the sense that it offers even better performance than full CQI reporting and with feedback reduced by 72%. At medium load, reporting the frequency selective CQI values for more than one CC may improve the downlink performance, but the feedback overhead is increased as well. In this situation, a compromise between the two should be made.

We also consider a case where 50% of the users are LTE-Rel’8 and the rest are LTE-Advanced users. The issue is here that LTE-Advanced users require feedback reduction to facilitate multi CC operation at the cell-edge. Using a simple threshold based on path loss, performance can be maintained close to full CQI reporting for both low and high cell load.

Although LTE-Advanced systems are used for performance evaluation, the findings and conclusions are in general valid in any OFDMA-based multi-carrier system, e.g. WiMax 802.16m.

ACKNOWLEDGMENT

The authors are grateful to Jens Steiner and Mads Brix of Nokia Siemens Networks for their valuable suggestions and help in carrying out this study.

REFERENCES