Decentralized User-Centric Scheduling with Low Rate Feedback for Mobile Small Cells

Wei Ni, *Member, IEEE*, Iain B. Collings, *Senior Member, IEEE*, and Ren Ping Liu, *Member, IEEE*

***Abstract*—Small cells with wireless backhaul are promising, whereas challenges of severe overlapping coverage and strong interference are yet to be addressed. Coordinating small cells could resolve the challenges; however, existing multicell coordinated beamforming techniques involve high cost of communication overhead, synchronization and backhaul. Such problems may deteriorate in practical cellular applications, where there could be many users, each generating high channel feedback overhead to compete for an opportunity of being scheduled, and the downlink data signals of the coordinated cells need to be precisely synchronized at each of the users. We propose a new scheme, which cuts the overhead by 80% and enables the coordination to be practically implemented in a decentralized manner. Our scheme is a user-centric downlink scheduling approach, where mobile terminals trigger and participate in the scheduling decisions of small cells. We also develop a new recursive algorithm to optimize the quantization levels of mobile terminals' feedback, minimizing the feedback requirement. Analysis, confirmed by simulations, shows that our scheme is able to achieve 94.4% of the sum-rate upper-bound which can only be approached by idealized centralized coordination. In terms of throughput, given the 80% reduced overhead, our scheme is 139.5% better than the idealized centralized coordination approach.**

***Index Terms*—Small cell; wireless backhaul; scheduling; channel quality indicator (CQI); CQI quantization.**

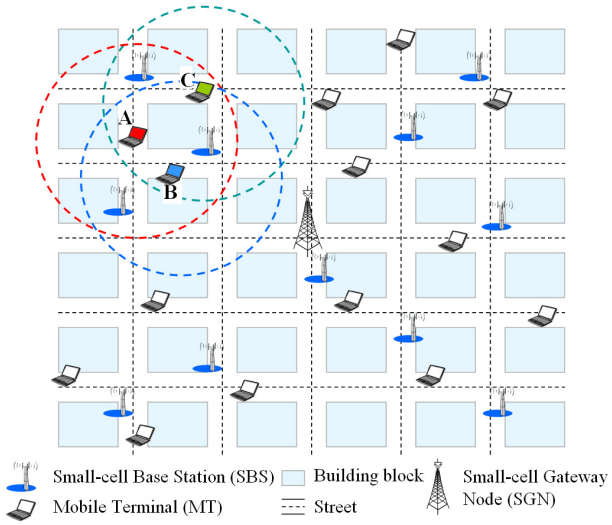

Fig. 1. An example of small cells with wireless backhaul, where there is a small-cell gateway node (SGN), and a number of distributed SBSs and mobile terminals (MT). The dash-line circles define a *user centric circle*, where the SBSs that potentially service a MT are enclosed. The red circle corresponds to MT A, the blue circle corresponds to MT B, and the green circle corresponds to MT C. Signals from other SBSs are too weak, and can be neglected.

I. Introduction

SMALL cells are the latest evolution of cellular systems, aiming to provide high throughput for hotspots [1, 2]. Recently, significant progress has been made in developing plug-and-play small-cell base stations (SBS), such as Alcatel-Lucent lightRadio^{TM} cube, Huawei Atom, and Nokia Siemens Networks Liquid Radio. Scheduling is a challenging task in this complex distributed network of small cells.

Several unique characteristics of small cells make scheduling challenging. One of them is demand oriented deployment, where SBSs are deployed in traffic hotspots. There are also a number of practical limitations on setting up SBSs, such as power availability and site accessibility; as well as the fact that some SBSs may be deployed by private companies or third party small cell service providers. As a result, SBSs are deployed in an ad-hoc manner deviating from traditional cellular structures [3].

Such ad-hoc deployment creates abundant overlapping coverage, as illustrated in Fig. 1, which imposes a challenge of strong *co-channel interference* to scheduling. Different carrier frequencies could be exploited between small cells in overlapping coverage areas to eliminate the interference and allow each SBS to carry out independent scheduling [4]. Unfortunately, this leads to a very poor throughput performance due to inefficient use of spectrum (as will be shown in this paper). It also causes frequent handover, and user experience of the mobile terminals (MT) would be interrupted.

Alternatively, coordinating SBSs to exploit multicell multiuser multiple-input multiple-output (MIMO) techniques [5–9] could help to suppress co-channel interference. However, the SBSs may be operated by a number of third parties, as mentioned earlier. This would prevent the SBSs from being coordinated. In addition, advanced multicell multiuser MIMO techniques, i.e., coordinated beamforming [5-9], would involve high cost of communication overhead, synchronization and backhaul for practical implementations. More specifically, those techniques require the SBSs to have accurate channel state information (CSI) [5, 9]; while for MTs, continuously varying CSI is too large to feed back [10]. This problem becomes even severer in practical cellular applications, where there would be many active MTs and each sends its CSI to compete for a downlink transmission opportunity. Coordinated beamforming is also very sensitive to synchronization errors [7]. Particularly, signals from different SBSs must be

Manuscript received December 14, 2012; revised April 15, 2013; accepted September 16, 2013. The associate editor coordinating the review of this paper and approving it for publication was M. L. Merani.

The authors are with CSIRO Computational Informatics (CCI), Marsfield, NSW, Australia, 2122 (e-mail: {wei.ni, iain.collings, ren.liu}@csiro.au).

Digital Object Identifier 10.1109/TWC.2013.101713.121966

synchronized at a sampling level, when received at each MT. Unfortunately, the timing error between a SBS and a MT is up to 24 samplings in practice [11]. Misaligned signal samplings will essentially have the phases of beamforming coefficients rotate unpredictably, degrading coordinated beamforming. The synchronization problem is difficult to solve, as synchronizing the signals from multiple SBSs at a MT would naturally lead to asynchronism at the others for a geometrical reason. Coordinated beamforming also requires high backhaul capacity to ensure that the data destined for all the MTs is error-free and ready for precoding at each SBS. There is another practical consideration preventing using coordinated beamforming, that is, commercial SBSs are IP devices, as specified by 3GPP system architecture evolution (SAE) [12]. Synchronizing different copies of a packet can hardly be guaranteed at different SBSs, even on a frame basis.

Several multicell coordinated MIMO techniques, named as coordinated multipoint transmission and reception (CoMP) in 3GPP long-term evolution (LTE) or LTE-advanced [13], have been implemented and tested in the field [14–18]. In the implemented systems, the stringent synchronization requirement between the SBSs and MTs was either bypassed by connecting all the SBSs and MTs with cables [15, 17, 18], or avoided by steering nulls to mitigate interference (in which case, only the SBSs need to be synchronized but extra feedback is required to indicate the locations of the MTs, as described in [14, 16]). Additionally, we note that only a few MTs were involved in the trials (e.g., two MTs in [14–16]), as the CSI overhead required is as high as 4.6 Mbits/s per MT for a 20 MHz system even in low mobility cases [15]. In typical urban mobile settings, there are many users and each sends its CSI to compete for an opportunity of being scheduled. Such a high CSI requirement would dramatically consume the available transmission bandwidths.

Another consequence of the ad-hoc deployment of small cells is that SBSs are often installed at places where optical fibre core network is not accessible, such as street lights and bus stops [2]. In such cases, small cells with wireless backhaul are feasible and promising. Such deployment is much faster and cheaper than installing new backhaul fibres. It also has the advantage of flexibility. The plug-and-play SBSs can be instantly moved to wherever a traffic growth is experienced. However, the lossy nature of wireless backhaul imposes further challenges to scheduling in small cells. An *idealized, centralized, multiuser MIMO scheduling scheme* was developed to overcome the lossy wireless backhaul [19, 20]. It is based on the channel quality indicators (CQI) of the MTs, as the CQI has a substantially lower feedback requirement than the CSI [10]. Unfortunately, the scheme requires prohibitively high overhead due to centralized operations, and therefore is of limited practical value.

In this paper, we propose a new decentralized downlink scheduling scheme that addresses both of the challenges of co-channel interference and lossy wireless backhaul in small cells. The key idea of our proposal is the user-centric approach to scheduling. In particular, a MT dynamically decides a single SBS (to be specific, one of the transmit (Tx) antennas of the SBS) for its service among all the candidate SBSs in its range based on a set of predefined criteria. This is in contrast with traditional cellular systems where the base stations make the scheduling decisions. The SBSs collect the pre-selection and channel quality feedback of the MTs, pick and send (in the same frequency) independent packets in a distributed manner. Every MT receives from one of the antennas of its associated SBSs at the same time. The antennas of the SBSs send different packets to the MTs, hence exploiting the spatial degree of freedom. Linear MIMO receivers are implemented at the MTs, detecting the packets destined for the MTs and cancelling the interference from other SBSs, as well as the interference from the other antennas of the same SBSs. This resolves the challenge of co-channel interference in small cells. The order of packet transmissions at each SBS relies on the SBS buffer status for the serviced MTs. This tackles the challenge of lossy wireless backhaul. Our proposed scheme is of practical value, as there is no critical requirement of stringent synchronization, and the channel quality feedback can be implemented with low rate feedback of CQI. It can be integrated in emerging small cell standards, such as small-cell components of 3GPP LTE [13]. Such integration may lead to further advance in future standards. The MTs' participation in scheduling can further cut the feedback overhead by 80% with a marginal sum-rate degradation.

We also optimize the CQI quantization levels for the proposed scheme. Based on the analytical characterization of the scheme, the optimization is achieved by using a new recursive method. Specifically, we recursively characterize all the quantization levels based on one of them. Only optimizing the chosen level, we can obtain all the optimal levels. The optimal levels substantially improve the sum rate, especially when quantization bits are few.

Our analysis, confirmed by simulations, shows that the optimized 1-bit CQI allows the proposed scheme to achieve 93.1% of the upper bound of the sum rate for small cells. Using 3-bit CQI, our scheme can achieve 97.6% of the upper bound. We also evaluate the throughput of our scheme, as it is an important performance indicator in a practical sense. In this case, the effect of allocating resources to sending the overhead is taken int account. Given the 80% reduced overhead, our scheme is able to achieve 139.5% higher throughput, compared to the idealized centralized scheme that provides the sum-rate upper bound for CQI based systems. As a result, bandwidth-demanding services and applications, such as mobile video, can be better supported.

The rest of this paper is organized as follows. In Section II, the protocol and the signalling requirement of the proposed scheme are elaborated on. In Section III, the optimal CQI quantization levels of the proposed scheme are derived, and the new recursive algorithm that can computationally efficiently achieves the optimality is presented. In Section IV, simulation and numerical results are provided, followed by concluding remarks in Section V.

II. Proposed Decentralized Scheduling Scheme

In this section, we provide an overview of the proposed scheme, followed by detailed protocol implementation.

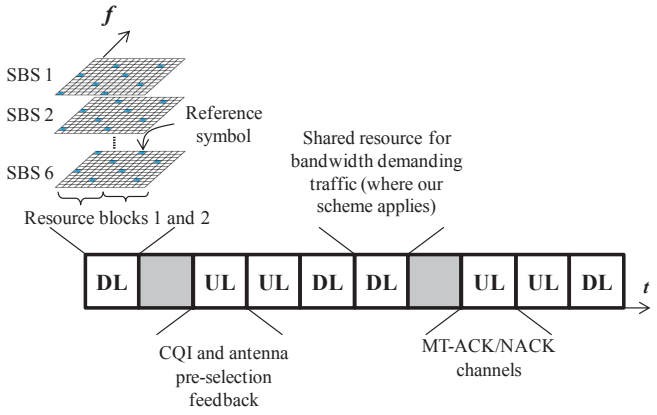

Fig. 2. An illustration of TDD-OFDM frame structure for implementing the proposed scheme, where downlink (DL) and uplink (UL) are in different subframes. The grey blocks are special subframes for DL-UL switching. Three new frame arrangements are proposed: (a) mobile-specific feedback channels, where each MT has a pre-allocated resource to feed back its SINR and antenna pre-selection; (b) a shared resource, where the SBSs use multicell multiuser MIMO to send packets as proposed; (c) packet-specific ACK/NACK channels, where the selected MTs feed back ACK/NACK in regards to every packet.

A. Scheme Overview

The ad-hoc small cell deployment scenario is shown in Fig. 1, where there is a small-cell gateway node (SGN) and a number of distributed SBSs and MTs. The SGN routes and forwards user data from the core network to the MTs, and also acts as the mobility anchor during handover (e.g., serving gateway specified by 3GPP LTE). The links between the SGN, SBSs, and MTs are wireless. For the purpose of illustrating the new user-centric approach to scheduling, we present it within the framework of 3GPP LTE, and consider time-division duplex (TDD) based orthogonal frequency-division multiple-access (OFDM) [13].

Fig. 2 illustrates an frame structure example of implementing our proposed scheme, where there are M SBSs that have an overlapping coverage area. In the downlink, we assign a shared radio resource to transmit bandwidth demanding traffic packets of all the M small cells. In the uplink, we reserve acknowledgement/non-acknowledgement (ACK/NACK) channels for the MTs pertaining to the packets that were transmitted in the shared downlink resource.

As a MT is likely to be in the overlapping coverage of the M SBSs, as discussed in Section I, we propose that the MT associates with all the SBSs in its range, and sends the registration information to the registration database in the SGN. When the MT moves, it disassociates the SBSs that fade away, and associates with new SBSs that come into its range. When the SGN sends a packet for a MT, it transmits the packet to all the SBSs associated with the MT. These SBSs will all detect the data, while only one of them will decide to send the data to the MT. The reason for the SGN transmitting to all the SBSs is that the SGN can hardly foresee which of the SBSs will best serve the MT in mobile environments. The transmission can be implemented with broadcast or multicast techniques [21], which can avoid the significant expenses of the SGN repeatedly sending the same data traffic to individual SBSs by exploiting the broadcast nature of wireless channels.

We also propose that the MT preselects an antenna of the SBSs that provide the overlapping coverage area. Based on the preselection and CQI feedback, each SBS independently selects one MT for each of its Tx antenna which will send packets to the MT. This particular way of associating each antenna with a MT is chosen because of its low CQI overhead and relatively loose synchronization requirement. In contrast, other multicell coordinated MIMO techniques, such as coordinated beamforming, would involve high cost of CSI overhead, synchronization and backhaul for practical implementations in urban mobile settings, as discussed in Section I.

Our scheme allows each SBS to transmit independently and eliminates the need of cell coordination, significantly reducing the feedback data rate. Our scheme is also able to achieve close to the maximum sum rate of CQI based mobile small cell systems, as will be shown in Section IV-B, and therefore is important for bandwidth demanding services. Besides, the dynamic association of our scheme can provide uninterrupted user experience for the MTs.

Note that accurate estimation of the channels from different SBSs are important for the MTs to make correct pre-selections. In LTE, to allow channel estimation, four downlink reference symbols are inserted in every resource block (which consists of 12 subcarriers and 7 OFDM symbols). The reference symbols of the SBSs that have coverage overlapping areas can relatively shift in frequency up to six subcarriers, thereby avoiding interference [22, Sec. 6.10.1], [23, Sec. 10.2.1.1], as demonstrated in the left-top corner of Fig. 2. In the case where there are more than six nearby SBSs, the SBSs that send their reference symbols in the same frequency-time elements will generate their reference symbols as two-dimensional pseudo-random sequences which retain significantly low cross-correlation in the time and frequency directions of the OFDM [22], [24, Sec. 16.2.2.1]. As a result, the MTs can mitigate the interference with correlation detection before estimating the channels. The use of these techniques allows for accurate channel measurements from multiple SBSs.

Also note that the CQI and pre-selection feedback of the MTs is assumed to be error free in this scheme. This assumption is reasonable, as many practical systems use reliable coding and modulation schemes to transmit fast feedback data, such as CQI and ACK/NACK [22, Sec. 5.4].

B. Detailed Implementation

The detailed design of our new scheme consists of two components. One is a new decentralized protocol. The other is a new buffer management method at the SBSs that underlies the decentralized protocol.

Fig. 3 provides a flow diagram of the new protocol, where there are three MTs in the overlapping coverage area of two SBSs. A single cycle of the protocol is demonstrated, which starts by the SGN's deciding the number of packets to send to the SBSs. The decision criteria will be described later. The MTs also measure the signal-to-interference-plus-noise ratios (SINR) from each SBS antenna based on the measurement of pilot signals that was demonstrated in Fig. 2.

As a key part of the proposed protocol, a new decentralized

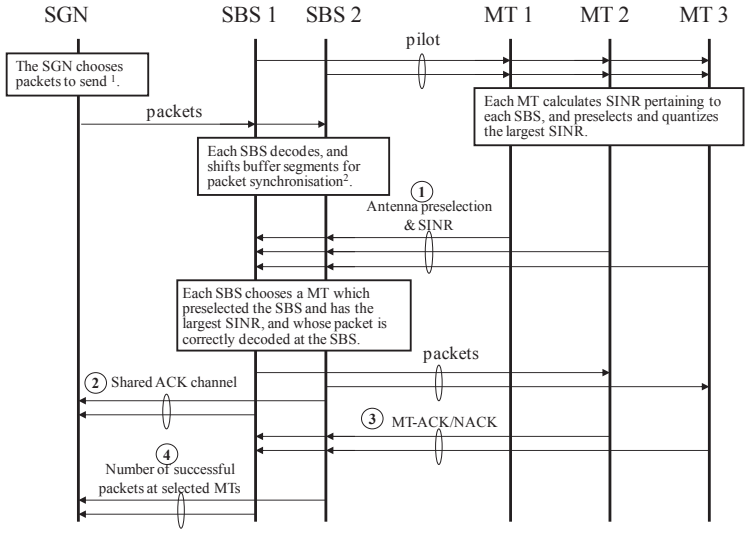

Fig. 3. A flow diagram of the new decentralized protocol, where two SBSs and three MTs are considered. We also provide the number of signaling bits pertaining to each signaling message: ① $(B + \lceil \log_2 N_T \rceil)K$; ② $N_T S_T$; ③ $N_T S$; and ④ $N_T(\lceil \log_2 K \rceil + \lceil \log_2 S_T \rceil)$, where B is the number of quantization bits per SINR, N_T is the total number of Tx antennas across all the SBSs that have an overlapping coverage area, K is the number of MTs in the overlapping coverage area, S_T is the maximum number of packets that each Tx antenna is able to send at a transmission opportunity, S is the maximum number of packets that a SBS buffers with regard to each MT, and $\lceil \cdot \rceil$ is the ceil function.

scheduling algorithm is then carried out at the MTs and SBSs. Following are the details.

Step 1 Each MT k preselects a SBS antenna, denoted by A_k, which contributes the largest SINR at the MT, denoted by γ_k.

$$\gamma_k = \max_{j=1,\cdots,N_T} \{\gamma_{k,j}\}$$
$$A_k = \arg\max_{j=1,\cdots,N_T} \{\gamma_{k,j}\}$$

where N_T is the total number of antennas of all the SBSs. $\gamma_{k,j}$ is the received SINR at MT k from SBS antenna j. The MT feeds back A_k and the quantized version of γ_k, denoted by $\hat{\gamma}_k$.

Step 2 For each SBS antenna $j \in \{1, \cdots, N_T\}$, its SBS selects a MT, denoted by π_j, from the set of MTs that preselected that antenna in Step 1, using the following criterion

$$\pi_j = \arg\max_{k \in \Phi_j \cap \Theta_j} \{\hat{\gamma}_k\}$$
$$\Phi_j = \{k : A_k = j, \forall k\}$$
$$\Theta_j = \{k : \theta_{k,j} \geq S_T, \forall k\}$$

where $\theta_{k,j}$ is the number of packets for MT k that have been successfully received at SBS antenna j from the SGN. Φ_j is the set of MTs that preselected antenna j in Step 1. Θ_j is the set of MTs for which SBS antennas have correctly received no less than S_T packets from the SGN.

Step 3 The SBS decides the number of packets for antenna j to transmit based on $\hat{\gamma}_{\pi_j}$, and transmits them independently in the shared resource.

Note that Step 2 specifies that the SBS chooses from the set of MTs for which the SBS has more than S_T packets to send. The threshold S_T is chosen to be larger than the maximum number of packets that a SBS antenna is able to send in a transmission opportunity, such that the transmission opportunities can be efficiently utilized. S_T can be predetermined by evaluating the radio resource size of a transmission opportunity, using the fastest modulation-coding scheme.

At the end of the protocol cycle, the selected MTs feed back ACK/NACK, denoted by "MT-ACK/NACK", in response to the SBSs' packet transmission. The MT-ACK/NACK messages are fed back in S pre-allocated MT-ACK/NACK channels. The SBSs interpret the MT-ACK/NACK messages to the numbers of successful packets, and forward them to the SGN. We also propose a new mapping mechanism relating to pre-allocating the S MT-ACK/NACK channels, which allows the SBSs to correctly discard successful packets in a decentralized manner. Details are provided later in this section.

Now, we can describe how the SGN decides the number of packets to send for the next protocol cycle, which is based on the ACK/NACK feedback of the SBSs and MTs. Consider a stable situation where the average system throughput is stationary. In response to the interpreted MT-ACK/NACK feedback that indicates the number of successful packets at a MT, the SGN will transmit an equal number of new packets destined for that MT. In response to the case where the SBSs all fail to decode a packet, the SGN will retransmit that packet. As such, packet overflows at the SBSs are avoided.

We also design a shared ACK channel with regard to each packet from the SGN, in which any SBS that successfully decoded the packet sends an ACK message. If the SGN receives signals in the shared ACK channel, it means the corresponding packet has been correctly decoded at some SBSs, and no retransmission is needed.

In terms of buffer management, we propose to configure K first-in-first-out (FIFO) buffers at each of the SBSs that have an overlapping coverage area, storing packets (from the SGN) destined for the K MTs in the area. Each buffer can

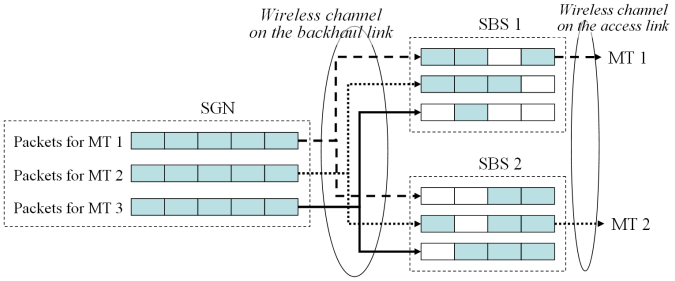

Fig. 4. An example of logical (buffer) segments at the SBSs, where two SBSs and three MTs are considered, and buffer i corresponds to MT i ($i = 1, 2, 3$). Each buffer consists of four (logical) segments, denoted by a block in the figure, each of which corresponds to a packet. A blue logical segment corresponds to a successfully decoded at the SBS, and a white logical segment corresponds to an unsuccessfully decoded packet at the SBS.

be mapped to S logical segments, indicating S consecutive packets that are yet to be delivered to the MT. If the SBS has successfully decoded a packet, the corresponding logical buffer segment is occupied; otherwise, it is empty (see Fig. 4). Since the SBSs receive the same packets from the SGN, the logical buffer segments (i.e., the packet sequence) can therefore be synchronized across the SBSs without centralized coordination, even in the case where the SBSs successfully decode different subsets of the packets.

We also propose to pre-allocate the aforementioned S MT-ACK/NACK channels one-to-one corresponding to the logical segments. The SBSs update the logical segments such that, once a MT-ACK message is received, the SBSs shift out the corresponding logical segments (occupied or not) independently. The actual buffer segments which correspond to the logical segments are released to store new incoming packets.

Such buffer management allows the SBSs to discard the successfully delivered packets in a decentralized manner. The requirement of central coordination and associated signaling is eliminated. In practice, the packet buffers can be implemented as linked circular buffers [25]. The shift operations are carried out by updating the logical segment pointers while the buffer segments stay unchanged. Additionally, buffer segments are only allocated to packets that are successfully decoded by SBSs. For the packets that are not decoded, the corresponding logical segment pointers will point to 'null', signifying empty segments without occupying buffer space.

Consider the example of Fig. 4, where SBSs 1 and 2 choose to transmit to MTs 1 and 2, respectively. If SBS 2 successfully delivered the rightmost two packets to MT 2 and the two SBSs have received two MT-ACK messages, then both the SBSs knock off their rightmost two logical segments and fill the gaps by shifting the remaining logical segments rightwards. As a result, SBS 1 discards successful packets of MT 2, though not involved in transmission to MT 2. The sequence of undelivered packets is correctly maintained at every SBS.

We can enhance fairness of our protocol by exploiting the threshold S_T. Specifically, the SGN can be designed to slow down transmitting packets destined for the MTs that have been selected more often than the others. As a consequence, the SBS buffer occupancy for those MTs becomes below the threshold, which gives other MTs opportunities to be scheduled. Our protocol can be further enhanced by setting a lifetime for every packet in the buffers. The lifetime increases as the packet progresses, until the packet is successfully delivered to the MT. If the lifetime becomes larger than a predefined value (e.g., 50 ms), the SBSs treat the buffer pertaining to the packet as a threshold exceeded buffer, and allow it to be scheduled. As a result, deadlock can be avoided in case where the SGN sends S packets successfully, whereas none of the SBSs succeed in decoding more than S_T packets. These enhancements are useful, but not discussed in this paper due to page limitation.

III. Optimization of CQI Quantization Levels

In this section, we start with the scenario where the CQI feedback of the MTs is imperfect (i.e., quantized), and optimize the CQI quantization levels, which maximize the sum rate of the proposed scheduling scheme. At the end of the section, we also discuss the scenario where the CQI is perfect (i.e., not quantized), which provides the upper bound to our optimized quantization levels.

In the case where the CQI is imperfect, to optimize the CQI quantization levels we first derive the probability mass function (PMF) of the number of MTs that can be chosen for each SBS Tx antenna in Step 2 of Section II-B (i.e., the cardinality of $\Phi_j \cap \Theta_j$). Let L_j denote this number in regards to SBS antenna $j \in \{1, \cdots, N_T\}$, and N_R denote the number of receive antennas at each MT. Let $N_T \leq N_R$, such that every MT is able to suppress interference by its own using MIMO detection methods. The PMF can be derived as follows.

For illustration convenience, on the backhaul link we consider independently and identically distributed (i.i.d.) fading. It is worth pointing out that the optimization algorithm described in this section is applicable to the general cases where the links are non-identical due to different link lengths. The probability of a packet being successfully decoded at a SBS, denoted by p_0, is identical for all SBSs. Then, the probability of more than S_T such packets at any SBS is given by

$$p_1 = \sum_{m=S_T}^{S} \binom{S}{m} p_0^m (1-p_0)^{S-m}.$$

Let p_e be the coded bit error rate on the backhaul link and W be the packet size (i.e., the size of a buffer segment in bits). When p_e is very small, $p_0 \approx 1 - p_e W$.

On the small-cell access link, we consider i.i.d. Rayleigh fading, as it models heavily built-up urban environments. Then, each Tx antenna is equally likely to be preselected by a MT (but may not be able to send packets to the MT, due to unsuccessful decoding on its backhaul). Let p_2 denote the likelihood. $p_2 = \frac{1}{N_T}$.

The joint PMF of $L_1, \cdots, L_{N_T}$ can be derived, given in (1), where $\mathcal{L}_i$ is the number of MTs that preselected the i-th SBS antenna, $\sum_{i=1}^{N_T} \mathcal{L}_i = K$, and l_i is the number of MTs out of $\mathcal{L}_i$ that have threshold exceeded buffers at antenna i for $0 \leq l_i \leq \mathcal{L}_i$. $\binom{K-\sum_{j=1}^{i-1}\mathcal{L}_j}{\mathcal{L}_i} p_2^{\mathcal{L}_i}$ is the probability of $\mathcal{L}_i$ MTs preselecting Tx antenna i, and $\binom{\mathcal{L}_i}{l_i} p_1^{l_i} (1-p_1)^{\mathcal{L}_i - l_i}$ is the probability of l_i of the $\mathcal{L}_i$ MTs having threshold exceeded buffers at the SBS of antenna i while the remaining $(\mathcal{L}_i - l_i)$ MTs do not. The product $\prod_{i=1}^{N_T}$ accounts for all N_T Tx antennas. The multiple

$$\Pr(L_1 = l_1, \cdots, L_{N_T} = l_{N_T}) = \sum_{\substack{\mathcal{L}_i : \sum_{i=1}^{N_T} \mathcal{L}_i = K \\ l_i : l_i \le \mathcal{L}_i,\, i=1,\cdots,N_T}} \cdots \sum \prod_{i=1}^{N_T} \left[\binom{K - \sum_{j=1}^{i-1} \mathcal{L}_j}{\mathcal{L}_i} p_2^{\mathcal{L}_i} \binom{\mathcal{L}_i}{l_i} p_1^{l_i} (1-p_1)^{\mathcal{L}_i - l_i} \right]$$
$$= \sum_{\substack{\mathcal{L}_i : \sum_{i=1}^{N_T} \mathcal{L}_i = K \\ l_i : l_i \le \mathcal{L}_i,\, i=1,\cdots,N_T}} \cdots \sum \left\{ p_1^{\sum_{i=1}^{N_T} l_i} (1-p_1)^{K - \sum_{i=1}^{N_T} l_i} p_2^K \prod_{i=1}^{N_T} \left[\binom{K - \sum_{j=1}^{i-1} \mathcal{L}_j}{\mathcal{L}_i} \binom{\mathcal{L}_i}{l_i} \right] \right\}, \quad (1)$$

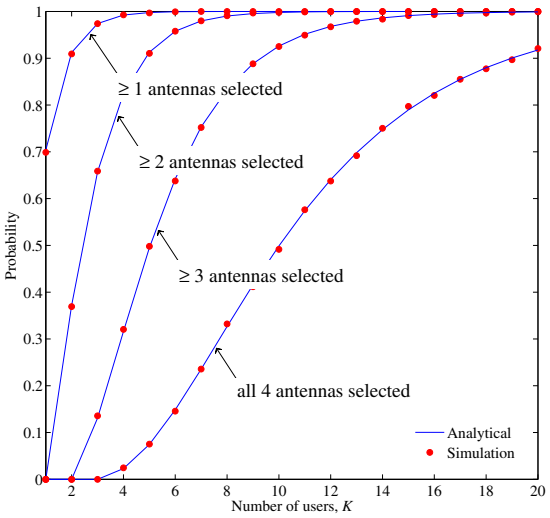

Fig. 5. Distribution of the number of scheduled antennas, where $M = N_T = N_R = 4$ (i.e., 4 single-antenna SBSs), $S = 20$, and $S_T = 12$. We consider a $\frac{3}{4}$-rate convolutional code with the polynomial (171, 133), 16 quadrature amplitude modulation (QAM), and an average SNR of 20 dB on the backhaul link. Each SBS uses a single antenna to receive. Therefore, $p_e = 1.8^{-4}$ [26, Fig. 10]. We also assume $W = 2048$.

sums collect all possible combinations of $l_1, \cdots, l_{N_T}$. Note that (1) still holds if the numbers of Tx antennas are different between the SBSs, since our scheduling is per-antenna based and independent between the SBS Tx antennas.

Fig. 5 plots the probability of the numbers of scheduled SBS antennas, showing analytical results from (1) and simulation results; and it confirms the accuracy of (1). Specifically, every analytical curve indicating more than 1 (2, 3, or 4) antenna selected is plotted by collectively adding up the PMFs of (1) with more than 1 (2, 3, or 4) non-zero l_i. The figure shows that the proposed scheme is able to make effective use of the SBS Tx antennas. When $K \ge 15$, it is almost certain that no less than three antennas are scheduled. When $K \ge 20$, the probability of having unscheduled antennas is less than 8%.

Next, we derive the sum rate of our scheme. Although current LTE-A and WiMax standards do not require user equipments to have MIMO capability, a range of MIMO receivers have been investigated in the industry, such as zero-forcing detector (ZFD), maximum mean square error (MMSE), maximum likelihood detector (MLD), QR decomposed MLD, and successive interference cancellation receivers [23, Sec. 5.5]. Amongst these techniques, ZFD and MMSE are linear MIMO receivers with complexity linear to the number of antennas [23, Sec. 5.5.2]. They are practically affordable to capability-limited user equipments (e.g. handset). Other techniques, though achieving lower bit error rates, require significantly high complexity which prevents them being practically implemented to the user equipments at the present [23, Sec. 5.5.3]. For this reason, we consider the ZFD in this paper. It is worth noting that our proposed scheduling approach and the technique of optimizing the CQI levels are also applicable to other MIMO receivers, such as MMSE.

With the ZFD, the received signal at MT k is given by

$$\mathbf{y}_k = \mathbf{h}_{k,n}\, x_n + \sum_{j=1,\, j \ne n}^{N_T} \mathbf{h}_{k,j}\, x_j + \mathbf{n}_k\,,$$

where $\mathbf{h}_{k,j}$ is the $N_R \times 1$ channel vector from antenna j to MT k, $\mathbf{n}_k$ is an $N_R \times 1$ additive white complex Gaussian noise vector, and x_j is the signal of antenna j. $E(|x_j|^2) = \frac{P}{N_T}$. P is the total Tx power of all the antennas.

The received SINR of MT k regarding antenna j is [27]

$$\gamma_{k,j} = \frac{\rho}{N_T [\mathbf{H}_k^\dagger \mathbf{H}_k]^{-1}_{j,\,j}}\,, \quad (2)$$

where $\mathbf{H}_k = (\mathbf{h}_{k,1} \cdots \mathbf{h}_{k,\,N_T})$ is the channel matrix of MT k, $\rho = \frac{P}{N_0}$ is the Tx signal-to-noise ratio (SNR), and † denotes the conjugate transpose. N_0 is the noise power.

Consider $N_T = N_R$, as this is the case where, for a given number of Rx antennas per MT, the number of MTs that can be selected at the same time is maximized. The subscripts "$_T$" and "$_R$" are suppressed for notational convenience. The probability distribution function (PDF) and the cumulative distribution function (CDF) of $\gamma_{k,\,j}$ can be given by [28]

$$f_{\gamma_{k,j}}(x) = \tfrac{N}{\rho}\, e^{-\frac{Nx}{\rho}};\; F_{\gamma_{k,j}}(x) = 1 - e^{-\frac{Nx}{\rho}} \quad (N \ge 2). \quad (3)$$

By definition, $\gamma_k = \max\{\gamma_{k,j}\}$. Exploiting Order Statistics [29] on (3), we have the PDF and CDF of γ_k as

$$f_{\gamma_k}(x) = \frac{N^2}{\rho} e^{-\frac{Nx}{\rho}} (1 - e^{-\frac{Nx}{\rho}})^{N-1}$$
$$F_{\gamma_k}(x) = (1 - e^{-\frac{Nx}{\rho}})^N. \quad (4)$$

Clearly, $\gamma_1, \cdots, \gamma_K$ are i.i.d. under the i.i.d. assumption on the small cell link, as mentioned earlier. We define a random variable γ which has an identical distribution as γ_k. Then, the PDF and CDF of γ, $f_\gamma(x)$ and $F_\gamma(x)$, are also given by (4).

Consider the B-bit quantization in Step 1 of Section II-B at each MT for that terminal's largest SINR, γ_k. The PMF of γ_k being quantized to the i-th CQI quantization level Q_i

$(i = 1, \cdots, 2^B)$ is

$$\begin{aligned}\Pr(\hat{\gamma}_k = Q_i) &= \int_{Q_i}^{Q_{i+1}} f_\gamma(x)dx \\ &= (1 - e^{-\frac{NQ_{i+1}}{\rho}})^N - (1 - e^{-\frac{NQ_i}{\rho}})^N,\end{aligned} \tag{5}$$

where $\hat{\gamma}_k$ is the quantized version of γ_k, i.e., if $Q_i \le \gamma_k < Q_{i+1}$ $(i = 1, \cdots, 2^B)$, then $\hat{\gamma}_k = Q_i$. $[Q_i, Q_{i+1})$ is the i-th CQI quantization interval. $Q_{2^B+1} = \infty$. We also have

$$\Pr(\hat{\gamma}_k < Q_i) = \int_0^{Q_i} f_\gamma(x)dx = (1 - e^{-\frac{NQ_i}{\rho}})^N. \tag{6}$$

In Step 2 of Section II-B, the SBSs choose the MTs for the SBS antennas. For SBS antenna n, there are L_n MTs to choose from. In the i.i.d. channel, the fed-back SINRs of the L_n MTs are also i.i.d.. The PDF of the maximum of the L_n SINRs, denoted by $\hat{\gamma}_{\pi_n}$, can be derived as

$$\begin{aligned}&\Pr(\hat{\gamma}_{\pi_n} = Q_i | L_n) \\ &\overset{(a)}{=} \Big[\Pr(\hat{\gamma}_k = Q_i) + \Pr(\hat{\gamma}_k < Q_i)\Big]^{L_n} - \Pr(\hat{\gamma}_k < Q_i)^{L_n} \\ &\overset{(b)}{=} \left(1 - e^{-\frac{NQ_{i+1}}{\rho}}\right)^{NL_n} - \left(1 - e^{-\frac{NQ_i}{\rho}}\right)^{NL_n},\end{aligned}$$

where π_n denotes the MT selected for the n-th SBS antenna in Step 2, as defined in Step 2 of Section II-B, and $\hat{\gamma}_{\pi_n}$ is the quantized SINR of that MT in regards to the n-th SBS antenna. (a) is obtained by exploiting Order Statistics of discrete random variables [29]. (b) is obtained by substituting (5) and (6) into (a). The PMF of L_n has been derived in (1).

Using binomial expansion, we rewrite $\Pr(\hat{\gamma}_{\pi_n} = Q_i | L_n)$ as

$$\begin{aligned}&\Pr(\hat{\gamma}_{\pi_n} = Q_i | L_n) = \\ &\sum_{k=0}^{NL_n} \left[\binom{NL_n}{k}(-1)^k \left(e^{-\frac{NQ_{i+1}}{\rho}k} - e^{-\frac{NQ_i}{\rho}k}\right)\right].\end{aligned}$$

In Step 3 of Section II-B, the SBSs send packets to the selected MTs. In practice, the SBSs carry out modulation and coding based on Q_i $(i \le 2^B)$. The closed-form expression for the sum rate can be derived, given in (7), where $P_{L_1, \cdots, L_N}$ is given by (1), and $\sum_{L_n: \sum_{n=1}^N L_n \le K} \cdots \sum$ adds up all the combinations of $L_1, \cdots, L_N$ satisfying $\sum_{n=1}^N L_n \le K$.

We proceed to optimize the quantization levels of the MTs in Step 2 of Section II-B, to maximize the sum rate of the new scheme (7). To do this, a constrained optimization problem is formulated with respect to Q_i $(i = 2, \cdots, 2^B)$, as given by

$$\begin{aligned}&\max R_{\mathcal{F}}(Q_2, \cdots, Q_{2^B}) \\ &s.t. \;\; 0 \le Q_i < Q_{i+1},\, i = 2, \cdots, 2^B\end{aligned} \tag{8}$$

where $R_{\mathcal{F}}(Q_2, \cdots, Q_{2^B})$ is provided by (7), since it is a function of $Q_2, \cdots, Q_{2^B}$.

Unfortunately, the inequality constraints and the non-convexity make it difficult for non-linear programming algorithms, such as Newton's methods [30], to solve (8). (The non-convexity will be discussed later in this section.)

We develop a new recursive algorithm to solve (8). First, rewrite (7) as

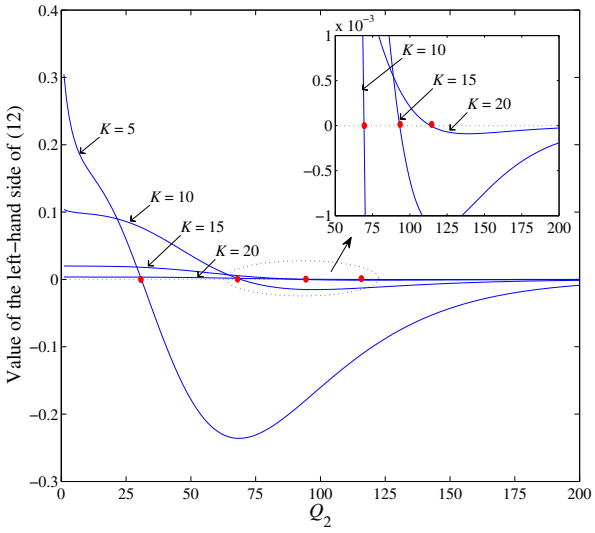

Fig. 6. The left-hand side of (12), where the optimal values of Q_2 are denoted by red dots. In the figure, on the backhaul link, we consider $p_e = 1.8^{-4}$, $W = 2048$, $S = 20$, and $S_T = 12$. On the access link, we consider $B = 1$, $N = 4$ and $\rho = 20$ dB. For visual clarity, we magnify part of the curves that is highlighted in the eclipse in the right-upper corner.

$$\begin{aligned}R_{\mathcal{F}} = \sum_{L_n: \sum_{n=1}^N L_n \le K} \cdots \sum \Bigg\{ P_{L_1, \cdots, L_N} \sum_{n=1}^N \Bigg\{ \Bigg[\left(1 - e^{-\frac{N}{\rho}Q_{i+1}}\right)^{NL_n} \\ - \left(1 - e^{-\frac{N}{\rho}Q_i}\right)^{NL_n} \Bigg] \log_2(1 + Q_i) + \left(1 - e^{-\frac{N}{\rho}Q_i}\right)^{NL_n} \\ \log_2(1 + Q_{i-1}) \Bigg\} \Bigg\} + G(Q_1, \cdots, Q_{i-1}, Q_{i+1}, \cdots, Q_{2^B}).\end{aligned}$$

where $G(\cdot)$ is a function independent of Q_i.

Apparently, the partial derivative of $R_{\mathcal{F}}$ with respect to Q_i is only dependent on its adjacent CQI quantization levels, Q_{i-1} and Q_{i+1} (in case of $i = 2^B$, only Q_{2^B-1}), and itself, Q_i. The partial derivative is given by (9).

For $i = 2^B$, we have $e^{-\frac{N}{\rho}Q_{i+1}} = 0$. Exploiting the first-order necessary condition of a local maximum (i.e., $\frac{\partial R_{\mathcal{F}}}{\partial Q_i} = 0$ [30]) on (9), the optimal Q_{2^B-1} is derived, given in (10). Likewise, for $1 \le i < 2^B - 1$, the optimal Q_i is given in (11). Q_i $(i = 2^B, 2^{B-1}, \cdots, 2)$ can be recursively calculated by using (10) and (11).

Our recursive algorithm can be summarized as follows.

- Substitute (10) in (11) recursively to express Q_1 with respect to Q_{2^B}.
- Solve Q_{2^B} since $Q_1 = 0$.
- Substitute Q_{2^B} in (10) and then (11) to calculate Q_{2^B-1} to Q_2 one after another.

The 2^B CQI quantization levels are all optimized.

Take an example of $B = 1$ to show a simple case of the

$$\begin{aligned} R_{\mathcal{F}} &= \sum_{i=1}^{2^B} \left[\log_2(1+Q_i) \sum_{L_n:\sum_{n=1}^N L_n \le K} \cdots \sum \left(P_{L_1,\cdots,L_N} \sum_{n=1}^{N} \Pr(\hat{\gamma}_{\pi_n} = Q_i | L_n) \right) \right] \\ &= \sum_{i=1}^{2^B} \sum_{L_n:\sum_{n=1}^N L_n \le K} \cdots \sum \sum_{n=1}^{N} \sum_{k=0}^{NL_n} \left[P_{L_1,\cdots,L_N} \binom{NL_n}{k} (-1)^k \left(e^{-\frac{NQ_{i+1}}{\rho}k} - e^{-\frac{NQ_i}{\rho}k} \right) \log_2(1+Q_i) \right]. \end{aligned} \tag{7}$$

$$\begin{aligned} \frac{\partial R_{\mathcal{F}}}{\partial Q_i} = \sum_{L_n:\sum_{n=1}^N L_n \le K} \cdots \sum \Bigg\{ P_{L_1,\cdots,L_N} \sum_{n=1}^{N} \Bigg[& -\frac{N^2 L_n}{\rho} e^{-\frac{N}{\rho}Q_i} \left(1 - e^{-\frac{N}{\rho}Q_i}\right)^{NL_n-1} \log_2(1+Q_i) + \\ & \frac{\left(1 - e^{-\frac{N}{\rho}Q_{i+1}}\right)^{NL_n} - \left(1 - e^{-\frac{N}{\rho}Q_i}\right)^{NL_n}}{(1+Q_i)\ln 2} + \frac{N^2 L_n}{\rho} e^{-\frac{N}{\rho}Q_i} \left(1 - e^{-\frac{N}{\rho}Q_i}\right)^{NL_n-1} \log_2(1+Q_{i-1}) \Bigg] \Bigg\}. \end{aligned} \tag{9}$$

$$Q_{2^B-1} = -1 + 2 \left\{ \frac{\sum_{L_n:\sum_{n=1}^N L_n \le K} \cdots \sum \left\{ P_{L_1,\cdots,L_N} \sum_{n=1}^N \left[\frac{1-\left(1-e^{-\frac{N}{\rho}Q_{2^B}}\right)^{NL_n}}{(1+Q_{2^B})\ln 2} - \frac{N^2 L_n \log_2(1+Q_{2^B})}{\rho e^{\frac{N}{\rho}Q_{2^B}}} \left(1-e^{-\frac{N}{\rho}Q_{2^B}}\right)^{NL_n-1} \right] \right\}}{\sum_{L_n:\sum_{n=1}^N L_n \le K} \cdots \sum \left\{ P_{L_1,\cdots,L_N} \sum_{n=1}^N \left[\frac{N^2 L_n}{\rho} e^{-\frac{N}{\rho}Q_{2^B}} \left(1-e^{-\frac{N}{\rho}Q_{2^B}}\right)^{NL_n-1} \right] \right\}} \right\}. \tag{10}$$

$$Q_i = -1 + 2 \left\{ \frac{\sum_{L_n:\sum_{n=1}^N L_n \le K} \cdots \sum \left\{ P_{L_1,\cdots,L_N} \sum_{n=1}^N \left[\frac{\left(1-e^{-\frac{N}{\rho}Q_{i+2}}\right)^{NL_n} - \left(1-e^{-\frac{N}{\rho}Q_{i+1}}\right)^{NL_n}}{(1+Q_{i+1})\ln 2} - \frac{N^2 L_n \log_2(1+Q_{i+1})}{\rho e^{\frac{N}{\rho}Q_{i+1}}} \left(1-e^{-\frac{N}{\rho}Q_{i+1}}\right)^{NL_n-1} \right] \right\}}{\sum_{L_n:\sum_{n=1}^N L_n \le K} \cdots \sum \left\{ P_{L_1,\cdots,L_N} \sum_{n=1}^N \left[\frac{N^2 L_n}{\rho} e^{-\frac{N}{\rho}Q_{i+1}} \left(1-e^{-\frac{N}{\rho}Q_{i+1}}\right)^{NL_n-1} \right] \right\}} \right\}. \tag{11}$$

recursive algorithm. From (10), we have

$$\begin{aligned} \sum_{L_n:\sum_{n=1}^N L_n \le K} \cdots \sum \Bigg\{ P_{L_1,\cdots,L_N} \sum_{n=1}^{N} \Bigg[& 1 - \left(1 - e^{-\frac{N}{\rho}Q_{2^B}}\right)^{NL_n} \\ & - \frac{N^2 L_n}{\rho} e^{-\frac{N}{\rho}Q_{2^B}} \left(1 - e^{-\frac{N}{\rho}Q_{2^B}}\right)^{NL_n-1} \times \\ & (1+Q_{2^B}) \ln(1+Q_{2^B}) \Bigg] \Bigg\} = 0, \end{aligned} \tag{12}$$

which can be solved by using a bisection method.

Fig. 6 plots the left-hand side of (12), and confirms that in the region of $[0, 2\rho]$ (ρ is 100 in decimal), (12) always has a root which is the optimal 1-bit CQI quantization level. The plotted curves are descendent when crossing the x axis. Note that the left-hand side of (12) is equal to (9) for $B = 1$. In other words, (9) is descendent. The second-order necessary condition of a local maximum [30], $\frac{\partial^2 R_{\mathcal{F}}}{\partial Q_{2^B}^2} < 0$, is satisfied at those roots. Therefore, the roots shown in the figure are locally optimal [30]. In fact, the roots are globally optimal, as they are the only roots (as shown in Fig. 6). The figure also reveals that (9) is not a monotonic function, which means $R_{\mathcal{F}}(Q_{2^B})$ is not convex or concave [30]. It would be difficult for other multivariate optimization methods [31] to solve (8). Our recursive algorithm is a simple and effective way to optimize the CQI levels.

For comparison purpose, we also derive the sum rate when the precise CQI is available at the SBSs. Exploiting Order Statistics [29] on (4), then differentiating it, and carrying out binomial expansion, we obtain the PDF of γ_{π_n}, as given by

$$f_{\gamma_{\pi_n}}(x) = -\frac{N^2 L_n}{\rho} \sum_{i=0}^{NL_n-1} \left[\binom{NL_n-1}{i} \left(-e^{-\frac{Nx}{\rho}}\right)^{i+1} \right]. \tag{13}$$

Given the precise CQI, the sum rate is given by

$$\begin{aligned} R_{\mathcal{I}} = -\frac{N}{\ln 2} \sum_{L_n:\sum_{n=1}^N L_n \le K} \cdots \sum \sum_{n=1}^{N} \sum_{i=0}^{NL_n-1} \Bigg[& \frac{P_{L_1,\cdots,L_N}}{i+1} \\ & L_n \binom{NL_n-1}{i} \left(-e^{\frac{N}{\rho}}\right)^{i+1} \Gamma\left(0, \tfrac{N}{\rho}(i+1)\right) \Bigg]. \end{aligned} \tag{14}$$

where $\Gamma(\cdot,\cdot)$ is the incomplete Gamma function and can be directly obtained in MATLAB.

IV. Numerical Results

In this section, we consider a practical small cell system where $M = N$, i.e., each SBS has a single Tx antenna. We assume the bit error rate $p_e = 1.8^{-4}$ on the backhaul link, which corresponds to a link with a SNR of 20 dB and using

$\frac{3}{4}$-rate 16QAM with conventional coding, as considered in Fig. 5. The packet size $W = 2048$. The number of segments per buffer $S = 20$. On the access link, we assume the buffer threshold $S_T = 12$.

We compare our proposed decentralized scheme with two other strategies that are applicable to small cells. The first strategy is a *ideally centralized multicell MIMO scheduling scheme* [19], which exploits linear MIMO techniques in the same way as the proposed decentralized scheme does. The difference is that this scheme requires the MTs to feed back the SINR from every SBS Tx antenna, based on which the SGN makes a globally optimal scheduling decision maximizing the sum rate. The scheme can provide the sum-rate upper-bounds for the small cells where multicell multiuser MIMO is exploited. Details are provided in Appendix A.

The second strategy is the *conventional frequency-reuse deployment scheme*, where the small cells that have an overlapping coverage area exploit different parts of the available radio resource to avoid interference [4]. Each SBS independently schedule its MTs based on their CQI feedback. Maximal ratio combining (MRC) is carried out at the MTs to detect downlink packets [32]. Details are provided in Appendix B.

We further compare the CQI quantization of our proposed decentralized scheme with two existing quantization designs. The two existing designs were originally developed for conventional cellular systems where the base stations have hard wired connections to the core networks (i.e., negligible decoding failure on the backhaul link) [33]. One design, referred to as the *pre-scheduling law*, was based on the CDF of the SINR before scheduling. Extended to our new scheme, the pre-scheduling CDF is given by (4). The other design, referred to as the *post-scheduling law*, was based on the CDF of the SINR after scheduling. Extended to our new scheme, the post-scheduling CDF can be obtained by first integrating (13) with respect to x to get the CDF for a given L_n $(n = 1, \cdots, N)$ and then calculating the mean for all possible L_n.

A. Throughput Evaluation

We first evaluate the throughput for its practical importance to the bandwidth-demanding traffic. The effect of allocating resources to sending the overhead is taken into account.

Here, we consider a super-frame structure where the access and backhaul links are multiplexed using a time-division multiplexing (TDM) technique. In other words, each super-frame consists of two consecutive frames. One frame is for the access link, details of which are described in Fig. 2. The other frame is for the backhaul link, which is also divided into an uplink sub-frame and a downlink sub-frame. We also assume that the duration of each super-frame is 20 ms.

Consider a single scheduling process (as demonstrated in Fig. 3). In every super frame, we allocate a total of R *subcarrier symbols* for the SBSs to transmit downlink packets and for all the nodes (i.e., the SGN, SBSs and MTs) to send the overhead. We allow the boundary between the two access and backhaul frames to be configurable (i.e., move leftwards or rightwards). In other words, the R subcarrier symbols can be divided unevenly between the two frames.

To ensure the reliability of signalling, we assume every overhead bit is spread to 4 bits using repetition coding, and transmitted using QPSK [34]. The SBSs exploit the remainder of the R subcarrier symbols, denoted by r subcarrier symbols, to transmit packets. We also assume that the wireless channels do not change between the r subcarrier symbols.

We also carry out Monte-Carlo simulations to validate the proposed scheme. At the beginning of each simulated super

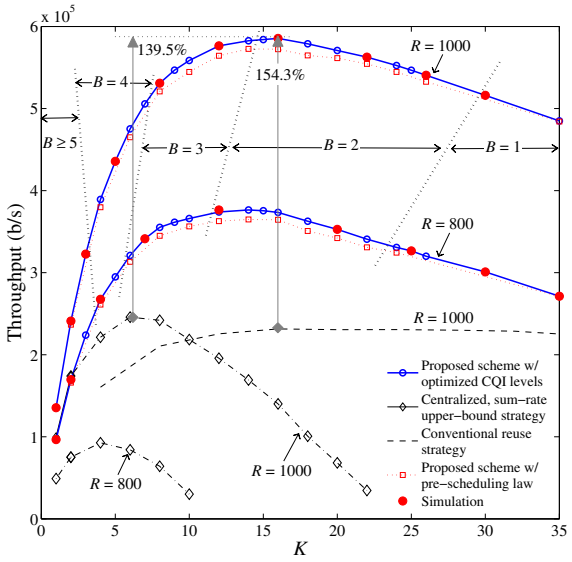

Fig. 7. Comparison on throughput, where $N = 4$, $\rho = 20$ dB, $S_T = 12$ and $S = 20$. For the proposed scheme, we analytically plot the throughput by multiplying the sum rate (7) and r. All the values of B are compared and the one contributing the largest throughput is plotted. For the scheme which can provide the sum-rate upper bounds, we plot its asymptotic upper bounds of the throughput, which are equal to the bigger one of the LUB (24) multiplied by the r corresponding to $B = 1$, and the GUB (20) multiplied the r corresponding to $B = 2$.

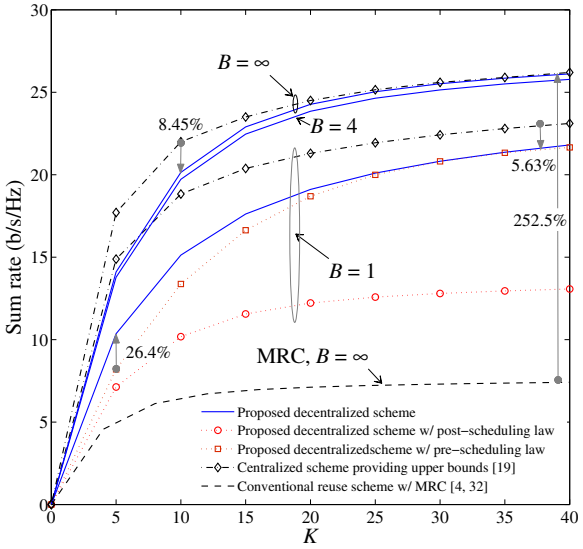

Fig. 8. Sum rate versus K, where $N = 4$, $\rho = 20$ dB, $S_T = 12$ and $S = 20$. The new scheme is analytically plotted by using (7) in the cases of $B = 1$ and 4, and (14) in the case of $B = \infty$. The upper bounds are analytically plotted by using (24) in the case of $B = 1$, and using (20) in the case of $B = \infty$.

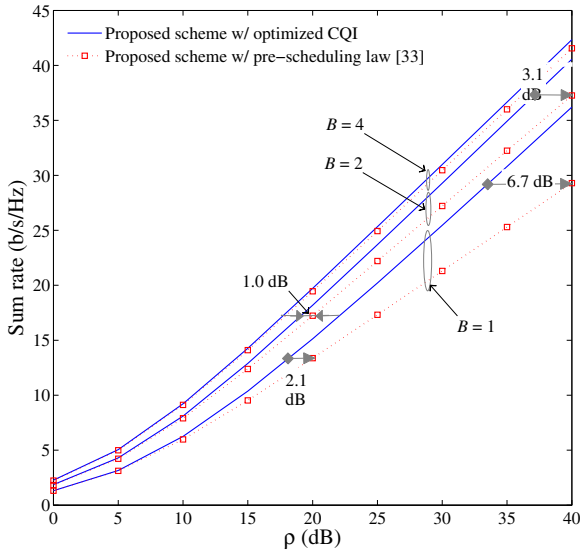

Fig. 9. Sum rate versus ρ, where $N = 4$, $K = 10$, $S_T = 12$ and $S = 20$.

frame, we generate binomially distributed numbers to indicate the new packets that the SBSs correctly decode. We then carry out the proposed protocol to schedule the packets. Since the numbers of packets the SBSs send on the small cell links are decided by the CQI levels, we assume all those packets are successful at the MTs. The number of successful packets will be used to generate the number of new packets in the next super frame. Rayleigh block fading channels are assumed on the small cell links; in other words, the channels do not change during a super frame, and change independently between super frames.

In Fig. 7, our analytical results, confirmed by simulations, show that, in terms of throughput, our proposed scheme is 139.5% higher than the centralized strategy which provides the sum-rate upper bounds. The reason is a substantially lower overhead of our scheme (as will be shown later), which not only spares more radio resources for packets transmission, but also allows for using more accurate CQI (e.g., $B \geq 3$) to improve the sum rate. In contrast, the sum-rate upper-bound centralized strategy can only use a smaller number of CQI quantization bits (i.e., $B \leq 2$), and degrades the sum rate.

Our proposed scheme is also 154.3% better than the conventional frequency-reuse strategy. This is because our scheme can achieve a significantly higher sum rate than the conventional frequency-reuse strategy, as will be discussed later.

Our optimized CQI quantization levels contribute to the significant throughput improvements. Specifically, they improve the peak throughput of the new scheme by 20 kb/s, compared to the existing pre-scheduling law, when $R = 1000$ subcarrier symbols. Note that the peak throughput indicates the balance point of the multiuser diversity gain and overhead resulting from increasing K. The optimized CQI levels push the balance point from $K = 14$ to 16, achieving higher multiuser diversity gain, compared to the pre-scheduling law.

Fig. 7 also shows that the proposed scheme supports much more MTs than the sum-rate upper-bound strategy. Given 1000 subcarrier symbols, our scheme is able to schedule more than 35 MTs, while the sum-rate upper-bound strategy can only schedule up to 22 MTs. Our scheme is also able to support a smaller size of radio resources (i.e., R) without decreasing the number of MTs it can schedule. In contrast, the sum-rate upper-bound strategy has to reduce from 22 MTs to 10 MTs, when R decreases from 1000 to 800 subcarrier symbols.

Note that the throughput in the figure corresponds to the case where the SBSs transmit as many bits as they are allowed to, which may consist of complete packets and up to two incomplete packets (which are at the beginning and/or the end of a transmission). This is of practical interest, as the same SBS is likely to be chosen to resume the incomplete packet due to time coherence of wireless channels.

Next, our evaluation is carried out separately from the aspects of overhead and sum rate, which substantiates the significant throughput improvement of the proposed scheme.

B. Sum Rate

Fig. 8 shows that the proposed scheme approaches the sum-rate upper bounds that can only be achieved by globally coordinating all the SBSs (see Appendix A). (a) Consider the case where $B > 1$. When $K \geq 15$, our scheme indistinguishably approaches the global upper bound (GUB) using precise CQI ($B = \infty$), and it can achieve 98% of the GUB using only 4-bit CQI ($B = 4$). The GUB is the maximum sum rate that the small cell system using linear MIMO techniques would achieve if the CQI is precise at the SBSs, and it is given by (20). When $K < 15$, our scheme is more likely to have antennas unscheduled, compared to the centralized strategy which provides the GUB. In this case, our scheme still achieves 91.55% of the GUB. (b) Proceed to consider the case where $B = 1$. Our scheme can achieve 94.4% of the lowest upper bound (LUB). The LUB is the maximum sum rate that the small cell system can achieve if the CQI is 1 bit, and it is given by (24).

The figure also shows that our scheme is 252.2% better than the conventional frequency-reuse strategy. Such an improvement leads to the substantial throughput gain our scheme has against the conventional strategy in Fig. 7, since the two approaches have similar overhead, as will be discussed later.

In the figure, we are also able to confirm that our optimized CQI design can substantially improve the sum rate in comparison with the existing designs of [33]. For $B = 1$, the sum-rate improvement is up to 26.4%. This substantiates the throughput improvement that the optimized CQI design made in Fig. 7.

In Fig. 9, we compare the sum rates of the optimized CQI design and the existing design from the perspective of ρ. We are able to further confirm that the sum-rate improvement that the optimized CQI design can make is increasingly large with ρ. Consider the growth of ρ from 20 dB to 40 dB. For $B = 1$, the sum-rate improvement increases from 2.1 dB to 6.5 dB. For $B = 2$, it increases from 1.0 dB to 3.1 dB.

There is also an interesting finding regarding the two existing CQI designs, that is, the pre-scheduling law outperforms the post-scheduling law, as shown in Fig. 8. This

TABLE I
OVERHEAD COMPARISON (IN KBPS)

	Proposed scheme		Centralized strategy		Reuse strategy	
B	$K = 20$	40	20	40	20	40
1	11.2	14.4	23.4	31.6	8.8	10.0
2	12.2	16.4	31.4	47.6	9.8	12.0
3	13.2	18.4	39.4	63.6	10.8	14.0
4	14.2	20.4	47.4	79.6	11.8	16.0

contradicts the conclusion drawn in [33]. The reason is that the post-scheduling CDF, which accounts for all combinations of $L_1, \cdots, L_N$ (resulting from the decoding failure at the SBSs), is substantially biased from the CDFs at individual Tx antennas with a given L_n ($n = 1, \cdots, N$). The conclusion we draw is that the post-scheduling law is susceptible to the decoding failure at the SBSs. Therefore, it is unsuitable for small cells with wireless backhaul. For this reason, we do not plot the post-scheduling law in Figs. 7 and 9.

C. Overhead Comparison

Table I shows that the proposed scheme can cut 80% of the overhead, compared to the centralized strategy which provides the GUB and LUB [19]. Here, the overhead is based on a single scheduling process (as illustrated in Fig. 3) and a super-frame duration of 20 ms, and it takes both the backhaul and access links into account. For example, when $B = 4$ and $K = 40$, the overhead is 20.4 Kbps for our scheme and 79.6 Kbps for the existing centralized strategy. Given the similar sum rates of the two approaches (as discussed earlier), the 80% reduced overhead accounts for the significant throughput improvement that our scheme achieved, as shown in Fig. 7.

The table also shows that the overhead of our scheme grows much more slowly with B and K, compared to the centralized strategy. This means that our scheme is able to schedule larger K than the centralized strategy, as concluded from Fig. 7.

The conventional reuse strategy requires less overhead, compared to the above two approaches. This is obvious, as the MTs only feed back a CQI and also the index of a MT consists of fewer bits (i.e., reduce $(B + \lceil \log_2 K \rceil)K$ to BK in Fig. 3 ①; reduce $N_T(\lceil \log_2 K \rceil + \lceil \log_2 S_T \rceil)$ to $N_T(\lceil \log_2 \frac{K}{N_T} \rceil + \lceil \log_2 S_T \rceil)$ in Fig. 3 ④). Compared to this strategy, Table I shows our scheme only requires $\frac{1}{4}$ extra overhead. Given the 252.5% higher sum rate our scheme can achieve (as shown in Fig. 8), such a small amount of extra overhead leads to a dramatic throughput increase of 154.3%, as shown in Fig. 7.

D. Small-cell Packet Size

In Fig. 7, we consider the scenario where the SBSs transmit as many bits as they are allowed to. In that case, it is possible that a SBS is selected to retransmit a packet which has been partly transmitted by another SBS. If the packet size is large, the transmission repetition is a waste; if the packet size is too small, the overhead of MT-ACK/NACK messages becomes overwhelming. Therefore, it is important to choose appropriate packet sizes on the small cell links.

Let w denote the packet size on the small cell access link. It is unnecessary for w to be equal to the packet size on the backhaul link, W. We test a range of w to determine the best packet size on the small cell link.

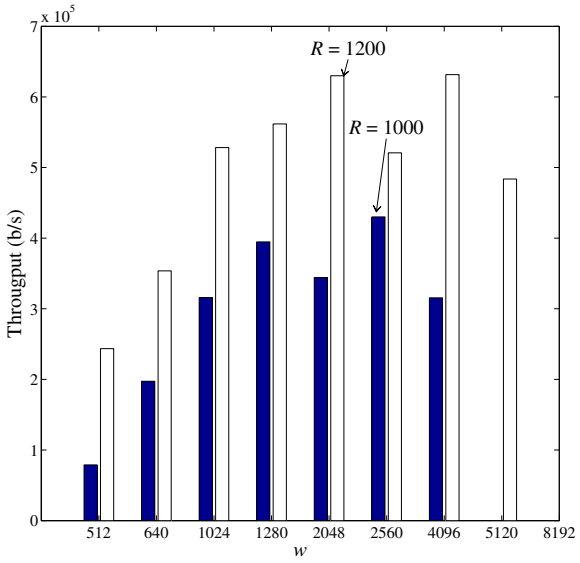

Fig. 10. Throughput versus w, where $K = 10$, $N = 4$, $\rho = 20$ dB, $S_T = 12$, $S = 20$, and $W = 2048$. In terms of w, we take the values that lead to an integral number of segments in the buffers, which result in an integral number of MT-ACK/NACK channels.

Fig. 10 reveals the best sizes of the small-cell packets. Specifically, when $R = 1000$ subcarrier symbols, the best size is 2560, which provides a throughput of 4.25×10^5 b/s. When $R = 1200$ subcarrier symbols, the best size is 2048, and the throughput we can achieve is 6.32×10^5 b/s. Note that in the figure, the throughput is analytically obtained by replacing $\log_2(1 + Q_i)$ with $\lfloor \frac{r}{w} \log_2(1 + Q_i) \rfloor w$ in (7). In other words, the SBSs only transmit complete packets, and the transmission repetition problem is eliminated.

Finally, it is worth noting that the method for determining the best packet size can also be used to optimize other parameters, such as the buffer size, buffer threshold, and packet size on the backhaul link.

V. CONCLUSION

In this paper, we proposed a new decentralized, user-centric, scheduling scheme for mobile small cells, which allows both the SBSs and MTs to participate in making the scheduling decisions in a distributed manner. We also optimized the CQI quantization levels, minimizing the feedback requirement. Analytical results, confirmed by simulations, show that the proposed scheme is able to improve the throughput by 139.5%, compared to the existing methods. This is the result of the 80% reduced overhead and 94.4% of the sum-rate upper bound that our scheme can achieve.

APPENDIX

A. Sum-Rate Upper Bounds

Fig. 11 illustrates the idealized centralized scheme which provides the GUB and LUB [19]. It is reformulated in a new way that leads to the analysis of the upper bounds. In the

scheme, every MT feeds back its CQI with regard to every Tx antenna. This information is gathered at the SGN. Using the information, the SGN serially selects the pairs of SBS Tx antenna and MT that can achieve the largest data rates.

A. GUB ($B = \infty$)

Let L_m denote the number of pairs of Tx antenna and MT prior to the m-th selection ($m = 1, \cdots, N$). In each pair, the SBS of the antenna has a threshold exceeded buffer for the MT. To derive the GUB, we need the PMF of L_m.

Define l_n to be the number of buffers exceeding the threshold at antenna n. $n = 1, \cdots, N$. Prior to the first selection, $L_1 = \sum_{n=1}^{N} l_n$. We can write the joint PMF of $l_1, \cdots, l_N$ as

$$\Pr(l_1, \cdots, l_N) = \prod_{n=1}^{N} \left[\binom{K}{l_n} p_1^{l_n} (1-p_1)^{K-l_n} \right], \quad (15)$$

where $\binom{K}{l_n} p_1^{l_n} (1-p_1)^{K-l_n}$ is the probability at antenna n.

Note that in this scheme, $l_1, \cdots, l_N$ are continually updated, i.e., once a pair of Tx antenna (i.e., n) and MT is selected, l_n is discarded and any other antenna m that has threshold exceeded buffers for the MT subtracts l_m by one.

We evaluate the probability of updating $l_1, \cdots, l_N$. Without loss of generality, we assume antennas 1 to m were previously selected. Then, the joint conditional probability of current l_n ($n = m+1, \cdots, N$) is given by

$$\begin{aligned} &\Pr(l_{m+1} - \nu_{m+1}, \cdots, l_N - \nu_N \,|\, l_m, \cdots, l_N) = \\ &\frac{1/\sum_{n=m}^{N} l_n}{\Pr(l_m, \cdots, l_N)} \underbrace{l_m \prod_{n=m+1}^{N} \left[p_1^{\nu_n} (1-p_1)^{1-\nu_n} \right]}_{(a)} \times \\ &\underbrace{\binom{K-m+1}{l_m} p_1^{l_m} (1-p_1)^{K-m-l_m+1}}_{(b)} \times \\ &\underbrace{\prod_{n=m+1}^{N} \left[\binom{K-m}{l_n - \nu_n} p_1^{l_n - \nu_n} (1-p_1)^{K-m-l_n+\nu_n} \right]}_{(c)} \end{aligned} \quad (16)$$

where $\nu_n \in \{0, 1\}$ indicates whether Tx antenna n has the buffer for the m-th selected MT, denoted by π_m, exceed the threshold $\frac{S_T}{S}$. If it does, $\nu_n = 1$; otherwise, $\nu_n = 0$. Besides, ν_n must satisfy $l_n - \nu_n \geq 0$. Details are following.

$1/\sum_{n=m}^{N} l_n$ is the probability of selecting a pair. (a) is the probability of the buffer for MT π_m exceeding the threshold at the antennas with $\nu_n = 1$. (b) is the probability of l_m threshold exceeded buffers at antenna m. (c) is $\Pr\left(l_{m+1} - \nu_{m+1}, \cdots, l_N - \nu_N\right)$, and can be given by (15).

We can obtain the PMF of L_m by recursively multiplying (16) to (15), since $L_{m+1} = \sum_{n=m+1}^{N} (l_n - \nu_n)$.

Next, we proceed to derive the GUB. We let $f_{\gamma_{(m)}}(x)$ and $F_{\gamma_{(m)}}(x)$ denote the PDF and CDF of the SINRs prior to the m-th selection, since $\gamma_{k,j}$ is i.i.d. ($k = 1, \cdots, K$ and $j = 1, \cdots, N$). We also let $f_{\gamma_{\pi_m}}(x)$ and $F_{\gamma_{\pi_m}}(x)$ denote the PDF and CDF of the SINR of the m-th selected pair.

Clearly, $\gamma_{(m)} \leq \gamma_{\pi_{m-1}}$. For $2 \leq m \leq N$, we also have

$$f_{\gamma_{\pi_{m-1}}}(x) = L_{m-1} f_{\gamma_{(m-1)}}(y) F_{\gamma_{(m-1)}}^{L_{m-1}-1}(x). \quad (17)$$

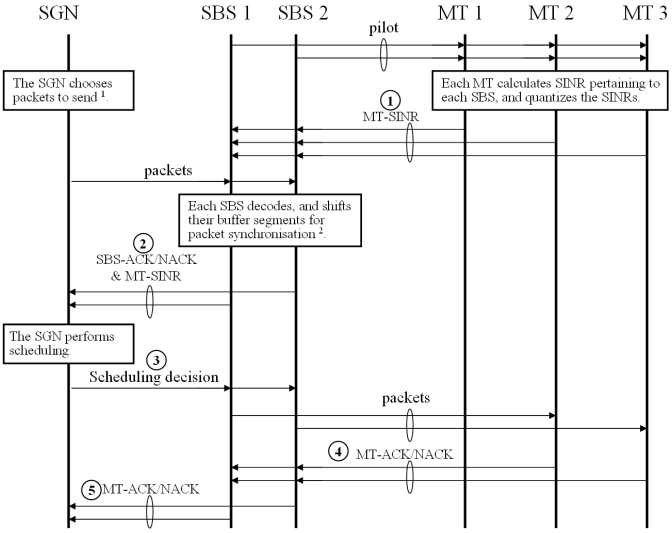

Fig. 11. The reformulated protocol of the centralized strategy of [19], where two SBSs and three MTs are considered. The numbers of signalling bits are summarized as: ① BN_TK; ② $S_TN_T^2 + BN_TK$; ③ $N_T\lceil \log_2 K \rceil$; ④ N_TS; and ⑤ $N_T\lceil \log_2 S_T \rceil$. Note that in this scheme we consider the same segment mapping mechanism and the MT-ACK/NACK channel pre-allocation as proposed for our decentralized scheme in Section II-B. Otherwise, ④ increases substantially to up to $N_T(N_T-1)S_T \log_2 S$, due to coordinating the SBSs to discarding successful packets.

Exploiting Order Statistics [27, Lemma 1], we can write the conditional PDF of $\gamma_{(m)}$ given $\gamma_{(m)} \leq \gamma_{\pi_{m-1}}$, as

$$f_{\gamma_{(m)}|\gamma_{\pi_{m-1}}}(x|\, y) = f_{\gamma_{(m-1)}}(x)/F_{\gamma_{(m-1)}}(y).$$

Integrating the product of $f_{\gamma_{(m)}|\gamma_{\pi_{m-1}}}(x|\, y)$ and $f_{\gamma_{\pi_{m-1}}}(y)$ with respect to y, we obtain

$$\begin{aligned} f_{\gamma_{(m)}}(x) &= \frac{L_{m-1}}{L_{m-1}-1} f_{\gamma_{(m-1)}}(x) \left[1 - F_{\gamma_{(m-1)}}^{L_{m-1}-1}(x) \right] \\ F_{\gamma_{(m)}}(x) &= \frac{L_{m-1}}{L_{m-1}-1} F_{\gamma_{(m-1)}}(x) - \frac{F_{\gamma_{(m-1)}}^{L_{m-1}}(x)}{L_{m-1}-1}. \end{aligned} \quad (18)$$

Recursively substituting (18) into (17), we obtain $f_{\gamma_{\pi_m}}(x)$ as

$$f_{\gamma_{\pi_1}}(x) = -\frac{NL_1}{\rho} \sum_{j_1=0}^{L_1-1} \left\{ \binom{L_1-1}{j_1} \left(-e^{-\frac{Nx}{\rho}} \right)^{j_1+1} \right\},$$

$$f_{\gamma_{\pi_2}}(x) = -\frac{NL_2}{\rho} \sum_{J_2=0}^{L_2-1} \sum_{j_1=2}^{L_1} \sum_{\sum_{i=2}^{L_1} z_{1,i} = J_2} \left\{ \frac{\prod_{i=2}^{L_1} \binom{L_1}{i}^{z_{1,i}}}{(1-L_1)^{J_2+1}} \right.$$

$$\left. \binom{L_2-1}{J_2} \binom{L_1-1}{j_1-1} \binom{J_2}{\substack{z_{1,i}: \\ i=2,\cdots,L_1}} \left(-e^{-\frac{Nx}{\rho}} \right)^{(j_1+\sum_{i=2}^{L_1} i z_{1,i})} \right\},$$

and (19). Finally, the GUB is given by (20), where $l_n^{(m)}$ is the number of threshold exceeded buffers for the n-th unselected Tx antenna prior to the m-th selection ($n = 1, \cdots, N-m+1$). Initially, $l_n^{(1)} = l_n$ for $n = 1, \cdots, N$. The conditional probability can be calculated using (16). $f_{\gamma_{\pi_m}}(x|\, L_1, \cdots, L_m)$ is obtained from (17), as (17) is on the condition of $L_1, \cdots, L_m$.

Using the integral identity [27, eq. 7], $C_{\pi_m}(L_1, \cdots, L_m)$

$$f_{\gamma_{\pi_m}}(x)=\frac{L_m}{L_{m-1}-1}\frac{N}{\rho}\sum_{J_m=0}^{L_{m-1}}\sum_{j_{m-1}=2}^{L_{m-1}}\sum_{\substack{k_{t,i}:\sum_{i=2}^{L_t}k_{t,i}=j_{t+1},\\ t=1,\cdots,m-2,i=2,\cdots,L_t}}\cdots\sum\sum_{\substack{z_{t,i}:\sum_{i=2}^{L_t}z_{t,i}=J_{t+1},\\ t=1,\cdots,m-1,i=2,\cdots,L_t}}\cdots\sum\left\{\frac{\binom{L_m-1}{J_m}\binom{L_{m-1}}{j_{m-1}}}{(1-L_m)^{J_m}}\left[\prod_{t=1}^{m-2}\prod_{i=2}^{L_t}\binom{L_t}{i}^{z_{t,i}+k_{t,i}}\right]\right.$$
$$\left.\frac{\binom{J_m}{z_{m-1,i}:\atop i=2,\cdots,L_{m-1}}\left[\prod_{i=2}^{L_{m-1}}\binom{L_{m-1}}{i}^{z_{m-1,i}}\right]\prod_{t=1}^{m-2}\left[\binom{j_{t+1}}{k_{t,i}:\atop i=2,\cdots,L_t}\binom{J_{t+1}}{z_{t,i}:\atop i=2,\cdots,L_t}\right]}{(1-L_{m-2})^{j_{m-1}+\sum_{i=2}^{L_{m-1}}iz_{m-1,i}}\prod_{t=2}^{m-2}(1-L_{t-1})^{\sum_{i=2}^{L_t}i(z_{t,i}+k_{t,i})}}\left(-e^{-\frac{Nx}{\rho}}\right)^{\sum_{i=2}^{L_1}i(z_{1,i}+k_{1,i})}\sum_{i=2}^{L_1}ik_{1,i}\right\}. \tag{19}$$

$$C_{\mathcal{I}}=\sum_{\substack{\forall l_i^{(j)}\geq 0,\, i=1,\cdots,N,\\ j=1,\cdots,N-i+1}}\cdots\sum\sum_{m=1}^{N}\left\{\Pr(l_1,\cdots,l_N)\prod_{n=2}^{m}\left[\Pr\left(l_1^{(n)},\cdots,l_{N-n+1}^{(n)}|\,l_1^{(n-1)},\cdots,l_{N-n+2}^{(n-1)}\right)\right.\right.$$
$$\left.\left.\times\underbrace{\int_0^{\infty}\log_2(1+x)\,f_{\gamma_{\pi_m}}\left(x|\underbrace{\sum_{n=1}^{N}l_n^{(1)}}_{L_1},\cdots,\underbrace{\sum_{n=1}^{N-m+1}l_n^{(m)}}_{L_m}\right)dx}_{C_{\pi_m}(L_1,\cdots,L_m)}\right]\right\}, \tag{20}$$

is given by (21), (22); and (23) for $m\geq 3$.

$$C_{\pi_1}(L_1)=L_1\sum_{j_1=0}^{L_1-1}\left[\binom{L_1-1}{j_1}\frac{e^{\frac{N}{\rho}(j_1+1)}\Gamma\left(0,\frac{N}{\rho}(j_1+1)\right)}{\log 2(-1)^{j_1}(j_1+1)}\right]. \tag{21}$$

B. LUB ($B=1$)

In this case, a CQI indicates whether a MT preselects a Tx antenna. For each Tx antenna, the SGN selects one of the MTs that preselected the antenna. The selection is arbitrary since no information is available for the SGN to differentiate the MTs. As a result, the scheduling is actually carried out one Tx antenna after another, choosing a MT for each antenna. The maximum number of unscheduled MTs for scheduling the m-th Tx antenna is $(K-m+1)$.

Using Discrete Order Statistics, we have

$$\Pr(\hat{\gamma}_{\pi_m}=Q_i)\leq \left(1-e^{-\frac{NQ_{i+1}}{\rho}}\right)^{K-m+1}-\left(1-e^{-\frac{NQ_i}{\rho}}\right)^{K-m+1}$$

where $\hat{\gamma}_{\pi_m}$ is the quantized SINR of the MT that is scheduled for the m-th Tx antenna. $i=1,2$. Then, the LUB is given by

$$C_{\mathcal{F}}\leq\log_2(1+Q_2)\sum_{m=1}^{N}\left[1-\left(1-e^{-\frac{NQ_2}{\rho}}\right)^{K-m+1}\right] \tag{24}$$

where the optimal Q_2 can be obtained by solving the first-order necessary condition: $\frac{\partial C_{\mathcal{F}}}{\partial Q_2}=0$. A bisection method can be used to solve the condition. The result is substituted to (24), achieving the LUB.

APPENDIX

B. Conventional Frequency-Reuse Scheme

The PDF and CDF of the received SNR of MRC are [32]

$$f_{\gamma}(x)=\frac{N}{(N-1)!\,\rho}\left(\frac{Nx}{\rho}\right)^{N-1}e^{-\frac{N}{\rho}x},$$
$$F_{\gamma}(x)=1-e^{-\frac{N}{\rho}x}\sum_{l=0}^{N-1}\frac{1}{l!}\left(\frac{Nx}{\rho}\right)^{l}.$$

For N small cells that have an overlapping coverage area, the frequency resource is divided by N, one for each cell. Also, we evenly divide the MTs that are within the overlapping coverage area into N groups. Each group is served by one SBS. Now the SINR of the MTs which are scheduled in those small cells has the PDF, as given by

$$f_{\gamma_{\pi_n}|L_n}(x|l)=lf_{\gamma}(x)F_{\gamma}^{l-1}(x)$$
$$=\frac{-l\,N}{(N-1)!\,\rho}\sum_{i=0}^{l-1}\sum_{\sum_{t=1}^{N}k_t=N}\left\{\frac{\binom{l}{i}\binom{i}{k_1,\cdots,k_N}}{\prod_{t=1}^{N}\left[(t-1)!\right]^{k_t}}\left(-e^{-\frac{N}{\rho}x}\right)^{i+1}\left(\frac{Nx}{\rho}\right)^{N-1+\sum_{t=1}^{N}(t-1)k_t}\right\}$$

where L_n is the number of threshold-exceeded buffers at SBS n. The PMF of L_n is

$$\Pr(L_n=l)=\binom{\frac{K}{N}}{l}p_1^l(1-p_1)^{\frac{K}{N}-l},$$

where $l=0,1,\cdots,\frac{K}{N}$. Using the integral identity [27, eq. 7], the sum rate is given by (25).

$$C_{\pi_2}(L_1,L_2)=\frac{L_2}{L_1-1}\sum_{J_2=0}^{L_2-1}\sum_{j_1=2}^{L_1}\sum_{\sum_{i=2}^{L_1} z_{1,i}=J_2} \left\{\frac{\binom{L_2-1}{J_2}\binom{L_1}{j_1}\binom{J_2}{z_{1,i}:\atop i=2,\cdots,L_1}\prod_{i=2}^{L_1}\binom{L_1}{i}^{z_{1,i}}}{(1-L_1)^{J_2}} j_1(-e^{\frac{N}{\rho}})^{j_1+\sum_{i=2}^{L_1} iz_{1,i}}\frac{\Gamma\left(0,\frac{N}{\rho}(j_1+\sum_{i=2}^{L_1} iz_{1,i})\right)}{j_1+\sum_{i=2}^{L_1} iz_{1,i}}\right\}, \tag{22}$$

$$C_{\pi_m}(L_1,\cdots,L_m)=\frac{L_m}{L_{m-1}-1}\sum_{J_m=0}^{L_m-1}\sum_{j_{m-1}=2}^{L_{m-1}}\sum_{\substack{k_{t,i}:\sum_{i=2}^{L_t} k_{t,i}=j_{t+1},\\ t=1,\cdots,m-2,i=2,\cdots,L_t}}\cdots\sum\sum_{\substack{z_{t,i}:\sum_{i=2}^{L_t} z_{t,i}=J_{t+1},\\ t=1,\cdots,m-1,i=2,\cdots,L_t}}\cdots\sum \left\{\frac{\binom{L_m-1}{J_m}\binom{L_{m-1}}{j_{m-1}}}{(1-L_m)^{J_m}}\right.$$
$$\frac{\left[\prod_{t=1}^{m-2}\prod_{i=2}^{L_t}\binom{L_t}{i}^{z_{t,i}+k_{t,i}}\right]\left[\prod_{i=2}^{L_{m-1}}\binom{L_{m-1}}{i}^{z_{m-1,i}}\right]\left\{\prod_{t=1}^{m-2}\left[\binom{j_{t+1}}{k_{t,i}:\atop i=2,\cdots,L_t}\binom{J_{t+1}}{z_{t,i}:\atop i=2,\cdots,L_t}\right]\right\}}{(1-L_{m-2})^{j_{m-1}+\sum_{i=2}^{L_{m-1}} iz_{m-1,i}}\prod_{t=2}^{m-2}(1-L_{t-1})^{\sum_{i=2}^{L_t} i(z_{t,i}+k_{t,i})}} \tag{23}$$
$$\left.\binom{J_m}{z_{m-1,i}:\atop i=2,\cdots,L_{m-1}}(-e^{\frac{N}{\rho}})^{\sum_{i=2}^{L_1} i(z_{1,i}+k_{1,i})}\frac{\Gamma\left(0,\frac{N}{\rho}\sum_{i=2}^{L_1} i(z_{1,i}+k_{1,i})\right)\sum_{i=2}^{L_1} ik_{1,i}}{\sum_{i=2}^{L_1} i(z_{1,i}+k_{1,i})}\right\}.$$

$$C_{MRC}(x)=\frac{1}{N!\ln 2}\sum_{n=1}^{N}\sum_{l=0}^{K/N}\sum_{i=0}^{l-1}\sum_{\sum_{t=1}^{N} k_t=N}\left\{\frac{1}{\prod_{t=1}^{N}\left[(t-1)!\right]^{k_t}}\frac{(-l)^i\Pr(l)\binom{l}{i}\binom{i}{k_1,\cdots,k_N}}{(\rho/N)^{N+\sum_{t=1}^{N}(t-1)k_t}}\left(N-1+\sum_{t=1}^{N}(t-1)k_t\right)!\right.$$
$$\left.e^{\frac{N}{\rho}(i+1)}\sum_{j=1}^{N+\sum_{t=1}^{N}(t-1)k_t}\frac{\Gamma\left(j-N-\sum_{t=1}^{N}(t-1)k_t,\frac{N}{\rho}(i+1)\right)}{\left(\frac{N}{\rho}(i+1)\right)^j}\right\} \tag{25}$$

REFERENCES

[1] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Trans. Veh. Technol.*, vol. 6, no. 1, pp. 37–43, Mar. 2011.

[2] "Metro cells: a cost-effective option for meeting growing capacity demands," Alcatel-Lucent, Tech. Rep., 2011.

[3] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.

[4] T. S. Rappaport, *Wireless Communications Principles and Practice*, 2nd ed. Prentice Hall, 2002.

[5] S. Jing, D. N. C. Tse, J. B. Soriaga, J. Hou, J. Smee, and R. Padovani, "Multicell downlink capacity with coordinated processing," *EURASIP J. Wireless Commun. Netw.*, Jan. 2008, article ID 586878.

[6] D. Wang, X. Wang, and X. Cai, "Optimal power control for multi-user relay networks over fading channels," *IEEE Trans. Wireless Commun.*, vol. 10, no. 1, pp. 199–207, Jan. 2011.

[7] E. Bjornson, N. Jalden, M. Bengtsson, and B. Ottersten, "Optimality properties, distributed strategies, and measurement-based evaluation of coordinated multicell ofdma transmission," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6086–6101, Dec. 2011.

[8] D. Gesbert *et al.*, "Multi-cell mimo cooperative networks: a new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[9] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, 2012, to appear.

[10] D. J. Love *et al.*, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[11] "LTE; Evolved universal terrestrial radio access (E-UTRA); requirements for support of radio resource management (3GPP TS 36.133 version 11.1.0 Release 11)," 3rd Generation Partership Project (3GPP)," technical specification, June 2012.

[12] P. Beming *et al.*, "LTE-SAE architecture and performance," *Ericcson Rev.*, no. 3, pp. 98–104, 2007.

[13] "Overview of 3GPP Release 10 v0.1.0," 3rd Generation Partership Project (3GPP), Tech. Rep., Apr. 2011.

[14] V. Jungnickel *et al.*, "Coordinated multipoint trials in the downlink," in *Proc. 2009 IEEE GLOBECOM Workshops*, pp. 1–7.

[15] ——, "Field trials using coordinated multi-point transmission in the downlink," in *Proc. 2010 IEEE International Symp. Personal, Indoor Mobile Radio Commun. Workshops*, pp. 440–445.

[16] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, 2011.

[17] "The advanced LTE toolbox for more efficient delivery of better user experience," Nokia Siemens Networks, LTE-Advanced, Karaportti 3, ESPOO, Finland, Technical white paper, 2011.

[18] P. Zetterberg and N. N. Moghadam, "An experimental investigation of SIMO, MIMO, interference-alignment (IA) and coordinated multi-point (CoMP)," in *Proc. 2012 International Conf. Syst., Signals Image Process.*, pp. 211–216.

[19] W. Ni, Z. Chen, I. B. Collings, and H. Suzuki, "Sum-rate scheduling of decode-and-forward broadcast channel with limited-feedback," in *Proc. 2010 PIMRC*, vol. 3, pp. 2460–2465.

[20] W. Ni, Z. Chen, H. Suzuki, and I. B. Collings, "Performance analysis of scheduling in decode-and-forward broadcast channel with limited-feedback," in *Proc. 2010 IEEE GLOBECOM*, pp. 1–5.

[21] D. Lecompte and F. Gabin, "Evolved multimedia broadcast/multicast service (eMBMS) in LTE-advanced: overview and Rel-11 enhancements," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 68–74, 2012.

[22] "LTE; evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (3gpp ts 36.211 version 10.0.0 release 10)," 3rd Generation Partership Project (3GPP), Technical Specification ETSI TS 136 211 v10.0.0 (2011-01), Jan. 2011.

[23] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. Elsevier, 2011.

[24] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G Evolution HSPA and LTE for Mobile Broadband*. Elsevier, 2007.

[25] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*, Chapter 28. Technical Publishing, 2002.

[26] P. Telagarapu, G. Naidu, and K. Chiranjeevi, "Analysis of coding techniques in WiMAX," *International J. Comput. Applications*, vol. 22, no. 3, pp. 19–26, May 2011.

[27] C. J. Chen and L. C. Wang, "Performance analysis of scheduling in

multiuser MIMO systems with zero-forcing receivers," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1435–1445, Sept. 2007.
[28] R. H. Y. Louie, M. R. McKay, and I. B. Collings, "Maximum sum-rate of MIMO multiuser scheduling with linear receivers," *IEEE Trans. Wireless Commun.*, vol. 57, no. 11, pp. 3500–3510, Nov. 2009.
[29] H. A. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Wiley, 2003.
[30] A. Mordecai, *Nonlinear Programming: Analysis and Methods*. Dover Publications, 2003.
[31] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control*. John Wiley & Sons, Inc., 2003.
[32] M. Torabi, D. Haccoun, and W. Ajib, "BER performance analysis of multiuser diversity with antenna selection in MRC MIMO systems," in *Proc. 2009 IEEE GLOBECOM*, pp. 1–6.
[33] J. L. Vicario *et al.*, "A throughput analysis for opportunistic beamforming with quantized feedback," in *Proc. 2006 PIMRC*, pp. 1–6.
[34] "Draft amendment to IEEE standard for local and metropolitan area networks, part 16: air interface for fixed and mobile broadband wireless access systems, multihop relay specification," IEEE, Tech. Rep. IEEE P802.16j, D2, Dec. 2007.

Wei Ni received the B.E. and Ph.D. degrees in Electronic Engineering from Fudan University, Shanghai, China, in 2000 and 2005, respectively. Currently he is a Senior Research Scientist in the Devision of CSIRO Computational Informatics (CCI), CSIRO, Australia. Prior to this he was a Research Scientist and Deputy Project Leader at the Bell Labs R&I Center, Alcatel-Lucent (2005-2008), and a Senior Research Engineer at Devices R&D, Nokia (2008-2009). His research interests include Multiuser MIMO, Relay Mesh Networks, Radio Resource Management, and Scheduling, etc. He serves as an Editorial Board Member for Hindawi Journal of Engineering since 2012.

Iain B. Collings (S92-M95-SM02) received the B.E. degree in Electrical and Electronic Engineering from the University of Melbourne in 1992, and the Ph.D. degree in Systems Engineering from the Australian National University in 1995. Currently he is the Deputy Chief (Research) of CCI in the CSIRO. Prior to this he was an Associate Professor at the University of Sydney (1999-2005); a Lecturer at the University of Melbourne (1996-1999); and a Research Fellow in the Australian Cooperative Research Center for Sensor Signal and Information Processing (1995). He has published over 250 research papers on mobile digital communications. He also served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the *Elsevier Physical Communication Journal*, and a Guest Editor for the *EURASIP Journal on Advanced Signal Processing*.

Ren Ping Liu received B.E. and M.E. degrees from Beijing University of Posts and Telecomms, China, and the Ph.D. degree in Electrical and Computer Engineering from the University of Newcastle, Australia. He joined CSIRO in 1995, where he is now a Principle Scientist. He has been heavily involved in commercial projects ranging from QoS design and TCP/IP inter-networking to next generation network architectures. His interests include modeling, resource allocation and analysis in IEEE 802.11, Mesh, Sensor and Cognitive Radio Networks.