Provisioning Statistical QoS for Coordinated Communications with Limited Feedback

Mohammad G. Khoshkholgh*, Keivan Navaie‡, Kang G. Shin†, Victor M. C. Leung*

*The University of British Columbia (m.g.khoshkholgh@gmail.com, vleung@ece.ubc.ca)
‡Lancaster University (k.navaie@lancaster.ac.uk), †The University of Michigan (kgshin@umich.edu)

Abstract—The capacity performance of ICIC has been extensively studied in coordinated multi-point transmissions (CoMP). In practice however, due to limited feedback, the acquired channel direction information (CDI), which is crucial for ICIC, is often partially available. Hence one may question whether the ICIC is able to meet the Quality-of-Service (QoS) requirements. This paper considers the optimal partitioning of the feedback bits in CoMP while accounting for the inter-cell interference cancellation (ICIC). In this paper, we adopt a statistical model of QoS in CoMP by using the notion of effective capacity (EC). Utilizing EC we then formulate the system function as an optimization problem with the objective of maximizing the total EC subject to the limited feedback available to the cluster of base stations (BSs). Analytical bounds are then obtained on the EC performance which are then utilized as the base for algorithms that assign feedback bits among the user equipments (UEs) and BSs. Using simulations we then investigate the accuracy of the obtained bounds and highlight practical system designs for dealing with stringent delay requirements. Of crucial practical importance, the findings of this paper also indicates that in CoMP there is an optimal cluster size for a given feedback capacity that maximizes the corresponding EC.

I. INTRODUCTION

Multi-antenna (MIMO) technology can increase the capacity of uplink/downlink communications in proportion to the number of antennas in the BSs. The capacity increase however is limited due to the inter-cell interference (ICI). An effective way to tackle the negative impact of the ICI is coordination/cooperation among adjacent BSs. In the coordinated multi-point (CoMP) systems, a number of BSs connected to each other/central processor (CP) through a high-speed, low-latency backhaul to facilitate clustered communications. Such a deployment is the building block of recently proposed cloud radio access networks (C-RANs) [1, 2].

Various technical aspects of CoMP have been investigated in the literature, see, e.g, [6–8]. Beamforming designs subject to the signal-to-interference plus noise (SINR) constraints at the user equipments (UEs) are explored in [6]. Clustering in CoMP is investigated in [7] where a novel dynamic clustering and interference coordinator system is proposed and prototyped based on distributed CSMA/CA protocol. Further [8] deals with multi-cell scheduling in CoMP systems. Energy-driven resource allocation (RA) and beamforming design for CoMP systems are also inspected in [9–11]. The energy efficient RA problem in [9, 11], is transformed into a subtractive form

which is analytically tractable and results in reasonable computational complexity. It is further shown in [9] that an energy efficient in CoMP results in an spectrum-efficient solution.

In practice however, the performance of CoMP systems is much smaller than that of suggested by the analysis. For instance, a large cluster size is required to achieve a high capacity growth by increasing transmission power. Nevertheless, the cluster-size is limited due to the excessive resources required for acquiring accurate channel direction information (CDI) [3]. This particularly happens in frequency division duplexing (FDD) systems, where BSs require the quantized UEs' CDI for beamforming design. It is further shown in [15] that limited feedback reduces network's degrees-of-freedom, where capacity is used as the main performance metric.

Negative effects of limited feedback and quantization inaccuracies on the capacity is further investigated in [12], where it is shown that careful feedback bit partitioning (FBP) significantly improves capacity. FBP obtained through a brute-force search is proposed in [14] to develop adaptive inter-cell interference cancellation (ICIC) techniques, while clusters with only three BSs are considered. Part of the issue causing small capacity performance is due to the fixed feedback capacity per cell as it is shown in [13]. An alternative approach is to optimally distribute the cluster's total assigned feedback resources amongst the cells/BSs to maximize the spectral efficiency and/or energy efficiency.

In addition to the spectral and energy efficiency, delay is equivalently important in many application scenarios such as those with very limited latency requirements envisaged in the 5G systems, see, e.g., [2]. To address this issue here we investigate the delay performance of the clustered structures with the ICIC beamforming. Here we focus on the statistical delay and formulate the system performance through the *effective capacity* (EC) concept. EC conceptualize the capacity in a system with delay constraints, see, e.g., [17, 18]. The EC concept is utilized to analyze the BS selection in the networked MIMO systems [16] however, the perfect CDI is considered ignoring the impact of limited feedback capacity.

To the best of our knowledge the delay performance in limited feedback CoMP systems has not been investigated in the related literature. In limited-feedback ICIC beamforming the exact evaluation of the EC performance is rather challenging mainly due to the residual intra-cluster interference caused by

the quantized CDIs. In our analysis, adopting techniques from stochastic geometry, we obtain analytical bounds on the EC performance. The accuracy of the bounds are then examined through extensive simulations. Based on the achieved bounds we then obtain sub-optimal feedback resource allocation that maximizes the system EC performance with limited feedback capacity in each cluster. The proposed technique is in fact a FBP method which allocates the available feedback resources amongst the cells and UEs.

Simulation results show that the proposed algorithm outperforms other FBP algorithms in the literature. Finally, we highlight the importance of dynamic clustering and power control in conjunction with FBP for enhancing the total system EC, especially in cases where the delay exponent is large.

II. SYSTEM MODEL

We consider downlink in a cluster consisting of $|\mathcal{C}|$ heterogenous base stations (BSs), indexed by $c \in \mathcal{C}$ and connected to a central processor (CP). Let the cluster be a disk-shaped with radius R meters. We assume that BSs are positioned on a disk with radius $0 < D \leq R$ meters from the center of the cluster (see Fig. 1-LHR for an example). BSs share the same portion of the cluster coverage as the cell area. Each BS is equipped with N_t transmit antennas. In CoMP $N_t \geq |\mathcal{C}|$ [12, 14]. There is a single-antenna UE c associated with each BS c. The path-loss attenuation between UE c and BS c' is $\rho_{cc'} = (1 + d_{cc'})^{-\alpha}$, where $d_{cc'}$ and $\alpha > 2$ are distance and path-loss exponent. The transmitted signal undergoes flat-fading, $\boldsymbol{h}_{cc'} \in \mathbb{C}^{N_t \times 1}$, with the CDI (direction) $\tilde{\boldsymbol{h}}_{cc'} = \boldsymbol{h}_{cc'}/\|\boldsymbol{h}_{cc'}\|$. We assume $\boldsymbol{h}_{cc'} \; \forall c' \neq c$ are independent, and their elements are complex Gaussian random variables with zero mean and unit variance, i.e., Rayleigh fading.

We focus on frequency division duplex (FDD) systems, such as LTE-Advance. For such systems, the acquisition of channel direction information (CDI) often requires a feedback channel in the uplink. Let B_{tot} be the total assigned feedback capacity (in number of bits) to the cluster [13]. Random vector quantization (RVQ) [4, 5] is considered. Here we assume that perfect CDI is available to the receivers. UE c separately quantizes CDIs according to given constructed quantization code-books for intended CDI, $\mathcal{W}_{cc}$, and interfering CDIs, $\{\mathcal{W}_{cc'}, \forall c' \neq c\}$ [12]. UEs then transmit the indices of the selected code-words to the corresponding BSs via the designated feedback channel. To do this, UE c requires $\sum_{c' \in \mathcal{C}} B_{cc'}$ feedback bits to quantize the CDIs. Let $\hat{\boldsymbol{h}}_{cc'} = \max_{l=1,\ldots,2^{B_{cc'}}} |\tilde{\boldsymbol{h}}_{cc'}^{\dagger}[\mathcal{W}_{cc'}]_{(:,l)}|^2$, where $[\mathcal{W}_{cc'}]_{(:,l)}$ is the l-th column of code-book $\mathcal{W}_{cc'}$, i.e., the l-th code-word, and superscript $\dagger$ indicates the transpose conjugate as the quantized CDI of $\tilde{\boldsymbol{h}}_{cc'}$. It then reports the indices $l_{cc'}^* = \arg\max_l |\tilde{\boldsymbol{h}}_{cc'}^{\dagger}[\mathcal{W}_{cc'}]_{(:,l)}|^2$ to the BS c which then dequantizes $l_{cc'}^*$ to $\hat{\boldsymbol{h}}_{cc'}$ and shares them with other BSs.

Other BSs also provide BS c with the interfering channel directions between BS c and selected UEs in other cells, $\hat{\boldsymbol{h}}_{c'c,n}$ $\forall c' \neq c$. Upon receiving required CDIs, BS c constructs matrix $\hat{\boldsymbol{G}}_c = [\hat{\boldsymbol{h}}_{c'c} \forall c' \neq c] \in \mathbb{C}^{N_t \times (|\mathcal{C}|-1)}$ and produces coordinated beamforming vector $\hat{\boldsymbol{f}}_c \in \mathbb{C}^{N_t \times 1}$ as $\hat{\boldsymbol{f}}_c = \frac{(\boldsymbol{I} - \mathcal{P}(\hat{\boldsymbol{G}}_c))\hat{\boldsymbol{h}}_{cc}}{\|(\boldsymbol{I} - \mathcal{P}(\hat{\boldsymbol{G}}_c))\hat{\boldsymbol{h}}_{cc}\|}$

where $\mathcal{P}(\hat{\boldsymbol{G}}_c) = \hat{\boldsymbol{G}}_c(\hat{\boldsymbol{G}}_c^{\dagger}\hat{\boldsymbol{G}}_c)^{-1}\hat{\boldsymbol{G}}_c^{\dagger}$ is the projection operator. This beamforming technique is commonly referred to as *inter-cell interference cancellation* (ICIC) [12, 14]. In this model, SINR at c is

$$\gamma_{cc} = \frac{\rho_{cc} \frac{P_c}{N_t} |\boldsymbol{h}_{cc}^{\dagger}\hat{\boldsymbol{f}}_c|^2}{\sigma^2 + \sum_{c' \in \mathcal{C}-c} \rho_{cc'} \frac{P_{c'}}{N_t} |\boldsymbol{g}_{cc'}^{\dagger}\hat{\boldsymbol{f}}_{c'}|^2}, \qquad (1)$$

where $\hat{\boldsymbol{f}}_{c'}$ is the coordinated beamforming vector constructed at BS c' and is independent of $\hat{\boldsymbol{f}}_c$, σ^2 is AWGN and an inter-cluster interference contribution, and P_c is the transmission power at BS c. Since BSs do not access the channel quality information (CQI), the allocated power on each subcarrier is simply divided by the number of antennas, N_t. Furthermore, due to the mismatch between $\hat{\boldsymbol{h}}_{cc'}$ and $\boldsymbol{h}_{cc'}$, we have $|\boldsymbol{h}_{cc'}^{\dagger}\hat{\boldsymbol{f}}_{c'}|^2 \neq 0$ while $|\hat{\boldsymbol{h}}_{cc'}^{\dagger}\hat{\boldsymbol{f}}_{c'}|^2 = 0$. To model this mismatch and incorporate its impact on the system design, we adopt Quantization Cell Approximation (QCA) [12, 19]. In the rest of this paper, we use the parameter $\delta_{c'c''} = 2^{-\frac{B_{c'c''}}{N_t-1}}$.

III. PARTITIONING FEEDBACK CAPACITY FOR OPTIMIZING EFFECTIVE CAPACITY

Spectral efficiency is usually considered in the design of CoMP systems. However, it overlooks some prominent aspects of the UE's QoS requirements, such as delay. One way to incorporate delay in the design of the RA problem is to use the notion of *statistical delay* by EC [17]. Introducing the delay exponent θ_c, which addresses the delay requirement of UE c, the EC, ec_{cc}, is expressed as $ec_{cc} = \frac{-1}{\theta_c}\log(\mathbb{E}e^{-\theta_c R_{cc}})$ where R_{cc} is the instantaneous transmission rate and is equal to $R_{cc} = \log(1+\gamma_{cc})$. EC is a flexible performance metric and can cover diverse performance metrics pertinent to the wireless communication systems including achievable capacity ($\theta_c \to 0$) and outage capacity ($\theta_c \to \infty$). Considering the limited feedback capacity, we formulate the following optimization

$$\max_{\boldsymbol{B}} \sum_{c \in \mathcal{C}} ec_{cc} \quad s.t. \quad \sum_{c \in \mathcal{C}} \left(B_{cc} + \sum_{c' \in \mathcal{C}-c} B_{cc'} \right) = B_{tot}. (2)$$

We first start by deriving the EC of a given link cc.

A. Evaluation of the Effective Capacity

Proposition 1: An approximate of UE c's EC is

$$ec_{cc} \approx \frac{-1}{\theta_c}\log\left(\frac{(-1)^{N_t-|\mathcal{C}|}}{\Gamma(N_t-|\mathcal{C}|)} \int_0^{\infty} \frac{z^{N_t-|\mathcal{C}|-1}}{(1+z)^{\theta_c}} \frac{\partial^{N_t-|\mathcal{C}|}O(z)}{\partial z^{N_t-|\mathcal{C}|}} dz \right),$$

where $O(z)$ is given in the Appendix.

Proof: See Appendix.

Obtaining ec_{cc} is highly computationally complex. We also provide a more computationally efficient approximation. But, before proceeding further, we need the following result on the average data rate of a given link.
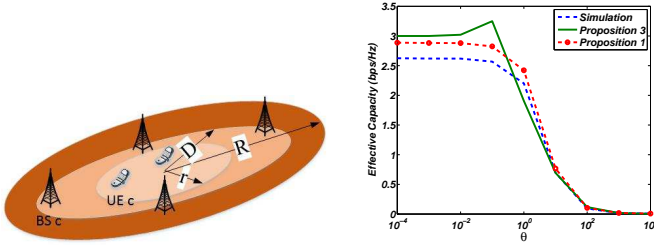
Fig. 1. LHS: A schematic of the system. RHS: The EC of the link versus θ_c for three BSs with equal transmission power, $N_t = 5$, and $\alpha = 4$. The allocated feedback bits are $B_{cc} = 8$ (distance 300m), $B_{cc'} = 6$ (distance 400m), and $B_{cc''} = 5$ (distance 500m).

Proposition 2: The average data rate of link cc is approximated by

$$\bar{R}_{cc} \approx \int_0^\infty \frac{e^{-N_t w \sigma^2}}{w} \prod_{c' \in \mathcal{C} - c} \frac{1}{1 + \rho_{cc'} P_{c'} \delta_{cc'} w} \times$$

$$\left(1 - \int_0^{\delta_{cc}} \frac{\frac{f_{\sin^2(\theta_{cc})}(x)}{1 + w P_c \rho_{cc} \delta_{cc}}}{(1 + w P_c \rho_{cc}(1-x))^{N_t - |\mathcal{C}|}} dx \right) dw. \quad (3)$$

where $f_{\sin^2(\theta_{cc})}(x)$ is the mismatch's pdf, assuming QCA [12, 19].

Proof: See Appendix.

Proposition 3: Another approximate of the EC is suggested by $ec_{cc} \approx \frac{-1}{\theta_c} \log(1 - \theta_c \bar{R}_{cc} + \frac{\theta_c^2}{2} \hat{R}_{cc})$, where $\bar{R}_{cc}$ is given in Proposition 2 and $\hat{R}_{cc}$ is obtained by

$$\int_0^\infty \int_0^\infty \frac{e^{-\sigma^2 N_t(w_1 + w_2)}}{\frac{w_1 w_2}{(1 - \hat{K}(w_1, w_2))}} \prod_{c' \in \mathcal{C} - c} \frac{1}{1 + \rho_{cc'} P_{c'} \delta_{cc'}(w_1 + w_2)} dw_1 dw_2$$

where function $\hat{K}(w_1, w_2)$ is given in Appendix.

Proof: See Appendix.

Note that the approximation in Proposition 3 is valid for $0 < 1 - \theta_c \bar{R}_{cc} + \frac{\theta_c^2}{2} \hat{R}_{cc,n} \leq 1$, which can be included as part of the FBP constraint.

Fig. 1-RHS illustrates the EC of a given link. As shown in this case, Propositions 1 and 3 also provide reasonably accurate approximations of the true EC, especially when θ_c is large enough. The accuracy of the derived approximations is lower for smaller θ_c. However, since for sufficiently small θ_c, the EC converges to the capacity, one may simply consider Proposition 2 as an approximation of the EC.

B. A Greedy Algorithm for Feedback Bit Partitioning

Having the EC specified by Proposition 1 or 3, we now aim at solving the optimization problem (2). However, due to complexity of the derived approximations of EC, analytical solution is challenging. Furthermore, exhaustive search is computationally infeasible as B_{tot} and/or $|\mathcal{C}|$ are usually very large. Here, we develop a greedy solution.

Our approach is to divide the greedy search into two layers: the first layer assigns bits to the cells by the developed

algorithm Cluster-level FBP (C-lFBP) in Alg. 1. The second layer partitions the assigned feedback bits to each cell for the quantization of $|\mathcal{C}|$ CDIs at each UE. This is done by conducting UE-level FBP (U-lFBP) in Alg. 2. Note that U-lFBP is previously considered in [20], where the objective was to minimize capacity gap between quantized and accurate scenarios. Furthermore, it is assumed there that per UE feedback capacity is given. Here we extend the algorithm for the cases where feedback capacity per cluster is unspecified, and we consider EC as the performance metric.

Algorithm 1 Cluster-level Feedback Bit Partitioning (C-lFBP)

1: Set $\boldsymbol{b} = [\check{b}_c, \forall c] = \mathbf{0}^{|\mathcal{C}| \times 1}$, $\boldsymbol{B} = \mathbf{0}^{|\mathcal{C}| \times |\mathcal{C}|}$, $\boldsymbol{ec}(\boldsymbol{b}) = \mathbf{0}^{|\mathcal{C}| \times 1}$, and $\check{\boldsymbol{B}}(\boldsymbol{b}) = \mathbf{0}^{|\mathcal{C}| \times |\mathcal{C}|}$
2: **while** $B > 0$ **do**
3:     **for** $c' \in \mathcal{C}$ **do**
4:         Set $\boldsymbol{b}^{(c')} = \{b_1, b_2, \ldots, b_{c'} + 1, \ldots, b_{|\mathcal{C}|}\}$
5:         **for** $\forall c \in C$ **do**
6:             Execute U-lFBA for UE c for $b_c = [\boldsymbol{b}^{(c')}]_c$ and set $[\boldsymbol{ec}(\boldsymbol{b}^{(c')})]_c$ and $[\check{\boldsymbol{B}}^{(c')}(\boldsymbol{b}^{(c')})]_{c,:}$ accordingly.
7:         **end for**
8:     **end for**
9:     $\dot{c} = \max_{c'} \left(\sum_{c \in \mathcal{C}} [\boldsymbol{ec}(\boldsymbol{b}^{(c')})]_c - \sum_{c \in \mathcal{C}} [\boldsymbol{ec}(\boldsymbol{b})]_c \right)$
10:     Update $ec = \sum_{c \in \mathcal{C}} [\boldsymbol{ec}(\boldsymbol{b}^{(\dot{c})})]_c$, $B = B - 1$, $\boldsymbol{B} = \check{\boldsymbol{B}}^{(\dot{c})}(\boldsymbol{b}^{(\dot{c})})$, and $\boldsymbol{b} = \boldsymbol{b}^{(\dot{c})}$
11: **end while**

Algorithm 2 *UE-level Feedback Bit Partitioning* (U-lFBP)

1: Set $B_{cc'} = 0, \forall c'$, $\boldsymbol{b}_{cc'} = \{B_{c1}, \ldots, B_{c|\mathcal{C}|}\}$
2: **while** $b_c > 0$ **do**
3:     **for** $c' \in C$ **do**
4:         Set $\boldsymbol{b}^{(c')} = \{B_{c1}, \ldots, B_{cc'} + 1, \ldots, B_{c|\mathcal{C}|}\}$
5:         Calculate ec_{cc} for $\boldsymbol{b}^{(c')}$ and update $[\boldsymbol{ec}_{cc}(\boldsymbol{b}^{(c')})]_{c'} = ec_{cc}$
6:     **end for**
7:     Find $\dot{c} = \max_{c'} \left([\boldsymbol{ec}_{cc}(\boldsymbol{b}^{(c')})]_{c'} - [\boldsymbol{ec}_{cc}(\boldsymbol{b})]_{c'} \right)$
8:     Update $ec_{cc} = [\boldsymbol{ec}_{cc}(\boldsymbol{b}^{(\dot{c})})]_{\dot{c}}$, $\boldsymbol{b}_{cc} = \boldsymbol{b}^{(\dot{c})}$, $b_c = b_c - 1$
9: **end while**

To implement the proposed greedy algorithm, the CP merely needs to know path-loss attenuations of each UE with respect to all BSs in the cluster, which is practically feasible and does not add much to the signaling overhead. After specifying FBP, BSs will respectively inform the selected UEs with available feedback resources for code-book constructions, and consequently CDI quantization/feedback.

IV. SIMULATION RESULTS

The simulation setup is depicted in Fig. 1-LHS. The cluster under consideration consists of $|\mathcal{C}|$ BSs. In each cluster, BSs are located in a circle centered at the cluster-center, where the PC is located, with radius $0 \leq D \leq R$ where R is the cluster radius. We set R to 500m. The rest of the parameters are $P_c = 10$W $\forall c$, $\sigma^2 = 10^{-14}$W, $\alpha = 3.76$. In both cases, UE c is randomly placed in the respective disk subject to the sector that is covered by BS c. We apply the Monte Carlo technique and average the performance over 10000 snapshots.

We compare our algorithm, referred to as *Sch. 1* , with three other FBP approaches: *Sch. 2* that not only equally partitions B_{tot} among the cells but also equally divides available $B_{tot}/|\mathcal{C}|$ to each UE among the interfering and attending BSs; *Sch. 3* that solves the following optimization for FBP

$$\Xi(B_{cc}^*) = \min \sum_{c' \in \mathcal{C}-c} \rho_{cc'} \delta_{cc'} \quad s.t. \quad \sum_{c' \in \mathcal{C}-c} B_{cc'} = B_c - B_{cc}^*.$$

Here we assume B_{tot} is equally partitioned across cells. Furthermore, $B_{cc}^* = \max_{B_{cc} \in \{0,1,...,B_c\}} \Xi(B_{cc})$. Sch.3 minimizes the residual interference due to interfering BSs at each UE. Note that the above optimization can be solved via a linear search on B_{cc} and Lagrange method for obtaining $B_{cc'}$ followed by simple rounding of the results, the details of which are omitted due to the space limitation. Finally, *Sch. 4* does FBP by solving the optimization problem $\max_B \sum_{c \in \mathcal{C}} \bar{R}_{cc}$ where $\bar{R}_{cc}$ is given by Proposition 2. The above optimization problem is solved via the developed greedy algorithm simply by replacing sum EC with sum average rate. For Sch. 2, Sch. 3, and Sch. 4, we obtain the corresponding FBPs and evaluate the sum ECs.

Fig. 2 illustrates the comparison results for two cases of $B_{tot} = 50$ (left) and $B_{tot} = 100$ (right). Sch. 1 is shown to outperform the other schemes. For the case when B_{tot} is large, there is a diminishing difference between Sch. 1 and Sch. 4. Further, increasing B_{tot} has a positive impact on increasing sum EC. Note that Fig. 2 also shows that by increasing θ, the EC is drastically decreased, indicating that in limited feedback CoMP systems, the system's outage capacity can be very low. As it is seen FBP alone is not able to completely overcome this phenomenon.
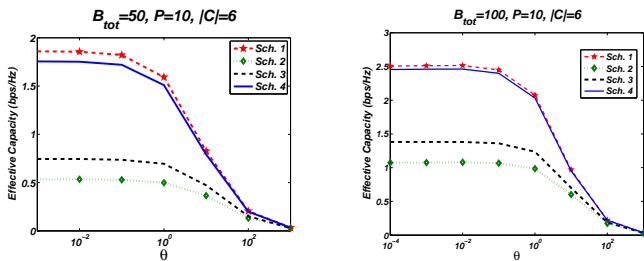

Fig. 2. Sum EC *vs.* delay exponent $\theta_c = \theta \ \forall c$.

One may consider two options to address the issue of low outage capacity, i.e., EC when $\theta >> 0.1$; (1) power control in conjunction with FBP, (2) dynamic clustering in conjunction with FBP, by properly switching off/on BSs, in order to manage the excessive residual interference due to the limited feedback. Thorough investigations of these options are left as our future work, while reporting some preliminary results in Fig. 3. As shown in the left panel of Fig. 3, increasing transmission power can dramatically improve the sum EC. However, very large transmission power is almost ineffective for the EC. This is previously reported for the capacity in [3]. Note that Sch. 2 and 3 could not absorb the benefits of higher

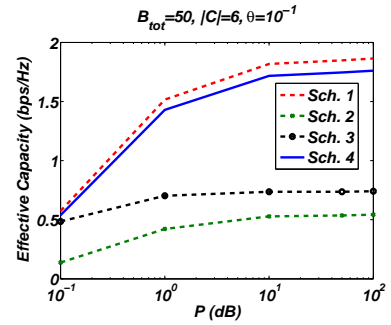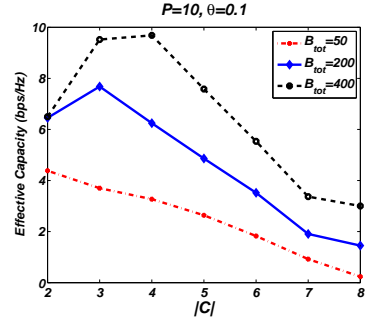Fig. 3. Sum EC *vs.* transmission power.

Fig. 4. Sch. 1's sum EC performance *vs.* cluster-size $|\mathcal{C}|$.

transmission power. Thus, it is important to join FBP with power control for full exploitation of the power control.

On the other hand, Fig. 3-RHS depicts sum EC versus the cluster size. It indicates that for given B_{tot} there is an optimal cluster-size that yields the maximum sum EC. For the case of $B_{tot} = 400$, there is a huge difference between cluster size of 4 and 7. This highlights the importance of clustering in enhancing the network performance.

Finally, Figs. 5 and 6 show the impact of parameter D (see Fig. 1-left) on sum EC, respectively, for $N_t = 8$ and $N_t = 12$. By increasing D, the sum EC steadily decreases. Note that increasing D results in reducing the power of signal and interference strengths, however, since the inter-cell interference is canceled, albeit partially, the negative impact of weak signal strength is superior. Comparing these figures, we further confirm that increasing N_t improves the sum EC.

V. CONCLUSIONS

In this paper we focused on developing a FBP algorithm suitable for provisioning statistical QoS for limited-feedback CoMP systems. We provided several analytical bounds on the EC and examined their accuracy. Further, the efficiency of our proposed greedy FBP algorithm against some existing approaches was confirmed. The main takeaways of this paper were that (*i*) it is very important to optimally divide the feedback capacity across cells and UEs, (*ii*) FBP is unable to meet the required stringent delay requirements, and it should be joined with other techniques such as power control and dynamic clustering, (*iii*) power control is highly effective in conjunction with optimal FBP; otherwise the performance boost is very limited, and (*iv*) for given feedback capacity
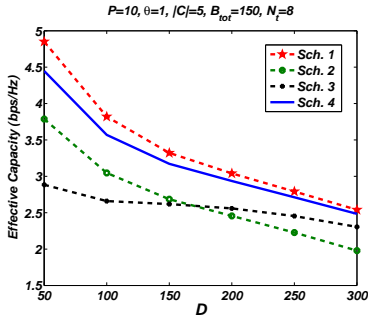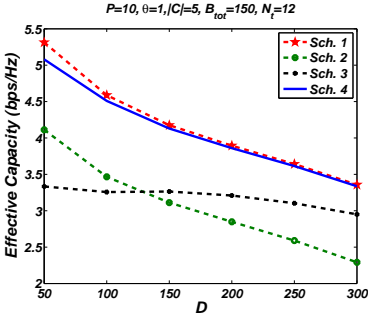
Fig. 5. Sum EC *vs.* distance D for $N_t = 8$.

Fig. 6. Sum EC *vs.* distance D for $N_t = 12$.

constraint per cluster, there is an optimal cluster size which renders the maximum EC.

APPENDIX

1) Proof of Proposition 1: Let random variable $J_{cc'}$ denote $J_{cc'} = \|\boldsymbol{h}_{cc'}\|^2 |\tilde{\boldsymbol{h}}_{cc'}^\dagger \hat{\boldsymbol{f}}_{c'}|^2$ and K_{cc} be $K_{cc} = \|\boldsymbol{h}_{cc}\|^2 |\tilde{\boldsymbol{h}}_{cc}^\dagger \hat{\boldsymbol{f}}_c|^2$. Decomposing vector $\tilde{\boldsymbol{h}}_{cc'}$ as $\tilde{\boldsymbol{h}}_{cc'} = \cos(\theta_{cc'})\hat{\boldsymbol{h}}_{cc'} + \sin(\theta_{cc'})\boldsymbol{v}_{cc'}$, where $\boldsymbol{v}_{cc'}$ is perpendicular to $\hat{\boldsymbol{h}}_{cc'}$ and $\theta_{cc'}$ is the angle between $\tilde{\boldsymbol{h}}_{cc'}$ and $\hat{\boldsymbol{h}}_{cc'}$, and adopting the developed theory of QCA quantization from [4, 12], $J_{cc'}$ is distributed as $J_{cc'} \propto \|\boldsymbol{h}_{cc'}\|^2 \sin^2(\theta_{cc'})B(1, N_t - 2)$ in which $B(1, N_t - 2)$ indicates a beta random variable with parameters 1 and $N_t - 2$. Utilizing the results of [21, 22] $J_{cc'}$ is an exponentially distributed random variable where parameter $\delta_{cc'}$. Note that random variables K_{cc} and $J_{cc'}$ $\forall c' \neq c$ are independent since the direct and interfering channel vectors are independent as well as the respective quantization code-books are constructed separately. We expand $\tilde{\boldsymbol{h}}_{cc}$ as $\tilde{\boldsymbol{h}}_{cc} = \cos(\theta_{cc})\hat{\boldsymbol{h}}_{cc} + \sin(\theta_{cc})\boldsymbol{s}_{cc}$, where $\boldsymbol{s}_{cc}$ is orthogonal to $\hat{\boldsymbol{h}}_{cc}$ and θ_{cc} is the angle between $\tilde{\boldsymbol{h}}_{cc}$ and $\hat{\boldsymbol{h}}_{cc}$. We then apply the following approximation:

$$K_{cc} \approx \cos^2(\theta_{cc})\|\boldsymbol{h}_{cc}\|^2 |\hat{\boldsymbol{h}}_{cc}^\dagger \hat{\boldsymbol{f}}_c|^2 + \sin^2(\theta_{cc})\|\boldsymbol{h}_{cc}\|^2 |\boldsymbol{s}_{cc}^\dagger \hat{\boldsymbol{f}}_c|^2. \tag{4}$$

As it is shown in [14] $\|\boldsymbol{h}_{cc}\|^2 |\hat{\boldsymbol{h}}_{cc}^\dagger \hat{\boldsymbol{f}}_c|^2 \sim \chi^2_{2(N_t - |\mathcal{C}| + 1)}$, which is a Chi-squared random variable with $2N_t - 2|\mathcal{C}| + 2$ degree-of-freedoms. Accordingly, we approximate the received signal strength by $\rho_{cc}P_c\|\boldsymbol{h}_{cc}\|^2 |\hat{\boldsymbol{h}}_{cc}^\dagger \hat{\boldsymbol{f}}_c|^2 \approx \tilde{\rho}_{cc}\chi^2_{2(N_t - |\mathcal{C}| + 1)}$. where $\tilde{\rho}_{cc} = \rho_{cc}P_c\left(1 - \frac{N_t - 1}{N_t}\delta_{cc}\right)$. We define random variable $Y = \sum_{c' \neq c} \rho_{cc'}P_{c'}J_{cc'}$, and introduce a new random variable defined as $Z = \tilde{\rho}_{cc}\chi^2_{2(N_t - |\mathcal{C}| + 1)}/(\sigma^2 N_t + Y)$. Let start by deriving an expression for the pdf of random variable Z as

follows:

$$f_Z(z) = \frac{\partial}{\partial z}\mathbb{P}\left\{\frac{\chi^2_{2(N_t - |\mathcal{C}| + 1)}}{\sigma^2 N_t + Y} \leq \frac{z}{\tilde{\rho}_{cc}}\right\}$$

$$= \mathbb{E}_Y\left[\frac{\sigma^2 N_t + Y}{\tilde{\rho}_{cc}}f_{\chi^2_{2(N_t - |\mathcal{C}| + 1)}}\left(z\frac{\sigma^2 N_t + Y}{\tilde{\rho}_{cc}}\right)\right]$$

$$= \mathbb{E}_Y\left[\left(\frac{\sigma^2 N_t + Y}{\tilde{\rho}_{cc}}\right)^{N_t - |\mathcal{C}| + 1}\frac{z^{N_t - |\mathcal{C}|}e^{-z\frac{\sigma^2 N_t + Y}{\tilde{\rho}_{cc}}}}{\Gamma(N_t - |\mathcal{C}| + 1)}\right]$$

$$= \frac{z^{N_t - |\mathcal{C}|}}{\Gamma(N_t - |\mathcal{C}| + 1)}\mathbb{E}_Y\left[(-1)^{N_t - |\mathcal{C}| + 1}\frac{\partial^{N_t - |\mathcal{C}| + 1}}{\partial z^{N_t - |\mathcal{C}| + 1}}e^{-z\frac{\sigma^2 N_t + Y}{\tilde{\rho}_{cc}}}\right],$$

where we substituted the pdf of random variable $\chi^2_{2(N_t - |\mathcal{C}| + 1)}$ and applied some straightforward manipulations. Consequently,

$$f_Z(z) = \frac{(-1)^{N_t - |\mathcal{C}| + 1}z^{N_t - |\mathcal{C}| + 1}}{\Gamma(N_t - |\mathcal{C}| + 1)}\frac{\partial^{N_t - |\mathcal{C}| + 1}}{\partial z^{N_t - |\mathcal{C}| + 1}}\mathbb{E}_Y\left[e^{-z\frac{\sigma^2 N_t + Y}{\tilde{\rho}_{cc}}}\right],$$

which, by recalling the definition of random variable Y can be written as

$$f_Z(z) = (-1)^{N_t - |\mathcal{C}| + 1}\frac{z^{N_t - |\mathcal{C}|}}{\Gamma(N_t - |\mathcal{C}| + 1)}\frac{\partial^{N_t - |\mathcal{C}| + 1}O(z)}{\partial z^{N_t - |\mathcal{C}| + 1}}, \tag{5}$$

where $O(z)$ is defined as

$$O(z) = e^{-z\frac{\sigma^2 N_t}{\tilde{\rho}_{cc}}}\prod_{c' \in \mathcal{C} - c}\frac{1}{1 + z\frac{\rho_{cc'}P_{c'}\delta_{cc'}}{\tilde{\rho}_{cc}}}. \tag{6}$$

Utilizing pdf (5), we get the following expression:

$$\mathbb{E}e^{-\theta_c \log(1 + Z)} = \frac{(-1)^{N_t - |\mathcal{C}| + 1}}{\Gamma(N_t - |\mathcal{C}| + 1)}\int_0^\infty \frac{z^{N_t - |\mathcal{C}| + 1}}{(1 + z)^{\theta_c}}\frac{\partial^{N_t - |\mathcal{C}| + 1}O(z)}{\partial z^{N_t - |\mathcal{C}| + 1}}dz.$$

Using this, we derive an approximate of the EC as suggested in Proposition 1.

2) Proof of Proposition 2: Referring to the proof of Proposition 1 we have $\bar{R}_{cc} = \mathbb{E}\log(1 + \frac{\rho_{cc}P_cK_{cc}}{\sigma^2 N_t + Y})$. According to the rate-splitting equality [23], it is observed that

$$\bar{R}_{cc} = \mathbb{E}\log\left(1 + \frac{\rho_{cc}P_cK_{cc} + Y}{\sigma^2 N_t}\right) - \mathbb{E}\log\left(1 + \frac{Y}{\sigma^2 N_t}\right). \tag{7}$$

Substituting the following identity $\log(1 + x) = \int_0^\infty \frac{e^{-w}}{w}(1 - e^{-wx})dw$ [24], into (7), and then applying some manipulations, we get

$$\bar{R}_{cc} = \int_0^\infty \frac{e^{-N_t w\sigma^2}}{w}\mathbb{E}e^{-wY}\left(1 - \mathbb{E}e^{-wP_c\rho_{cc}K_{cc}}\right)dw.$$

$$= \int_0^\infty \frac{e^{-N_t w\sigma^2}}{w}\prod_{c' \in \mathcal{C} - c}\frac{1}{1 + \rho_{cc'}P_{c'}\delta_{cc'}w}\left(1 - \mathbb{E}e^{-wP_c\rho_{cc}K_{cc}}\right)dw. \tag{8}$$

where in the last step we used the fact that $J_{cc'}$ is an exponential random variable with parameter $\delta_{cc'}$. Now, we calculate $\mathbb{E}e^{-wP_c\rho_{cc}K_{cc}}$ for fixed w. Random variable $|\boldsymbol{s}_{cc}^\dagger \hat{\boldsymbol{f}}_c|^2$ is beta with parameters 1 and $N_t - 2$ [4]. Besides, random

variable $\sin^2(\theta_{cc})\|\boldsymbol{h}_{cc}\|^2|\boldsymbol{s}_{cc}^\dagger\hat{\boldsymbol{f}}_c|^2$ is exponentially distributed with parameter δ_{cc}. Thus,

$$\mathbb{E}e^{-wP_c\rho_{cc}K_{cc}} \approx \mathbb{E}e^{-wP_c\rho_{cc}\cos^2(\theta_{cc})\chi^2_{2(N_t-|\mathcal{C}|+1)}}\mathbb{E}e^{-wP_c\rho_{cc}\delta_{cc}\chi^2_2}, \tag{9}$$

in which for mathematical tractability we have assumed that random variable $\cos^2(\theta_{cc})\chi^2_{2(N_t-|\mathcal{C}|+1)}$ and $\delta_{cc}\chi^2_2$ are independent, which clearly are not. Combining (9) and (8), the following approximation on the achievable capacity can be suggested

$$\bar{R}_{cc} \approx \int_0^\infty \frac{e^{-N_t w\sigma^2}}{w}\prod_{c'\in\mathcal{C}-c}\frac{1}{1+\rho_{cc'}P_{c'}\delta_{cc'}w}\times$$

$$\left(1 - \frac{\mathbb{E}e^{-wP_c\rho_{cc}\cos^2(\theta_{cc})\chi^2_{2(N_t-|\mathcal{C}|+1)}}}{1+\rho_{cc}P_c\delta_{cc}w}\right)dw. \tag{10}$$

What remains is to substitute the pdf of QCA mismatch in (10) that yields the desired result.

3) Proof of Proposition 3: It is straightforward to show that

$$\mathbb{E}e^{-\theta_c R_{cc}} = 1 - \theta_c\bar{R}_{cc} + \sum_{k=2}^\infty \frac{(-\theta_c)^k}{k!}\mathbb{E}\left(\log(1+\gamma_{cc})\right)^k, \tag{11}$$

where $\bar{R}_{cc}$ is already known from Proposition 2. Recalling the definitions of the random variables J_{cc} and K_{cc} from the first part of the Appendix, we get

$$\mathbb{E}R_{cc}^k = \mathbb{E}\left(\int_0^\infty \frac{e^{-w}}{w}(1-e^{-w\gamma_{cc}})dw\right)^k$$

$$= \mathbb{E}\int_0^\infty\cdots\int_0^\infty\prod_{i=1}^k\frac{e^{-w_i}}{w_i}(1-e^{-w_i\gamma_{cc}})dw_i = \int_0^\infty\cdots\int_0^\infty\left(\prod_{i=1}^k\frac{e^{-\sigma^2 N_t w_i}}{w_i}\right)$$

$$\times\mathbb{E}e^{-Y\sum_{i=1}^k w_i}\mathbb{E}\prod_{i=1}^k\left(1-e^{-\rho_{cc}P_c K_{cc}w_i}\right)dw_1\ldots dw_k$$

$$= \int_0^\infty\cdots\int_0^\infty\frac{e^{-\sigma^2 N_t\sum_{i=1}^k w_i}}{\prod_{i=1}^k w_i}\prod_{c'\in\mathcal{C}-c}\frac{1}{1+\rho_{cc'}P_{c'}\delta_{cc'}\sum_{i=1}^k w_i}$$

$$\times\mathbb{E}\prod_{i=1}^k\left(1-e^{-\rho_{cc}P_c K_{cc}w_i}\right)dw_1\ldots dw_k$$

For the general case, it is too complicated to derive a closed-form expression for this integral. We therefore assume $k = 2$, and denote $\hat{R}_{cc} = \mathbb{E}\left(\log(1+\gamma_{cc,n})\right)^2$. As a result, $\hat{R}_{cc}$ is obtained as

$$\int_0^\infty\int_0^\infty\frac{e^{-\sigma^2 N_t(w_1+w_2)}}{\frac{w_1 w_2}{(1-\hat{K}(w_1,w_2))}}\prod_{c'\in\mathcal{C}-c}\frac{1}{1+\tilde{\rho}_{cc'}P_{c'}\delta_{cc'}(w_1+w_2)}dw_1 dw_2. \tag{12}$$

where, by following the lines presented in the proof of Proposition 2, it is straightforward to confirm that $\hat{K}(w_1,w_2)$ is obtained as

$$\int_0^{\delta_{cc}}\left[\frac{\frac{1}{1+w_1 P_c\rho_{cc}\delta_{cc}}}{(1+\frac{w_1 P_c\rho_{cc}}{(1-x)^{-1}})^{N_t-|\mathcal{C}|+1}} + \frac{\frac{1}{1+w_2 P_c\rho_{cc}\delta_{cc}}}{(1+w_2 P_c\rho_{cc}(1-x))^{N_t-|\mathcal{C}|+1}}\right.$$

$$\left. - \frac{\frac{1}{1+(w_1+w_2)P_c\rho_{xc}\delta_{cc}}}{(1+(w_1+w_2)P_c\rho_{cc}(1-x))^{N_t-|\mathcal{C}|+1}}\right]\mathrm{f}_{\sin^2(\theta_{cc})}(x)dx.$$

Substituting (12) into (11), the proposed approximation in Proposition 3 is finally obtained.

REFERENCES

[1] D. Gesbert, *et al.*, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE JSAC*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[2] J. G. Andrews *et al.*, "What will 5G be?" *IEEE JSAC*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sep. 2013.

[4] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[5] G. Caire *et al.*, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[6] H. Dahrouj and W. Yu, "Coordinated beamforming for the multi-cell multi-antenna wireless system," *IEEE Trans. Wireless Comm.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[7] X. Zhang *et al.*, "NEMOx: Scalable network MIMO for wireless networks," in *Proc. of MobiCom'13*, Sep. 2013, pp. 453–464.

[8] S. Kaviani and W. A. Krzymien, "Multicell scheduling in network MIMO," *IEEE GLOBECOM'10*, pp. 1 – 5, Dec. 2010.

[9] S. He *et al.*, "Coordinated multicell multiuser precoding for maximizing weighted sum energy efficiency," *IEEE Trans. Sig. Proc.*, vol. 62, no. 3, pp. 741–751, Feb. 2014.

[10] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization through user association and beamforming in C-RAN," *IEEE Trans. Wireless Comm.*, vol. 14, no. 1, pp. 494–508, Jan. 2015.

[11] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Trans. Wireless Comm.*, vol. 11, no. 10, pp. 3618–3631, Oct. 2012.

[12] R. Bhagavatula and R. W. Heath, "Adaptive bit partitioning for multicell intercell interference nulling with delayed limited feedback," *IEEE Trans. Sig. Proc.*, vol. 59, no. 8, pp. 3824–3836, Aug. 2011.

[13] S. Yu *et al.*, "Novel feedback bit allocation methods for multi-cell joint processing systems," *IEEE Trans. Wireless Comm.*, vol. 11, no. 9, pp. 3030–3036, Sep. 2012.

[14] J. Zhang and J. G. Andrews, "Adaptive spatial intercell interference cancellation in multicell wireless networks," *IEEE JSAC*, vol. 28, no. 9, pp. 1455–1468, Dec. 2010.

[15] P. Kerret and D. Gesbert, "Degrees of freedom of the network MIMO channel with distributed sci," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 6806–6824, Nov. 2012.

[16] Q. Du and X. Zhang, "Base-station selections for QoS provisioning over distributed multi-user MIMO links in wireless networks," *IEEE INFOCOM*, pp. 3038–3046, Apr. 2011.

[17] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Trans. Wireless Comm.*, vol. 2, no. 4, pp. 630–643, Jul. 2003.

[18] J. Tang and X. Zhang, "Cross-layer resource allocation over wireless relay networks for quality of service provisioning," *IEEE JSAC*, vol. 25, no. 4, pp. 645–656, May 2007.

[19] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE JSAC*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.

[20] X. Chen and C. Yuen, "Performance analysis and optimization for interference alignment over MIMO interference channels with limited feedback," *IEEE Trans. Sig. Proc.*, vol. 62, no. 7, pp. 1785–1795, Apr. 2014.

[21] J. Zhang *et al.*, "Multi-mode transmission for the MIMO broadcast channel with imperfect channel state information," *IEEE Trans. Comm.*, vol. 59, no. 3, pp. 803–814, Mar. 2011.

[22] M. Kountouris and J. G. Andrews, "Downlink SDMA with limited feedback in interference-limited wireless networks," *IEEE Trans. Wireless Comm.*, vol. 11, no. 8, pp. 2730–2741, Aug. 2012.

[23] N. Jindal and A. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates," *IEEE Trans. Inf. Theory*, vol. 49, no. 11, pp. 2895–2909, Nov. 2003.

[24] K. Hamdi, "Capacity of mrc on correlated rician fading channels," *IEEE Trans. Comm.*, vol. 56, no. 5, pp. 708–711, May 2008.