Small-Macro Cell Cooperation for HetNet Uplink Transmission: Spectral Efficiency and Reliability Analyses

Ahmad Abu Al Haija, *Member, IEEE*, and Chintha Tellambura, *Fellow, IEEE*

***Abstract*— We investigate the impact of small-macro cell cooperation (SMC) in improving the spectral efficiency and reliability of uplink transmission in a heterogeneous network. We consider a network of two user equipments (UEs), a macro-cell base station (BS) and a small-cell BS. Joint SMC involves macro-to-small quantized feedback and decode-forward relaying from small to macro cell. This cooperation utilizes full-duplex transmission and intra-network spectrum sharing. We first propose a transmission scheme based on superposition block Markov encoding at each UE, coherent decode-forward relaying and sliding window decoding at the small-cell BS, and quantize-forward relaying and backward decoding at the macro-cell BS. Second, we derive the optimal macro-cell quantization to maximize the whole spectral efficiency. Third, for a certain non-fading scenario, we prove that the proposed scheme asymptotically achieves the capacity (maximum spectral efficiency) by reaching the cut-set bound as macro-cell power approaches infinity. Fourth, we formulate the outage probability over block fading channels, considering the outage events at the small and macro-cell BSs and the channel variations over different blocks. Last, we generalize the proposed scheme to an N (> 2)-UE case. As macro-cell power increases, the results show that the proposed scheme achieves a full diversity order of two and outperforms all existing non-SMC schemes. These strong results suggest the utility of the proposed scheme for potential deployment in 5G cellular networks.**

***Index Terms*— Superposition block Markov encoding, decode-forward (DF) and quantize-forward (QF) relaying, backward and sliding window decoding, capacity analysis, block fading channels, outage analysis.**

I. Introduction

ULTRA-HIGH spectral efficiency and reliability ($> 99\%$) are two important requirements for 5G (fifth generation) cellular standards [1], [2]. Hence, to improve the network throughput and reliability, the concept of a heterogeneous network (HetNet) with wireless backhaul, full-duplex transmission and spectrum sharing confers several benefits. First, coverage and high-speed mobility are enhanced by combining small and macro cells. Second, unlike conventional half-duplex (HD) mode, full-duplex (FD) allows simultaneous transmission and reception on the same frequency slot, which significantly improves spectral efficiency. Last, spectrum sharing allows efficient flexible use of the spectrum based on throughput and reliability requirements. To meet a specific requirement, different operators (inter-spectrum sharing) can share their licensed spectrum [2], or different nodes in a single operator (intra-spectrum sharing) can share their slotted bands (resource blocks) [3]. Clearly, an optimal combination of these technologies could improve the spectral efficiency and network reliability sufficient to meet the 5G requirements.

Manuscript received May 7, 2016; revised August 13, 2016 and October 20, 2016; accepted November 6, 2016. Date of publication November 24, 2016; date of current version January 12, 2017. This work was supported by the Natural Science and Engineering Research Council of Canada. This work was presented at the IEEE International Conference on Communications 2016.

The authors are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada (e-mail: sabualhai@ualberta.edu; ct4@ualberta.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2016.2633039

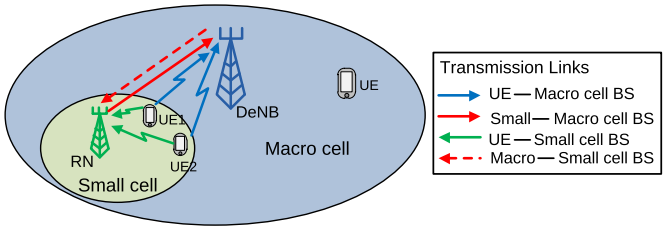

Fig. 1. HetNet uplink transmission.

A. Problem Statement

Consider the HetNet uplink transmission (Fig. 1) of two random UEs in a small-cell with transmission rates R_1 and R_2. The two UEs communicate with some other entities somewhere outside the small-cell. The UE signals are received by both the small and macro cells. In addition, the small-cell provides a relaying function utilizing the quantized feedback from the macro-cell (see Section I.C for more details). The goal is to improve their uplink spectral efficiency, i.e., enlarging the rate region (R_1, R_2). This channel is theoretically defined as a multiple access relay channel (MARC) [4] where the two UEs resemble the sources, the small-cell resembles the relay ($\mathcal{R}$) and the macro-cell resembles the destination ($\mathcal{D}$). In this paper, we design a communication scheme for this channel with high spectral efficiency and reliability. To reach these improvements, the scheme combines the following techniques.

1) Densification and randomness of small cells: In a 5G network, the small-cell location is random (not necessarily on the macro-cell boundary). Hence, when the small-cell is close to the macro-cell, the link qualities from UEs to the small and macro cells are roughly similar. Therefore, the macro BS receives UEs' signals not only through the small-cell relaying as in dual-hop transmission [5] but also directly through both UEs as in MARC [4].

2) Wireless backhaul: Small and macro-cells are connected through wired or wireless backhaul networks. While wireless backhaul has been standardized in LTE-A release 10 [5], 5G wide deployment is expected due to the massive densification of small-cells that makes wireless backhaul much more cost-effective than wired backhaul [1], [2]. Thus, we can exploit the wireless backhaul channel to enable small-cell and macro-cell cooperation.

3) Full duplex radio: The spectral efficiency can potentially be doubled when the small-cell deploys FD relaying, which will be deployed in 5G cellular networks [2]. The main drawback of FD is self-interference, which is the interference a radio node experiences from its own transmission to the received signal. Fortunately, much progress has been made in self-interference suppression by using millimeter waves [1] and cancellation techniques [6]–[9], including passive suppression, and analog and digital cancellation. These techniques can reduce the self-interference by 110 dB in WiFi radios [9]. Therefore, FD transmission must be exploited to improve the network performance in Fig. 1.

4) Macro-cell feedback: Since the macro BS has more capabilities and much more transmission power ($(20-40)$W) than the small-cell BS ($(0.02-2)$W) and UEs ($(16-50)$mW) [5], feedback from macro to small-cell improves the spectral efficiency and reliability. This small-macro cell cooperation (SMC) is achieved over the same band using full-duplex spectrum sharing.

B. Related Work

Several studies have utilized spectrum sharing to improve the coverage probability and energy efficiency of HetNets, while others have analyzed the different transmission schemes of HetNets and MARC and studied their spectral efficiency and reliability performances.

1) Spectrum Sharing: Intra-operator spectrum sharing techniques can improve HetNet performance [10]–[14]. For the downlink, spectrum sharing helps to manage interference and improve coverage and spectral efficiency though feedback from macro to small-cell [10], [11] or coordinated and cooperative transmissions among small and macro-cells [13]–[18]. For the uplink, compared to resource partitioning, spectrum sharing improves the average network utility per UE in the uplink transmission despite interference from other UEs in the network [12]. However, [12] considers no SMC for the data signals. Thus, joint processing among small and macro cells suggests itself for data signals in uplink transmission.

2) Transmission Schemes and Spectral Efficiency: Using the basic relaying techniques in [19], several studies propose different coding schemes for classical MARC (without relay-destination cooperation (RDC)), including full and partial decode-forward (DF) relaying [4], [20]–[22] and compress and quantize-forward (QF) relaying [23]–[27]. These studies derive the spectral efficiency [4], [20]–[22] and the optimal quantization for the QF relaying schemes [26], [27]. Reference [28] derives the sum rate capacity for the degraded Gaussian MARC.

When $\mathcal{D}$ cooperates with $\mathcal{R}$ or the sources, the cooperation can improve spectral efficiency for cooperative multiple access channel (MAC) [29], interference channel [30] and relay channel [31], [32]. It can even achieve the capacity asymptotically for the relay channel with RDC as $\mathcal{D}$ power approaches infinity [32]. This fact motivates us to investigate the impact of RDC (e.g. SMC) in improving the spectral efficiency of the HetNet uplink channel.

The theoretical studies in [4] and [20]–[22] apply directly to HetNets with FD relaying. By further considering spectrum sharing for SMC, [32] shows the capacity achievement for a single UE only. Hence, for multiple UEs, it is of interest to investigate the spectral efficiency and reliability of HetNet with full-duplex relaying and spectrum sharing.

3) Reliability: An outage of the uplink HetNets was considered for DF [33], [34], and QF [24] relaying. While [24] and [34] consider HD transmission, FD improves performance and will be standardized in 5G networks [2]. The DF scheme in [33] considers FD transmission but with wired backhaul of capacity C between small and macro-cells. However, wireless backhaul must be considered because it will be widely used in 5G networks [2]. Moreover, the small-cell in [33] only relays the successfully decoded message(s), since this relaying strategy achieves a full-diversity order [35] but requires UEs to know the instantaneous relaying link amplitudes. Knowledge of only the link order between direct and relaying links at UEs can also help achieve a full diversity order of the relay channel [36]–[38]. Overall, this paper aims to analyze the outage of uplink HetNet transmission with SMC achieved through wireless backhaul.

C. Main Results and Contributions

This paper comprehensively investigates the uplink transmission of a HetNet with SMC (Fig. 1) and extends our initial results in [39]. We propose a new transmission scheme and derive its spectral efficiency and reliability along with the optimal macro-cell signaling; we also show its asymptotic capacity achievement at high macro-cell power. Our analyses for the basic HetNet form the basis for network-wide scenarios with multiple small and macro-cells.

1) Transmission scheme: : We propose a transmission scheme that is carried over $B \gg 1$ blocks where each UE performs superposition block Markov encoding; the small-cell employs DF relaying and joint sliding window decoding over two blocks, and the macro-cell employs quantize-forward (QF) relaying and backward decoding. While the small-cell (e.g. $\mathcal{R}$) in [4] and [20] decodes the information from both UEs directly at the end of each block, the small-cell in the proposed scheme waits one more block before decoding UEs' information in order to receive the quantized signal from the macro-cell. Hence, transmission in block $k \in \{1, 2, ..., B\}$ in the proposed scheme depends on the information transmitted in block $k-2$.

2) Spectral efficiency: : We derive the spectral efficiency by determining the rate constraints that ensure reliable decoding at small and macro cells. Results show the impact of SMC in improving the spectral efficiency compared with non-SMC

DF relaying schemes [4], [20]. The improvement increases as the macro-cell power increases and as the quality of each UE-small cell link gets closer to that of its corresponding UE-macro cell link.

3) Optimal macro-cell quantization: : Due to its importance from a practical implementation perspective, we derive the optimal quantization at the macro-cell for spectral efficiency maximization. Considering the small-cell rate constraints, we determine the optimal quantization that maximizes each individual UE rate and their sum rate (throughput). These three quantizations lead to three small-cell spectral efficiency regions. Next, by considering their time-shared spectral efficiency region [40], we determine the quantization that leads to a region outside it. Hence, we derive all possible quantization ranges at the macro-cell that maximize the whole spectral efficiency.

4) Asymptotic capacity achievement: : We further analyze the asymptotic performance of the proposed scheme when the macro-cell power approaches infinity. We prove that this asymptotic region is the capacity since it matches the cut-set bound when the link amplitude ratio from one UE to macro and small cells is equal to that from the other UE. As for the relay channel with RDC [32], this result agrees with the intuition that when the macro-cell power approaches infinity, the macro-cell virtually joins the small-cell in one entity and the channel becomes similar to a MAC that has a known capacity [40]. For non-asymptotic regimes, we provide comparisons between the proposed and NNC [41] schemes for HetNet with SMC, the DF relaying scheme for MARC without RDC (e.g. SMC) [4], [20] and the cut-set bound. Results show that the proposed scheme outperforms all existing schemes as the macro-cell power increases.

5) Reliability: : We formulate the outage probability of the proposed scheme over block fading channels assuming full CSI at receivers (CSIR) (e.g., small and macro cells) and limited CSI at transmitters (CSIT) (e.g., UEs and small-cell). More specifically, for any transmission block, each of the small and macro cells knows the phases and amplitudes of its receiving links, which is possible through channel estimation. As transmitters, to perform coherent transmission, UEs and the small-cell each knows the phase of its link to the macro-cell, which is a standard assumption in coherent relaying [19], [40] and can be obtained via feedback from the macro-cell [42]. Furthermore, the macro-cell knows the amplitude of its feedback link to the small-cell. This additional information allows the macro-cell to perform optimal quantization at each block.

We analyze the outage probability considering the outage events at the small and macro-cells. We also consider channel variation over different blocks since each UE message sent in block k, the small-cell decodes it using its received signals in blocks k and $k+1$ while the macro-cell decodes it in block $k+2$. Results show that SMC significantly improves outage performance. Moreover, as the macro-cell power increases, the proposed scheme achieves a full diversity order of two without requiring UEs to have full CSI of their links [35] or even the link orders [36]–[38] through feedback from $\mathcal{D}$. This reduction of the feedback non-data signals satisfies the 5G ultra lean design requirement [2].

6) Generalization to N-UE: : While the previous results explain in detail the analysis of a case with two UEs, we further generalize these results to an N-UE ($N > 2$) case in Section VII.

Controlling the spectral efficiency and reliability of HetNet uplink transmission through the macro-cell has many advantages. First, while achieving the same spectral efficiency, UEs can transmit with lower power, which increases their battery life and reduces their interference to other UEs. Second, the macro-cell can control the interference of its feedback signal on other UEs in the same macro-cell. More specifically, since the macro-cell knows the quantization index and other UEs' messages, it can encode the quantization index against the other UEs' messages using dirty paper coding (DPC) [43] to remove interference from the quantization index signal on other UEs' signals as shown in [40, Ch. 8] for a broadcast channel. Last, the macro-cell can control the interference on other macro-cells' UEs through multi-cell processing techniques like interference coordination or cooperative transmission [44], as the macro-cells can exchange their information through the backhaul network connecting them.

D. Paper Organization

The remainder of this paper is organized as follows. Section II presents the HetNet uplink channel model with SMC. Section III describes the proposed scheme for HetNet uplink transmission with SMC and derives its spectral efficiency region. Section IV derives the optimal macro-cell quantization that is sufficient to achieve the whole spectral efficiency region. Section V investigates the spectral efficiency as the macro-cell power increases, and then proves that the proposed scheme can asymptotically archive capacity by reaching the cut-set bound as the macro-cell power approaches infinity. Section VI derives both individual and common outage probabilities, taking into account outages at the small and macro cells and the channel variation over different blocks. Section VII generalizes the proposed scheme to the N-UE scenario. Section VIII presents numerical results and Section IX concludes the paper.

II. Channel Model

The HetNet uplink channel with SMC consists of two UEs (UE1 and UE2) communicating with a macro-cell with the help of a small-cell where the small and macro cells cooperate to improve the spectral efficiency of both UEs. To further improve spectral efficiency, the nodes share their spectrum instead of partitioning it among them, and both small and macro cells work in FD mode. Fig. 2 is equivalent to Fig. 1 and shows the channel model for FD uplink HetNet with SMC. For any transmission over B blocks, the discrete-time channel

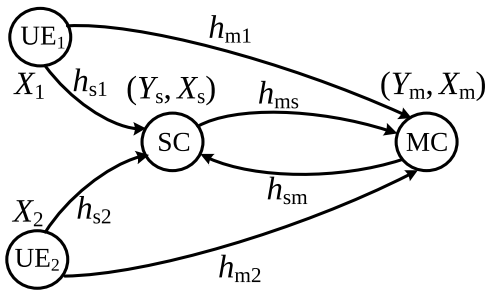

Fig. 2. Channel model for HetNet uplink transmission with SMC (SC: small-cell, MC: macro-cell).

model at block $k \in \{1, 2, ..., B\}$ is given as follows:

$$Y_{s,k} = h_{s1,k}X_{1,k} + h_{s2,k}X_{2,k} + h_{sm,k}X_{m,k} + Z_{s,k},$$
$$Y_{m,k} = h_{m1,k}X_{1,k} + h_{m2,k}X_{2,k} + h_{ms,k}X_{s,k} + Z_{m,k}, \quad (1)$$

where for node $j \in \{1, 2, s, m\}$ (UE1, UE2, small or macro-cell) and node $i \in \{s, m\}$, h_{ij} is the complex link coefficient from node j to i; $X_{j,k}$ is the signal transmitted from node j with transmission power P_j in block k; $Y_{i,k}$ is the signal received at node i in block k; $Z_{s,k}$ and $Z_{m,k} \sim CN(0, 1)$ are independent complex AWGN. Each link coefficient has a complex value $h_{ij,k} = g_{ij,k}e^{\sqrt{-1}\theta_{ij}}$ where $g_{ij,k}$ is the real amplitude gain and θ_{ij} is the phase. In AWGN, the channel links are the same over all transmission blocks ($g_{ij,k} = g_{ij}$). However, over block fading channels, the links remain constant in each transmission block and change independently in the next block (see Section VI-A). Using the standard assumption of coherent relaying [19], [40], we assume that UE1, UE2 and the small-cell each knows the phase of its respective link to the macro-cell and they can transmit coherently. The CSI (phase knowledge) at transmitters can be obtained via feedback from the macro-cell [42]. We further assume that the full channel coefficients are known at the respective receivers (small and macro-cells).

Although the FD transmissions suffer from self-interference, it is alleviated with millimeter waves [1] and analog and digital cancellation techniques [9], which demonstrate a 110 dB reduction of self-interference in WiFi radios with a transmission power of 100mW. With higher transmission power, more reduction in self-interference is expected. However, hardware imperfections leave a small residual self-interference signal, which is seen as additive noise by the small-cell [45] and can be easily incorporated into an outage analysis by increasing the received noise power [37]. In addition to self-interference, the small-cell receives other interference from the macro-cell transmission to other small-cells and macro UEs. The macro-cell can alleviate this interference at the small-cell by using precoding and dirty paper coding techniques [43], [46]. Both self and external interference are important topics that could be considered in the future work.

III. Transmission Scheme for HetNet Uplink Transmission With SMC

The proposed scheme utilizes the macro-cell feedback to the small-cell to increase the information rates (R_1 for UE1 and R_2 for UE2). Thus, the small-cell can decode more information which can be sent to the macro-cell, which enlarges the overall spectral efficiency. Similar to the scheme of relay channel with RDC [32], the transmission scheme is carried over B independent blocks ($B >> 1$, $B \in \mathbb{N}$) where each UE aims to send $B - 2$ messages through B blocks.[1] The scheme is based on superposition block Markov encoding at the UEs; DF relaying and sliding window decoding at the small-cell; and QF relaying and backward decoding at the macro-cell.

A. Transmission Scheme

In a transmission block $k \in \{1, 2, ..., B\}$, UE1 transmits its new and old information $(w_{1,k}, w_{1,k-2})$[2] using superposition block Markov encoding [40, Ch. 16].[3] UE1 first generates a codeword U_1 for $w_{1,k-2}$ and then superposes $w_{1,k}$ over U_1 and generates the codeword X_1. UE2 performs similar encoding and generates the codewords U_2 and X_2. The small-cell forwards U_1 and U_2 to the macro-cell, since it has already decoded $w_{1,k-2}$ and $w_{2,k-2}$ using the received signals in blocks $k-2$ and $k-1$. The macro-cell has already quantized the received signal in block $k-1$ and determined the quantization index (l_{k-1}). The macro-cell then generates a codeword X_m for l_{k-1} and transmits X_m to the small-cell in block k.

1) Transmit Signals: In block k, UE1, UE2, the small-cell and macro-cell respectively transmit $X_1(w_{1,k}, w_{1,k-2})$, $X_2(w_{2,k}, w_{2,k-2})$, $X_s(w_{1,k-2}, w_{2,k-2})$ and $X_m(l_{k-1})$. They construct their transmit signals as follows:

$$X_1 = \sqrt{\rho_{n1}}V_1(w_{1,k}) + \sqrt{\rho_{o1}}U_1(w_{1,k-2}),$$
$$X_2 = \sqrt{\rho_{n2}}V_2(w_{2,k}) + \sqrt{\rho_{o2}}U_2(w_{2,k-2}),$$
$$X_s = \sqrt{\rho_{s1}}U_1(w_{1,k-2}) + \sqrt{\rho_{s2}}U_2(w_{2,k-2}),$$
$$X_m = \sqrt{P_m}V_m(l_{k-1}),\ \hat{Y}_m = Y_m + \hat{Z}_m, \quad (2)$$

where $\hat{Y}_m$ is the quantized version of Y_m and $\hat{Z}_m$ is the quantization noise with zero mean and Q variance $\sim \mathcal{CN}(0, Q)$. The signals V_1, U_1, V_2, U_2, and V_m are all i.i.d Gaussian signals $\sim \mathcal{CN}(0, 1)$ that convey the codewords of the messages $w_{1,k}$, $w_{1,k-2}$, $w_{2,k}$, $w_{2,k-2}$ and the bin index l_{k-1}, respectively. While the macro-cell transmits the quantization index with power P_m, the power allocation parameters by UE1 (ρ_{n1}, ρ_{o1}), UE2 (ρ_{n2}, ρ_{o2}), and the small-cell (ρ_{s1}, ρ_{s2}) satisfy the following constraints:

$$\rho_{n1} + \rho_{o1} = P_1,\ \rho_{n2} + \rho_{o2} = P_2,\ \rho_{s1} + \rho_{s2} = P_s. \quad (3)$$

where P_1, P_2, P_s and P_m are the transmit powers from UE1, UE2, the small-cell and the macro-cell, respectively.

2) Decoding: The small and macro cells can decode both UEs' information using joint typicality (JT) [40] or maximum likelihood (ML) [47] decoding. A brief description of the decoding techniques will be given here, with the full analysis in Appendix I.

[1] This may reduce the average achievable spectral efficiency by a factor of (2/B). However, this factor becomes negligible as $B \to \infty$ [40].

[2] In the first (last) two blocks, the old (new) information is 1 by convention, i.e., $w_{1,-1} = w_{1,0} = w_{1,B-1} = w_{1,B} = 1$ [32], [40].

[3] In block Markov encoding, the codeword sent in each block depends not only on the new information but also on old information from an earlier block. Sending two or more information parts is performed through superposition coding where for each transmit codeword of the old information part, a group of codewords is generated for the new information parts [40 Ch. 16].

At the Small-Cell: The small-cell performs sliding window decoding to decode the information from both UEs and the quantization index from the macro-cell. At the end of block $k+1$, the small-cell has already estimated $\hat{w}_{1,k-2}$, $\hat{w}_{2,k-2}$ and l_{k-2} ($\hat{w}_{1,k-1}$, $\hat{w}_{2,k-1}$ and l_{k-1}) from the decoding in blocks $k-2$ and $k-1$ ($k-1$ and k). It then simultaneously utilizes the received signals in blocks k and $k+1$ ($Y_{s,k}, Y_{s,k+1}$) to jointly decode both UEs' information ($w_{1,k}, w_{2,k}$) and the quantization index (l_k). The transmission rates for UE1 and UE2 information that ensure reliable decoding at the small-cell are derived as follows (see Appendix I):

$$\begin{aligned} R_1 \le \min\{\acute{I}_1, \acute{I}_2\}, \quad R_2 \le \min\{\acute{I}_4, \acute{I}_5\}, \\ R_1 + R_2 \le \min\{\acute{I}_7, \acute{I}_8\}, \end{aligned} \tag{4}$$

where

$$\begin{aligned} \acute{I}_1 &= \mathcal{C}\left(g_{s1,k}^2\rho_{n1} + \frac{g_{m1,k}^2\rho_{n1}}{1+Q}\right), \\ \acute{I}_2 &= \mathcal{C}\left(\frac{g_{sm,k+1}^2 P_m}{1+g_{s1,k+1}^2\rho_{n1}+g_{s2,k+1}^2\rho_{n2}}\right) \\ &\quad + \mathcal{C}(g_{s1,k}^2\rho_{n1}) - \mathcal{C}(\frac{1}{Q}), \\ \acute{I}_4 &= \mathcal{C}\left(g_{s2,k}^2\rho_{n2} + \frac{g_{m2,k}^2\rho_{n2}}{1+Q}\right), \\ \acute{I}_5 &= \mathcal{C}\left(\frac{g_{sm,k+1}^2 P_m}{1+g_{s1,k+1}^2\rho_{n1}+g_{s2,k+1}^2\rho_{n2}}\right) \\ &\quad + \mathcal{C}(g_{s2,k}^2\rho_{n2}) - \mathcal{C}(\frac{1}{Q}), \\ \acute{I}_7 &= \mathcal{C}\left(g_{s1,k}^2\rho_{n1} + g_{s2,k}^2\rho_{n2} + \frac{g_{m1,k}^2\rho_{n1}+g_{m2,k}^2\rho_{n2}}{1+Q}\right), \\ \acute{I}_8 &= \mathcal{C}\left(\frac{g_{sm,k+1}^2 P_m}{1+g_{s1,k+1}^2\rho_{n1}+g_{s2,k+1}^2\rho_{n2}}\right) \\ &\quad + \mathcal{C}(g_{s1,k}^2\rho_{n1} + g_{s2,k}^2\rho_{n2}) - \mathcal{C}(\frac{1}{Q}), \end{aligned} \tag{5}$$

where $\mathcal{C}(x) = \log(1+x)$ and Q is the quantization noise variance as shown in (2) and is subject to

$$Q \ge \frac{1+g_{s1,k+1}^2\rho_{n1}+g_{s2,k+1}^2\rho_{n2}}{g_{sm,k+1}^2 P_m} \tag{6}$$

In (6), the bound on the quantization noise variance is obtained since the transmission rate for the quantization index sent by the macro-cell is bounded by the link quality from the macro to small-cell.

At the Macro-Cell: The macro-cell performs backward decoding for both UEs' information. In block $k+2$, it has already estimated the new information $\tilde{w}_{1,k+2}$ and $\tilde{w}_{2,k+2}$ from the decoding in block $k+4$. It then utilizes the received signal in block $k+2$ ($Y_{m,k+2}$) to jointly decode both UEs' old information ($w_{1,k}, w_{2,k}$) at the following rates:

$$\begin{aligned} R_1 &\le \mathcal{C}\big(g_{m1,k+2}^2 P_1 + g_{ms,k+2}^2\rho_{s1} \\ &\qquad + 2g_{m1,k+2}g_{ms,k+2}\sqrt{\rho_{o1}\rho_{s1}}\big) = \acute{I}_3, \\ R_2 &\le \mathcal{C}\big(g_{m2,k+2}^2 P_2 + g_{ms,k+2}^2\rho_{s2} \\ &\qquad + 2g_{m2,k+2}g_{ms,k+2}\sqrt{\rho_{o2}\rho_{s2}}\big) = \acute{I}_6, \\ R_1 + R_2 &\le \mathcal{C}\Big(g_{m1,k+2}^2 P_1 + g_{m2,k+2}^2 P_2 + g_{ms,k+2}^2 P_s \\ &\qquad + 2g_{m1,k+2}g_{ms,k+2}\sqrt{\rho_{o1}\rho_{s1}} \\ &\qquad + 2g_{m2,k+2}g_{ms,k+2}\sqrt{\rho_{o2}\rho_{s2}}\Big) = \acute{I}_9, \end{aligned} \tag{7}$$

where the term $2g_{m1,k}g_{dr,k}\sqrt{\rho_{o1}\rho_{s1}}$ ($2g_{m2,k}g_{dr,k}\sqrt{\rho_{o2}\rho_{s2}}$) shows the beamforming gain of the coherent transmission from UE1 (UE2) and the small-cell.

B. Spectral Efficiency Region

The error analyses for the decoding rules at the small and macro cells lead to some rate constraints in (4) and (7) that determine the spectral efficiency region in the following Theorem:

Theorem 1: For the HetNet uplink transmission with SMC, the spectral efficiency region consists of all rate pairs (R_1, R_2) *satisfying*

$$\begin{aligned} R_1 \le \min\{I_1, I_2, I_3\}, \quad R_2 \le \min\{I_4, I_5, I_6\}, \\ R_1 + R_2 \le \min\{I_7, I_8, I_9\}, \end{aligned} \tag{8}$$

$$\text{subject to } Q \ge Q_c, \quad Q_c = \frac{1+g_{s1}^2\rho_{n1}+g_{s2}^2\rho_{n2}}{g_{sm}^2 P_m}, \tag{9}$$

where $I_i = \acute{I}_u$, $u \in \{1, 2, ..., 9\}$ *but after setting all* $g_{ij,k} = g_{ij,k+1} = g_{ij,k+2} = g_{ij}$ *for* $i \in \{s, m\}$ *and* $j \in \{1, 2, m, s\}$ *as the channel is constant over AWGN. The power allocation parameters satisfy (3).*

Proof: Constraints I_1, I_2, I_4, I_5, I_7, I_8 and the quantiation noise bound are obtained from decoding at the small-cell while I_3, I_6 and I_9 are obtained from decoding at the macro-cell. For detailed proof, see Appendix I. □

The impact of each encoding/decoding technique can be seen in Theorem 1 as follows:

- Block Markov encoding at UEs and DF relaying at the small-cell facilitate coherent transmission between each UE and the samll-cell [40]. This coherent transmission leads to the beamforming gain appears in I_3, I_6 and I_9.
- Backward decoding is used with blcok Markov encoding [40] and it is stronger than direct decoding or sliding window decoding [21]. Besides, for outage performance, backward decoding has simpler analysis and better performance than sliding window decoding even when they both achieve the same throughput [37], [48].
- QF relaying helps improve the decoding at the small-cell since it will receive signals not only from the UEs but also from the macro-cell thorugh the feedback link.

- Joint sliding window decoding helps the small-cell utlize the signals from UEs (in block k) and the macro-cell (in block $k+1$) to decode UEs messages. This decoding is better than considering one block only. While the backward decoding is stornger, it has long delay and the small-cell needs to forward UEs messages to the macro-cell.
- Joint deocding for the bin index and UEs messages is stronger than sequential deocding (bin index first and then UEs messages) [32]. Besides, for single UE [32], joint decoding helps achieve the capacity for some channel conditions, which we analyze in Section V.
- I_1, I_2, I_4, I_5, I_7, and I_8 show the impact of joint sliding window decoding at the small-cell and QF relaying at the macro-cell. Without the quantized signal from the macro-cell through the feedback link (See Remark 3), the small-cell decodes UEs directly at the end of each block (no sliding window decoding) as in classical MAC. Then, we have $I_1 = I_2 = \mathcal{C}(g_{s1}^2\rho_{n1})$, $I_4 = I_5 = \mathcal{C}(g_{s2}^2\rho_{n2})$ and $I_7 = I_8 = \mathcal{C}(g_{s1}^2\rho_{n1} + g_{s2}^2\rho_{n2})$.

C. Discussion

We have some remarks on the proposed scheme.

Remark 1: As in [32], this scheme uses a second-order block Markov encoding where the transmitted codeword in block k depends on the information transmitted in block $k-2$. This is because the small-cell does not decode UE1 and UE2 information in block k directly, but rather waits another block $k+1$ to receive the quantized signal from the macro-cell and then decodes the UE1 and UE2 information using the signals it received in blocks k and $k+1$. It then forwards this information in block $k+2$.

Remark 2: The proposed scheme can be generalized by using partial DF relaying at the small-cell, where each UE splits its message into two parts: private (decoded at the macro-cell only) and public (decoded at both the small and macro cells). However, we use only full DF relaying, since our focus is on understanding the impact of SMC in enlarging the spectral efficiency region. Besides, for practical implementation, full DF relaying is simpler than partial DF relaying, which requires further processing at each UE for optimal rate splitting and power allocation between its private and public message parts.

Remark 3: Regarding the MARC, which resembles HetNet uplink transmission, the proposed scheme includes the following existing schemes as special cases:

- The DF scheme for MARC without RDC (e.g. SMC) [4], [20]. This can be verified by setting $X_m = \hat{Y}_m = \emptyset$ and $P_m = 0$.
- The DF scheme for the basic relay channel [19] by setting $X_2 = X_m = \hat{Y}_m = \emptyset$ and $P_2 = P_m = 0$.

Moreover, if we generalize the scheme to partial DF relaying as in Remark 2, it will include the partial DF relaying with RDC for relay channel described in [32] if we set $X_2 = \emptyset$ and $P_2 = 0$.

Remark 4: A noisy network coding (NNC) scheme [41] can be applied to the channel model in Fig. 2. This scheme is based on message repetition at the sources in all transmission blocks, QF relaying at the small and macro cells, and simultaneous joint decoding over all transmission blocks at the macro-cell. By applying [41, Th. 1] to HetNet with SMC, we obtain the spectral efficiency region that consists of all rate pairs (R_1, R_2) satisfying

$$R_1 \le \min\{J_1, J_2\},\ R_2 \le \min\{J_3, J_4\},$$
$$R_1 + R_2 \le \min\{J_5, J_6\} \text{ where}$$
$$J_1 = \mathcal{C}\left(g_{m1}^2 P_1 + \frac{g_{s1}^2 P_1}{1+Q_r}\right),\ J_3 = \mathcal{C}\left(g_{m2}^2 P_2 + \frac{g_{s2}^2 P_2}{1+Q_r}\right),$$
$$J_2 = \mathcal{C}\left(g_{m1}^2 P_1 + g_{ms}^2 P_s\right) - \mathcal{C}(\frac{1}{Q_r}),$$
$$J_4 = \mathcal{C}\left(g_{m2}^2 P_2 + g_{ms}^2 P_s\right) - \mathcal{C}(\frac{1}{Q_r}),$$
$$J_5 = \mathcal{C}\Big(g_{m1}^2 P_1 + g_{m2}^2 P_2 + \frac{g_{s1}^2 P_1 + g_{s2}^2 P_2 + (g_{s1}g_{m2} - g_{s2}g_{m1})^2 P_1 P_2}{1+Q_r}\Big),$$
$$J_6 = \mathcal{C}\left(g_{m1}^2 P_1 + g_{m2}^2 P_2 + g_{ms}^2 P_s\right) - \mathcal{C}(\frac{1}{Q_r}),$$

where Q_r is the quantization noise variance obtained from quantization at the small-cell. The optimal Q_r^* for individual and sum rates are obtained in a similar way to Q_1^*, Q_2^* and Q_s^* in Corollary 1. Unlike the proposed scheme with SMC, the feedback from macro to small-cell in the NNC scheme does not improve the spectral efficiency region. This is because in the proposed scheme, the small-cell decodes the users' information, which is improved with the signal received from the macro-cell. In the NNC scheme, however, the small-cell quantizes its received signal and the macro-cell won't benefit from receiving a quantized version of a signal it knows already. Hence, NNC schemes for classical HetNet uplink transmission with or without SMC have the same spectral efficiency region.

IV. Optimal and Active Quantization

From a practical implementation perspective, it is important to derive, at the macro-cell, the optimal quantization that maximizes the spectral efficiency of the proposed scheme. This section derives the optimal Q^* values that maximize the individual rates (R_1, R_2) and the sum rate $(R_1 + R_2)$ in Theorem 1. In addition to these values, we derive all other Q values that determine the whole spectral efficiency region. Since Q appears only on the rate constraints at the small-cell, the Q values that maximize the whole spectral efficiency region in Theorem 1 are those that maximize it at the small-cell.

A. Optimal Q_1^*, Q_2^* and Q_s^*

Since I_1 is a decreasing function with respect to Q while I_2 is a an increasing function, the optimal Q_1^* that maximizes R_1 is obtained from the intersection between I_1 and I_2. The same holds true for the optimal Q_2^* of R_2 and Q_s^* of $R_1 + R_2$. Q_1^*, Q_2^* and Q_s^* are given as follows:

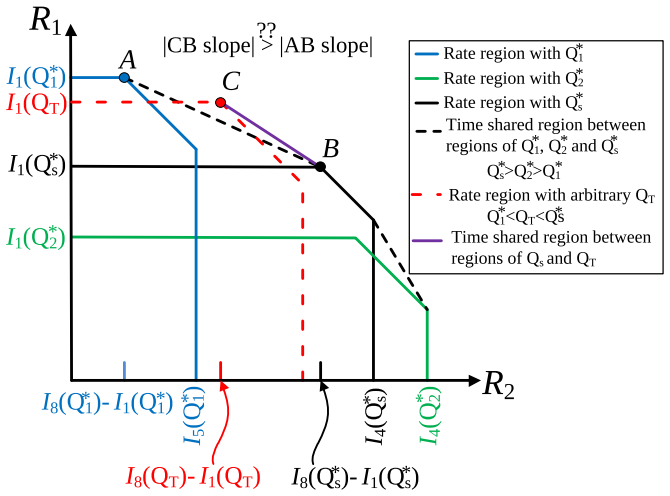

Fig. 3. Time-shared rate region at the macro-cell between Q_1^*, Q_2^* and Q_s^* regions.

Corollary 1: The optimal Q_1^*, Q_2^* and Q_s^* that respectively maximize R_1, R_2 and R_1+R_2 in Theorem 1 are given as

$$Q_1^* = \frac{\left(1+\frac{g_{s2}^2\rho_{n2}}{1+g_{s1}^2\rho_{n1}}\right)\left(1+(g_{s1}^2+g_{m1}^2)\rho_{n1}\right)}{g_{sm}^2 P_m},$$
$$Q_2^* = \frac{\left(1+\frac{g_{s1}^2\rho_{n1}}{1+g_{s2}^2\rho_{n2}}\right)\left(1+(g_{s2}^2+g_{m2}^2)\rho_{n2}\right)}{g_{sm}^2 P_m},$$
$$Q_s^* = \frac{1+(g_{s1}^2+g_{m1}^2)\rho_{n1}+(g_{s2}^2+g_{m2}^2)\rho_{n2}}{g_{sm}^2 P_m}. \quad (10)$$

Proof: In Theorem 1, R_1 is maximized when $I_1 = I_2$, since I_1 decreases while I_2 increases as Q increases. By setting $I_1 = I_2$ and solving for Q, we obtain Q_1^* in (10). Similarly, R_2 is maximized with a Q_2^* value that makes $I_4 = I_5$ while R_1+R_2 is maximized with a Q_s^* value that makes $I_7 = I_8$ in (8). □

Remark 5: Q_1^*, Q_2^* and Q_s^* in (10) satisfy the condition in (9) since they are greater than Q_c.

B. Active Q Ranges

Corollary 1 shows the optimal quantization that maximizes the individual and sum rates. Let $\mathcal{G}(Q_i^*)$ for $i \in \{1,2,s\}$ be the spectral efficiency region obtained with Q_i^* and let $\mathcal{G}(Q_1^*, Q_2^*, Q_s^*)$ be the time-shared region between $\mathcal{G}(Q_1^*)$, $\mathcal{G}(Q_2^*)$ and $\mathcal{G}(Q_s^*)$.[4] In addition, define Q_T as a value of Q that belongs to the active ranges of Q in which $\mathcal{G}(Q_T)$ lies outside $\mathcal{G}(Q_1^*, Q_2^*, Q_s^*)$ as shown in Fig. 3. The active ranges of Q can then be determined as in the following Theorem:

Theorem 2: For each set of power parameters, the spectral efficiency region in Theorem 1 is determined through a time-shared region between $\mathcal{G}(Q_1^*)$, $\mathcal{G}(Q_2^*)$ and $\mathcal{G}(Q_s^*)$ and all $\mathcal{G}(Q_T)$ for the following Q_T ranges:

$$Q_T \in \begin{cases} (Q_1^*, Q_1^{1T}) \cup (Q_2^*, Q_2^{4T}) & \text{if } \max\{Q_1^*, Q_2^*\} < Q_s^* \\ (Q_1^{2T}, Q_1^*) \cup (Q_2^{5T}, Q_2^*) & \text{if } Q_s^* < \min\{Q_1^*, Q_2^*\} \\ (Q_1^*, Q_1^{1T}) \cup (Q_2^{5T}, Q_2^*) & \text{if } Q_1^* < Q_s^* < Q_2^* \\ (Q_1^{2T}, Q_1^*) \cup (Q_2^*, Q_2^{4T}) & \text{if } Q_2^* < Q_s^* < Q_1^*, \end{cases} \quad (11)$$

where Q_i^{lT} for $i \in \{1,2\}$ and $l \in \{1,2,4,5\}$ is the solution of $f_{l,i,j}(Q_T) = 0$ where $j = 7$ (8) if $l \in \{2,5\}$ ($\{1,4\}$) and

$$f_{l,i,j}(Q_T) = \delta_{ij} I_l(Q_s^*) + (1-\delta_{ij}) I_l(Q_i^*) - I_l(Q_T),$$
$$\delta_{ij} = \frac{I_j(Q_T) - I_j(Q_i^*)}{I_j(Q_s^*) - I_j(Q_i^*)}. \quad (12)$$

Sketch of the Proof:. See Appendix II for the detailed proof. It is obtained by showing that for any values of $Q > Q_c$ where Q_c is given in (9), only Q_T defined in (11) leads to a spectral efficiency region outside $\mathcal{G}(Q_1^*, Q_2^*, Q_s^*)$. First, we show that for any $Q < \min\{Q_1^*, Q_2^*, Q_s^*\}$ or $Q > \max\{Q_1^*, Q_2^*, Q_s^*\}$, $\mathcal{G}(Q)$ is included inside $\mathcal{G}(Q_1^*)$, $\mathcal{G}(Q_2^*)$ or $\mathcal{G}(Q_s^*)$. This can be proven on the basis of the fact that I_1, I_4 and I_7 (I_2, I_5 and I_8) are decreasing (increasing) functions of Q.

Second, when $\min\{Q_1^*, Q_2^*, Q_s^*\} < Q < \max\{Q_1^*, Q_2^*, Q_s^*\}$, $\mathcal{G}(Q)$ is outside $\mathcal{G}(Q_1^*, Q_2^*, Q_s^*)$ only for $Q \in Q_T$ in (11). This can be proven with help from Fig. 3. Assume that $Q_1^* < Q < Q_s^*$, $\mathcal{G}(Q)$ is then outside $\mathcal{G}(Q_1^*, Q_2^*, Q_s^*)$ in point C (Fig. 3) if $|SL_1| < |SL_2|$ where $|x|$ is the absolute value of x, SL_1 is the slope of the time-shared line between $\mathcal{G}(Q_1^*)$ and $\mathcal{G}(Q_s^*)$ (AB line) and SL_2 is the slope of the time-shared line between $\mathcal{G}(Q)$ and $\mathcal{G}(Q_s^*)$ (CB line). Considering the coordinates of points A, B and C, the formula $|SL_1| < |SL_2|$ becomes

$$\delta_{18} I_1(Q_s^*) + (1-\delta_{18}) I_1(Q_1^*) < I_1(Q_T),$$
$$\delta_{18} = \frac{I_8(Q_T) - I_8(Q_1^*)}{I_8(Q_s^*) - I_8(Q_1^*)}. \quad (13)$$

Since $Q_1^* < Q < Q_s^*$, $I_8(Q_1^*) < I_8(Q) < I_8(Q_s^*)$ as $I_8(Q)$ is an increasing function of Q and hence, $0 < \delta_{18} < 1$. Then, since I_1 is a decreasing function of Q, formula (13) is satisfied when $Q \in (Q_1^*, Q_1^{1T})$. A similar analysis holds for all other cases where $\min\{Q_1^*, Q_2^*, Q_s^*\} < Q < \max\{Q_1^*, Q_2^*, Q_s^*\}$. □

Theorem 2 and Corollary 1 imply the following about the constraint $Q > Q_c$ in Theorem 1:

Proposition 1: In Theorem 1, the constraint on the quantization noise variance $Q > Q_c$ is inactive.

Proof: Obtained, since Q_1^*, Q_2^*, Q_s^* and Q_T in (10) and (11) are sufficient to obtain the whole spectral efficiency region and they are all greater than Q_c. □

Theorem 2 and Corollary 1 further imply that as the macro-cell power increases, the quantization noise variance decreases, since the macro-cell increases the quantization levels of its received signal. Moreover, the macro-cell transmits the quantization index with higher power, which improves the spectral efficiency at the small-cell. Hence, the next section analyzes the asymptotic spectral efficiency region as the macro-cell power approaches infinity.

[4]Time sharing is an encoding technique used to achieve the convex closure of two regions [40]. Considering $\mathcal{G}(Q_1^*)$ and $\mathcal{G}(Q_s^*)$, to reach a point on the AB line in Fig. 3, the macro-cell sends a codeword that includes $\alpha \in (0,1)$ portion of the bin index codeword pertaining to Q_s^* and $(1-\alpha)$ portion of the bin index codeword pertaining to Q_1^*. By varying α, we obtain the whole AB line [40].

V. Capacity Achieving at High Macro-Cell Power

As mentioned in Section I, the macro-cell transmission power in HetNet is about 200x that of the small-cell and 1000x that of the UE. Therefore, the macro-cell can cooperate with the small-cell at high power to enlarge the spectral efficiency region. Although high transmission power in full-duplex mode increases self-interference, the interference cancellation technique in [9] can reduce it by much more than the 110 dB that is achieved for WiFi radio with 100mW of transmission power. Moreover, as a theoretical limit, this section shows that the spectral efficiency of the proposed scheme can asymptotically achieve capacity as $P_m \to \infty$ by reaching the cut-set bound when the link amplitude ratios $\frac{g_{m1}}{g_{s1}} = \frac{g_{m2}}{g_{s2}}$.

A. Achievable Spectral Efficiency at $P_m \to \infty$

When the macro-cell has enough power ($P_m \to \infty$), it can reduce the quantization noise by increasing the number of bin indices. This allows it to send a very clear version of its received signal to the small-cell to improve its decoding, which leads to the following spectral efficiency region:

Corollary 2: The proposed HetNet transmission scheme with $P_m \to \infty$ achieves the spectral efficiency region in Theorem 1 with

$$I_1 \to \mathcal{C}\left((g_{s1}^2+g_{m1}^2)\rho_{n1}\right), \quad I_4 \to \mathcal{C}\left((g_{s2}^2+g_{m2}^2)\rho_{n2}\right),$$
$$I_7 \to \mathcal{C}\left((g_{s1}^2+g_{m1}^2)\rho_{n1}+(g_{s2}^2+g_{m2}^2)\rho_{n2}\right), \tag{14}$$

$I_2 = I_1$, $I_4 = I_5$, and $I_7 = I_8$ while I_3, I_6, and I_9 are given as in Theorem 1.

Proof: In Theorem 1, Corollary 1 shows that the individual and sum rates are maximized at Q_1^*, Q_2^* and Q_s^* values in (10) that make $I_1 = I_2$, $I_4 = I_5$ and $I_7 = I_8$. However, as $P_m \to \infty$, Q_c in Theorem 1 and all Q_1^*, Q_2^* and Q_s^* values in Corollary 1 approach zero $(Q_c, Q_1^*, Q_2^*, Q_s^*) \to 0$. Therefore, as $P_m \to \infty$, the whole spectral efficiency is maximized as $Q \to 0$. As $Q \to 0$, I_1, I_4 and I_7 in Theorem 1 become as in (14). □

B. Cut-Set Outer Bound

For the uplink HetNet with SMC in (1), there are 6 cutsets as shown in Fig. 4. By determining by the maximum information flowing through these cutsets, the cut-set bound is then given as follows [40].

Corollary 3: The capacity region of the uplink HetNet with SMC is upper bounded by the rate pairs (R_1, R_2) satisfying

$$R_1 \le \min\{C_1, C_2\}, \quad R_2 \le \min\{C_3, C_4\},$$
$$R_1 + R_2 \le \min\{C_5, C_6\}, \tag{15}$$

where

$$C_1 = \mathcal{C}\left((g_{s1}^2+g_{m1}^2)\beta_1 P_1\right), \quad C_3 = \mathcal{C}\left((g_{s2}^2+g_{m2}^2)\gamma_1 P_2\right),$$
$$C_2 = \mathcal{C}\left(g_{m1}^2\beta_2 P_1 + g_{ms}^2\beta_3 P_s + 2g_{m1}g_{ms}\beta_4\sqrt{P_1 P_s}\right),$$
$$C_4 = \mathcal{C}\left(g_{m2}^2\gamma_2 P_2 + g_{ms}^2\gamma_3 P_s + 2g_{m2}g_{ms}\gamma_4\sqrt{P_2 P_s}\right),$$

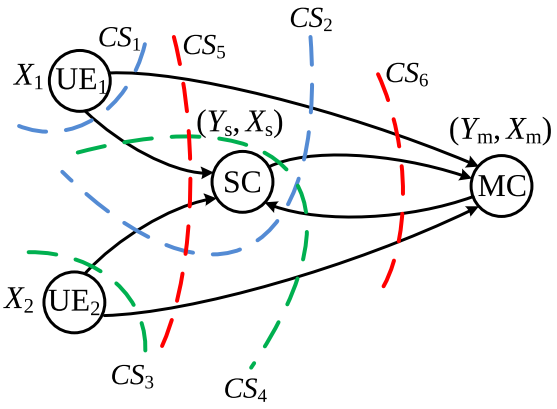

Fig. 4. The 6 cutsets for uplink HetNet with SMC.

$$C_5 = \mathcal{C}\Big((g_{s1}^2+g_{m1}^2)\mu_1 P_1 + (g_{s2}^2+g_{m2}^2)\mu_2 P_2 - 2(g_{m1}g_{m2}+g_{s1}g_{s2})\mu_3\sqrt{P_1 P_2} + (g_{s1}g_{m2}-g_{s2}g_{m1})^2\mu_4 P_1 P_2\Big),$$
$$C_6 = \mathcal{C}\Big(g_{m1}^2 P_1(1-\delta_{1m}^2)+g_{m2}^2 P_2(1-\delta_{2m}^2)+g_{ms}^2 P_s(1-\delta_{sm}^2) - 2g_{m1}g_{m2}\delta_{1m}\delta_{2m}\sqrt{P_1 P_2} + 2g_{m1}g_{ms}\sqrt{P_1 P_s}(\delta_{1s}-\delta_{1m}\delta_{sm}) + 2g_{m2}g_{ms}\sqrt{P_2 P_s}(\delta_{2s}-\delta_{2m}\delta_{sm})\Big), \tag{16}$$

and

$$\beta_1 = \frac{\eta}{1-\delta_{2s}^2-\delta_{2m}^2-\delta_{sm}^2+2\delta_{2s}\delta_{2m}\delta_{sm}},$$
$$\eta = 1-\delta_{1s}^2-\delta_{1m}^2-\delta_{2s}^2-\delta_{2m}^2-\delta_{sm}^2+\delta_{1m}^2\delta_{2s}^2+\delta_{2m}^2\delta_{1s}^2 + 2\delta_{1s}\delta_{1m}\delta_{sm}+2\delta_{2s}\delta_{2m}\delta_{sm}-2\delta_{1s}\delta_{1m}\delta_{2s}\delta_{2m},$$
$$\beta_2 = 1-\frac{\delta_{1m}^2}{1-\delta_{2m}^2}, \quad \beta_3 = 1-\frac{\delta_{2s}^2+\delta_{sm}^2-2\delta_{2s}\delta_{2m}\delta_{sm}}{1-\delta_{2m}^2}$$
$$\beta_4 = \delta_{1s}-\frac{\delta_{1m}\delta_{sm}-\delta_{1m}\delta_{2m}\delta_{2s}}{1-\delta_{2m}^2},$$
$$\mu_1 = 1-\frac{\delta_{1m}^2+\delta_{1s}^2}{1-\delta_{sm}^2}, \quad \mu_2 = 1-\frac{\delta_{2m}^2+\delta_{2s}^2}{1-\delta_{sm}^2}, \quad \mu_= \frac{\eta}{1-\delta_{sm}^2},$$
$$\mu_3 = \frac{\delta_{1m}\delta_{2m}+\delta_{1s}\delta_{2s}-\delta_{sm}(\delta_{1m}\delta_{2s}+\delta_{2m}\delta_{1s})}{1-\delta_{sm}^2}, \tag{17}$$

where γ_i, for $i \in \{1,2,3,4\}$ is similar to β_i except switching all indices from $1 \to 2$ and $2 \to 1$. $\delta_{un} \in [-1,+1]$ for $u \in \{1,2,s\}$ and $n \in \{s,m\}$ is the correlation factor between X_u and X_n and

$$\eta \ge 0. \tag{18}$$

Proof: By determining the mutual information of the 6 cutsets for the discrete memoryless MARC-RDC; applying them into the uplink HetNet with SMC Gaussian channel in (1); and showing that the optimal input distribution (X_1, X_2, X_s, X_m) is jointly Gaussian $\sim \mathcal{N}(0, \Sigma)$ where (18) ensures that Σ is a positive semi-definite covariance matrix. For detailed proof, see Appendix III. □

Remark 6: The cut-set bound for classical MARC (uplink HetNet without SMC) is given as in Corollary 3 but with $\delta_{1m} = \delta_{2m} = \delta_{sm} = 0$. In both bounds with or without SMC,

$0 \leq (\beta_i, \gamma_i, \mu_i) \leq 1$ for $i \in \{1, 2, 3, 4\}$. Therefore, the two bounds are the same.

From Corollaries 2 and 3, we obtain the following theorem:

Theorem 3: The proposed scheme for uplink HetNet with SMC achieves capacity as $P_m \to \infty$ by reaching the cut-set bound when the amplitude ratio of one UE-small to UE-macro cell link is equal to that of the other UE, i.e., $g_{m1}/g_{s1} = g_{m2}/g_{s2}$.

Proof: By comparing spectral efficiency regions in Corollaries 2 and 3. The region in Corollary 3 consists of the constraints in (15) with correlation factors satisfying (18) while the region in Corollary 2 consists of the constraints in Theorem 1 and their updates in (14) with power allocation parameters satisfying (3). After some mathematical manipulations, we can show that they have almost identical constraints ($I_1 \triangleq C_1$, $I_3 \triangleq C_2$, $I_4 \triangleq C_3$, $I_6 \triangleq C_4$, $I_9 = C_6$) except for $I_7 \neq C_5$ because of the term $(g_{s1}g_{m2} - g_{s2}g_{m1})^2 \mu_4 P_1 P_2$ in C_5 in (15). However, this term is 0 when $g_{m1}/g_{s1} = g_{m2}/g_{s2}$. Note that the negative term in C_5 is equal to zero for some correlation factors that make $\mu_3 = 0$. □

Theorem 3 shows that our scheme achieves the capacity without requiring the channel to be physically degraded with a specific correlation factor between small and macro cell noises (Z_s and Z_m) [28].

Remark 7: Theorem 3 is interesting for practical designers of 5G cellular systems. The theorem implies that two UEs in a small-cell can guarantee the maximum possible throughput when macro and small cells cooperate. However, the two UEs need to pay for such a service since the macro-cell will cooperate with high power.

VI. Outage Performance Over Fading Channel

High reliability (low outage probability) is required by many wireless applications. For some applications, a minimum target information rate is required, below which the service is unsustainable. For example, the 5G standard can support 50 Mbps with very high reliability (> 99%) [49] and VoIP service in LTE release 8 can tolerate up to 2% outage probability [50]. In fading channels, outage probability is the probability that the rate supported by the fading channel falls below the target rate.

In this section, we analyze the outage probability for uplink transmission in HetNet with SMC over block fading channels. This assumption simplifies the analysis but requires a transmission bandwidth less than the channel coherence bandwidth. Hence, with carrier aggregation, the block fading assumption is valid when the aggregated bands are less than the coherence bandwidth and when using contiguous carrier aggregation (intra-band), since non-contiguous carriers have different channel realization over different bands.

We take into account the outage events at both small and macro cells. At the small-cell, we further consider the channel variation over different blocks that results from the sliding window decoding over two consecutive blocks.

A. Fading Channel Model

We consider a block fading channel where the links remain constant in each transmission block and change independently in the next block. Hence, in block k, each link gain in (1)is modeled by Rayleigh fading and pathloss as follows:

$$\begin{aligned} h_{ij,k} &= \tilde{h}_{ij,k}/(d_{ij}^{\alpha_{ij}/2}), \\ &= g_{ij,k} e^{\sqrt{-1}\theta_{ij}},\ i \in \{s, m\},\ j \in \{1, 2, s, m\} \end{aligned} \quad (19)$$

where $\tilde{h}_{ij,k} \sim \mathcal{CN}(0, 1)$ represents the small scale fading. The large scale fading is captured by pathloss where d_{ij} is the distance between nodes i and j and α_{ij} is the attenuation factor. Let $g_{ij,k}$ and $\theta_{ij,k}$ be the amplitude and phase of a link coefficient in block k, then $g_{ij,k} = |\tilde{h}_{ij,k}|/d_{ij}^{\alpha_{ij}/2}$ is Rayleigh distributed while $\theta_{ij,k}$ is uniform in $[0, 2\pi]$. We assume full CSI at the receivers and partial CSI at the transmitters. This assumption implies the following: as receivers, the small-cell knows $h_{s1,k}$, $h_{s2,k}$ and $h_{sm,k}$ while the macro-cell knows $h_{m1,k}$, $h_{m2,k}$ and $h_{ms,k}$. As transmitters, since the outage probability depends on the statistics of the channel links, statistical (not instantaneous) knowledge of the links is required at both UEs and the small-cell such that they can optimize their transmit powers in (3) to minimize outage probability. Moreover, to perform coherent transmission, UE1, UE2 and the small-cell each knows the phase of its respective links to the macro-cell. This knowledge can be obtained via feedback from the macro-cell [42]. The macro-cell knows the links' amplitudes in each transmission block in order to perform the optimal quantization as explained in Section IV.

Unlike DF relaying schemes [35]–[37], [51], UEs need not know the instantaneous link orders at the beginning of each transmission block as in [36], [37], and [51] or whether there is an outage at the small-cell (e.g. relay) as in [35]. While these requirements are necessary to achieve a full diversity order of two for the schemes in [35]–[37] and [51], our scheme can achieve a full diversity order without this knowledge as shown in Section VIII-C, which is a further benefit of SMC.

B. Outage Probability

For a transmission scheme with different messages sent from each UE, an outage can occur for either or both UEs' messages. Hence, individual and common outages are defined in [52] where an individual outage pertains to incorrect decoding of one UE message regardless of the other UE message, while a common outage pertains to incorrect decoding of either or both UEs' messages.

For the proposed scheme with SMC, an outage for a UE message can occur at the small or macro cell; the average individual and common outage probabilities are then given as follows:

Theorem 4: For a given target rate pair (R_{1t}, R_{2t}) in the proposed scheme of uplink HetNet with SMC, the average common $(\bar{\mathbb{P}}_c)$ and individual $(\bar{\mathbb{P}}_1, \bar{\mathbb{P}}_2)$ outage probabilities are given as follows:

$$\begin{aligned} &\bar{\mathbb{P}}_c = \mathbb{P}_{sc} + (1 - \mathbb{P}_{sc})\mathbb{P}_{mc}, \\ &\bar{\mathbb{P}}_1 = \mathbb{P}_{s1} + (1 - \mathbb{P}_{s1})\mathbb{P}_{m1}, \quad \bar{\mathbb{P}}_2 = \mathbb{P}_{s2} + (1 - \mathbb{P}_{s2})\mathbb{P}_{m2}, \end{aligned} \quad (20)$$

where $\mathbb{P}_{sc}$ $(\mathbb{P}_{mc})$ is the common outage probability at the small (macro) cell. Similarly, $\mathbb{P}_{s1}$ $(\mathbb{P}_{m1})$ and $\mathbb{P}_{s2}$ $(\mathbb{P}_{m2})$ are

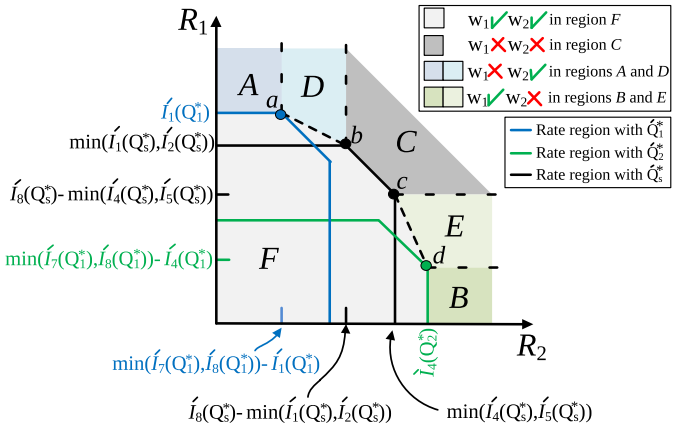

Fig. 5. Outage regions at the small-cell.

the individual outage probabilities at the small (macro) cell. The formulation for the common and individual outage probabilities at the small and macro cells are given in lemmas 1 and 2, respectively.

Proof: Obtained by considering the outage events at the small-cell and at the macro-cell when there is no outage at the small-cell. □

Note that the outages at the small and macro cells are independent because for given messages sent in block k, they are decoded in blocks k and $k+1$ at the small-cell while they are decoded in block $k+2$ at the macro-cell as shown in Section III-A. Because of block fading channels, the channel links in these blocks are independent, which makes the outage events at small and macro cells independent.

C. Outage at the Small-Cell

The common and individual outages at the small-cell can be derived with the help of Fig. 5, which shows the spectral efficiency region at the small-cell. To simplify the analyses, we consider only the time-shared region between $\mathcal{G}(Q_1^*)$, $\mathcal{G}(Q_2^*)$ and $\mathcal{G}(Q_s^*)$. Moreover, Q_1^*, Q_2^* and Q_s^* in (10) are slightly changed to consider the channel variation over different blocks. By setting $\acute{I}_1 = \acute{I}_2$, $\acute{I}_4 = \acute{I}_5$ and $\acute{I}_7 = \acute{I}_8$, in (4), we respectively obtain $\acute{Q}_1^*$, $\acute{Q}_2^*$ and $\acute{Q}_s^*$ as follows:

$$\acute{Q}_1^* = \frac{\left(1+g_{s1,k+1}^2\rho_{n1}+g_{s2,k+1}^2\rho_{n2}\right)\left(1+(g_{s1,k}^2+g_{m1,k}^2)\rho_{n1}\right)}{(1+g_{s1,k}^2\rho_{n1})g_{sm,k+1}^2P_m},$$

$$\acute{Q}_2^* = \frac{\left(1+g_{s1,k+1}^2\rho_{n1}+g_{s2,k+1}^2\rho_{n2}\right)\left(1+(g_{s2,k}^2+g_{m2,k}^2)\rho_{n2}\right)}{(1+g_{s2,k}^2\rho_{n2})g_{sm,k+1}^2P_m},$$

$$\acute{Q}_s^* = \frac{\left(1+g_{s1,k+1}^2\rho_{n1}+g_{s2,k+1}^2\rho_{n2}\right)}{\left(1+g_{s1,k}^2\rho_{n1}+g_{s2,k}^2\rho_{n2}\right)g_{sm,k+1}^2P_m} \times\left(1+(g_{s1,k}^2+g_{m1,k}^2)\rho_{n1}+(g_{s2,k}^2+g_{m2,k}^2)\rho_{n2}\right). \quad (21)$$

Note that $\acute{Q}_1^*$, $\acute{Q}_2^*$ and $\acute{Q}_s^*$ do not change the results of Theorem 2. Then, considering the time-shared region at the small-cell in Fig. 5, an outage occurs if the target rate pair (R_{1t}, R_{2t}) lies in region A, B, C, D, or E; the outage probability is then given as follows:

Lemma 1: For a target rate pair (R_{1t}, R_{2t}), the individual and common outage probabilities at the small-cell are given as follows:

$$\begin{aligned}\mathbb{P}_{s1} &= \mathbb{P}_s[A]+\mathbb{P}_s[D]+\mathbb{P}_s[C],\\ \mathbb{P}_{s2} &= \mathbb{P}_s[B]+\mathbb{P}_s[E]+\mathbb{P}_s[C],\\ \mathbb{P}_{sc} &= \mathbb{P}_s[A]+\mathbb{P}_s[B]+\mathbb{P}_s[C]+\mathbb{P}_s[D]+\mathbb{P}_s[E]. \quad (22)\end{aligned}$$

The probabilities for regions A–E are given as follows

$$\begin{aligned}\mathbb{P}_s[A] &= \mathbb{P}\big[R_{1t} > \acute{I}_1(\acute{Q}_1^*),\\ &\qquad R_{2t} \le \min\left\{\acute{I}_7(\acute{Q}_1^*), \acute{I}_8(\acute{Q}_1^*)\right\} - \acute{I}_1(\acute{Q}_1^*)\big] \quad (23)\\ \mathbb{P}_s[B] &= \mathbb{P}\big[R_{1t} \le \min\left\{\acute{I}_7(\acute{Q}_2^*), \acute{I}_8(\acute{Q}_2^*)\right\} - \acute{I}_4(\acute{Q}_2^*),\\ &\qquad R_{2t} > \acute{I}_4(\acute{Q}_2^*)\big]\\ \mathbb{P}_s[C] &= \mathbb{P}\big[R_{1t} > \acute{I}_7(\acute{Q}_s^*) - \min\left\{\acute{I}_4(\acute{Q}_s^*), \acute{I}_5(\acute{Q}_s^*)\right\},\\ &\qquad R_{2t} > \acute{I}_7(\acute{Q}_s^*) - \min\left\{\acute{I}_1(\acute{Q}_s^*), \acute{I}_2(\acute{Q}_s^*)\right\},\\ &\qquad R_{1t}+R_{2t} > \acute{I}_7(\acute{Q}_s^*)\big],\\ \mathbb{P}_s[D] &= \mathbb{P}\big[R_{2t} > \min\left\{\acute{I}_7(\acute{Q}_1^*), \acute{I}_8(\acute{Q}_1^*)\right\} - \acute{I}_2(\acute{Q}_1^*),\\ &\qquad R_{2t} \le \acute{I}_7(\acute{Q}_s^*) - \min\left\{\acute{I}_1(\acute{Q}_s^*), \acute{I}_2(\acute{Q}_s^*)\right\},\\ &\qquad R_{1t}+\zeta_1 R_{2t} > \vartheta_1\big],\\ \mathbb{P}_s[E] &= \mathbb{P}\big[R_{1t} > \min\left\{\acute{I}_7(\acute{Q}_2^*), \acute{I}_8(\acute{Q}_2^*)\right\} - \acute{I}_4(\acute{Q}_2^*),\\ &\qquad R_{1t} \le \acute{I}_7(\acute{Q}_s^*) - \min\left\{\acute{I}_4(\acute{Q}_s^*), \acute{I}_5(\acute{Q}_s^*)\right\},\\ &\qquad R_{1t}+\zeta_2 R_{2t} > \vartheta_2\big],\end{aligned}$$

where $\acute{I}_u(\acute{Q}_v^*)$, $u \in \{1,2,4,5,7,8\}$ and $v \in \{1,2,s\}$ is the rate constraint given in (4) with $\acute{Q}_v^*$ given in (21); ζ_1 (ζ_2) is the slope of the ab (cd) time-shared line in Fig. 5 while ϑ_1 (ϑ_2) is the intersection of this line with R_2 axis. These parameters are given as follows:

$$\begin{aligned}\zeta_1 &= \frac{I_1(Q_1^*) - \min\left\{\acute{I}_1(\acute{Q}_s^*), \acute{I}_2(\acute{Q}_s^*)\right\}}{\zeta_{11}-\zeta_{12}},\\ \zeta_{11} &= I_7(Q_s^*) - \min\left\{\acute{I}_1(\acute{Q}_s^*), \acute{I}_2(\acute{Q}_s^*)\right\},\\ \zeta_{12} &= \min\left\{\acute{I}_7(\acute{Q}_1^*), \acute{I}_8(\acute{Q}_1^*)\right\} - I_1(Q_1^*),\\ \vartheta_1 &= I_1(Q_1^*) + \zeta_1\zeta_{12},\\ \zeta_2 &= \frac{\zeta_{21}-\zeta_{22}}{I_4(Q_2^*) - \min\left\{\acute{I}_4(\acute{Q}_s^*), \acute{I}_5(\acute{Q}_s^*)\right\}},\\ \zeta_{21} &= I_7(Q_s^*) - \min\left\{\acute{I}_4(\acute{Q}_s^*), \acute{I}_5(\acute{Q}_s^*)\right\},\\ \zeta_{22} &= \min\left\{\acute{I}_7(\acute{Q}_2^*), \acute{I}_8(\acute{Q}_2^*)\right\} - I_4(Q_2^*),\\ \vartheta_2 &= \zeta_{22} + \zeta_2 I_4(Q_2^*). \quad (24)\end{aligned}$$

Proof: By determining the probabilities of regions A, B, C, D, and E and conducting similar analysis in [52], $w_{1,k}$ ($w_{2,k}$) is in outage if $(R_{1t}, R_{2,t})$ lies in region A (B), D (E) or C, while either or both messages are in outage if $(R_{1t}, R_{2,t})$ lies in region A, B, D, E or C. □

D. Outage at the Macro-Cell

The outage analyses at the macro-cell are quite similar to that of the classical MAC [52]. However, in the proposed scheme, UEs' messages sent in block k are decoded in block

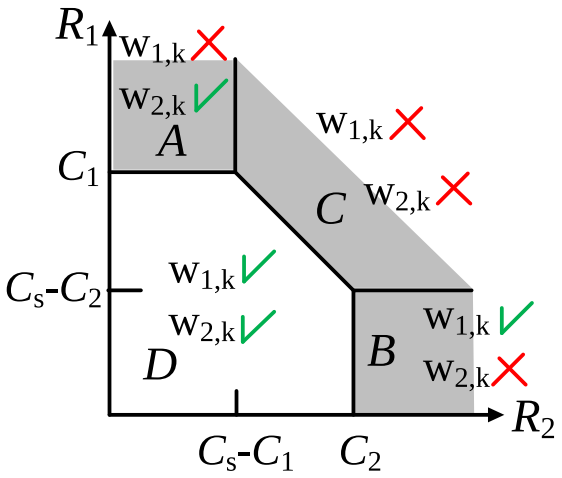

Fig. 6. Outage regions at the macro-cell BS.

$k+2$ at the macro-cell instead of the same block k as in classical MAC.

For UEs' messages sent in block k, the macro-cell reliably decodes them in block $k+2$ if their transmission rates (R_1, R_2) satisfy the constraints in (7). These constraints are illustrated in Fig. 6. With help from this figure, the outage probability can be derived such that for a given target rate pair $(R_{1t}, R_{2,t})$, there is no outage if the rate pair is in region D, an outage for UE1 (UE2) message $(w_{1,k})$ $((w_{2,k}))$ if the rate pair is in region A (B), and an outage for both UE1 and UE2 messages $(w_{1,k}$ and $w_{2,k})$ if the rate pair is in region C [52]. The rate pair can lie in these regions A, B and C with probabilities $\mathbb{P}_m[A]$, $\mathbb{P}_m[B]$, and $\mathbb{P}_m[C]$, respectively. Then, the individual and common outage probabilities are given as follows [52]:

Lemma 2: For a target rate pair (R_{1t}, R_{2t}), the individual and common outage probabilities at the macro-cell are given as follows [52]:

$$\begin{aligned}
&\mathbb{P}_{m1} = \mathbb{P}_m[A] + \mathbb{P}_m[C], \quad \mathbb{P}_{m2} = \mathbb{P}_m[B] + \mathbb{P}_m[C],\\
&\mathbb{P}_{mc} = \mathbb{P}_m[A] + \mathbb{P}_m[B] + \mathbb{P}_m[C], \text{ where,}\\
&\mathbb{P}_m[A] = \mathbb{P}\left[R_{1t} > I_3(k+2),\, R_{2t} \le I_9(k+2) - I_3(k+2)\right],\\
&\mathbb{P}_m[B] = \mathbb{P}\left[R_{1t} \le I_9(k+2) - I_6(k+2),\, R_{2t} > I_6(k+2)\right],\\
&\mathbb{P}_m[C] = \mathbb{P}\big[R_{1t} > I_9(k+2) - I_6(k+2),\\
&\quad R_{2t} > I_9(k+2) - I_3(k+2),\, R_{1t} + R_{2t} > I_9(k+2)\big],
\end{aligned} \tag{25}$$

and $I_3(k+2)$, $I_6(k+2)$ and $I_9(k+2)$ are respectively given as $\acute{I}_3$, $\acute{I}_6$ and $\acute{I}_9$ in (7).

Proof: By determining the probabilities of regions A, B, and C and conducting a similar analysis as in [52], $w_{1,k}$ $(w_{2,k})$ is in outage if $(R_{1t}, R_{2,t})$ lies in region A (B), or C while either or both messages are in outage if $(R_{1t}, R_{2,t})$ lies in region A, B or C. □

In Lemma 2, $I_i(k+2)$, $i \in \{3, 6, 9\}$ can be replaced with $I_i(k)$ since the channel links in blocks k and $k+1$ have the same statistics. However, we use $k+2$ since it comes directly from the backward decoding at the macro-cell and it explains the independency of the outage events at small and macro cells.

Remark 8: The diversity order of a transmission scheme is obtained by analyzing the asymptotic outage behavior at high SNR. However, because of space limitation, we only intuitively show that as P_m increases, the diversity becomes 2 for each UE's message despite the decoding at the small-cell. This is because for any UE's message to be lost, it always requires at least 2 links to be weak [35]. At the macro-cell, $w_{1,k}$ is in outage if $g_{m1,k+1}$ and $g_{ms,k+2}$ links are weak as deduced from $\acute{I}_3$ in (7). At the small-cell, $w_{1,k}$ is in outage if $g_{s1,k}$ and $g_{m1,k}$ are weak as deduced from $\acute{I}_1$ in (4) and provided that the quantization noise variance (Q) is very small. However, from (21), the optimal Q decreases as P_m increases until $Q \to 0$ as $P_m \to \infty$ and we obtain I_1 I_4, and I_7 in Corollary 2. A similar explanation holds for $w_{2,k}$.

VII. Generalization to N-UE

The previous sections consider the two-UE case as the basic multi-user model. However, it is common in practice to have many UEs located in a small cell. In this section, we generalize the two-UE scheme to an N-UE scheme and study its spectral efficiency, optimal quantization and reliability performance.

A. Transmission Scheme and Spectral Efficiency

The N-UE scheme is quite similar to the two-UE case where the transmission is carried over B blocks. In each block k, each UE (UEi for $i \in \{1, 2, ..., N\}$) transmits its new and old information $(w_{i,k}, w_{i,k-2})$ with powers ρ_{ni} and ρ_{oi}. The small-cell transmits the old information of all UEs with powers $(\rho_{s1}, \rho_{s2}, ..., \rho_{sN})$ and the macro cell transmits the quantization index with power P_m. Using encoding and decoding techniques similar to the two-UE case, we obtain the following spectral efficiency region:

Lemma 3: For the HetNet uplink transmission of N-UE with SMC, the spectral efficiency region consists of all N-tuples rate pairs $(R_1, R_2, \ldots, R_N)$ satisfying

$$\begin{aligned}
R_\Upsilon &\le \min\{I_{1,\Upsilon}, I_{2,\Upsilon}, I_{3,\Upsilon}\}, \qquad (26)\\
I_{1,\Upsilon} &= \mathcal{C}\left(\sum_{i\in\Upsilon} g_{si,k}^2\rho_{ni} + \frac{g_{mi,k}^2\rho_{ni}}{1+Q}\right),\\
I_{2,\Upsilon} &= \mathcal{C}\left(\frac{g_{sm,k+1}^2 P_m}{1+\sum_{i=1}^N g_{si,k+1}^2\rho_{ni}}\right)\\
&\quad + \mathcal{C}\left(\sum_{i\in\Upsilon} g_{si,k}^2\rho_{ni}\right) - \mathcal{C}\left(\frac{1}{Q}\right),\\
I_{3,\Upsilon} &= \mathcal{C}\Big(\sum_{i\in\Upsilon} g_{mi,k+2}^2 P_i + g_{ms,k+2}^2\rho_{si}\\
&\quad + 2g_{mi,k+2}g_{ms,k+2}\sqrt{\rho_{oi}\rho_{si}}\Big),
\end{aligned}$$

Subject to

$$Q > Q_N, \quad Q_N = \frac{1+\sum_{i=1}^N g_{si,k+1}^2\rho_{ni}}{g_{sm,k+1}^2 P_m} \tag{27}$$

for all subsets $\Upsilon \subseteq [1:N]$ where $R_\Upsilon = \sum_{i\in\Upsilon} R_i$, the power allocation set (ρ_{oi}, ρ_{ni}) for a UE i satisfying $\rho_{oi}+\rho_{ni} = P_i$ and

$(\rho_{s1}, \rho_{s2}, ..., \rho_{sN})$ for the small-cell satisfying $\sum_{i=1}^{N} \rho_{si} = P_s$ while the macro-cell sends its signal at power P_m.

Proof: Similar to the Proof of Theorem 1. □

It is straightforward to get Theorem 1 by setting $N = 2$ where the possible subsets Υ are $\{1\}$, $\{2\}$, and $\{1, 2\}$.

1) Optimal Quantization: In Lemma 3, similar to the two-UE case, $I_{1,\Upsilon}$ is a decreasing function of Q while $I_{2,\Upsilon}$ is an increasing function. Hence, the optimal quantization is given as follows:

Lemma 4: For each subset $\Upsilon \subseteq [1 : N]$ in Lemma 3, the optimal Q^*_Υ that maximizes the spectral efficiency is given as follows:

$$\acute{Q}^*_\Upsilon = \frac{\left(1 + \sum_{i=1}^{N} g^2_{si,k+1}\rho_{ni}\right)\left(1 + \sum_{i\in\Upsilon}(g^2_{si,k} + g^2_{mi,k})\rho_{ni}\right)}{(1 + \sum_{i\in\Upsilon} g^2_{si,k}\rho_{ni})g^2_{sm,k+1}P_m} \tag{28}$$

Proof: $\acute{Q}^*_\Upsilon$ is obtained from the solution of $I_{1,\Upsilon} = I_{2,\Upsilon}$ in (26). □

Note that $\acute{Q}^*_\Upsilon$ in (4) satisfies the condition in (27) since $\acute{Q}^*_\Upsilon > Q_N$ for each subset $\Upsilon \subseteq [1 : N]$.

2) Cut-Set Outer Bound: For the uplink HetNet with SMC and N-UE, there are $(N + 1) \times N$ cut-sets. These cut-sets determine the cut-set bound from the maximum information flowing through them as follows:

Lemma 5: The capacity region of the uplink HetNet with SMC and N-UE is upper bounded by the rate tuples $(R_1, R_2, \ldots, R_N)$ *satisfying*

$$R_\Upsilon \le \min\{C_{1,\Upsilon}, C_{2,\Upsilon}\}, \tag{29}$$

$$\begin{aligned} C_{1,\Upsilon} = \mathcal{C}\Big(&\sum_{i\in\Upsilon}(g^2_{si,k} + g^2_{mi})P_i\big(1 - \frac{\delta^2_{is}}{1 - \sum_{j\in\overline{\Upsilon}}\delta^2_{js}}\big) \\ &+ \sum_{(i,j)\in\Upsilon, j>i} -2(g_{mi}g_{mj} + g_{si}g_{sj})\delta_{is}\delta_{js}\sqrt{P_iP_j} \\ &+(g_{si}g_{mj} - g_{mi}g_{sj})^2(1 - \delta^2_{is} - \delta^2_{js})P_iP_j\Big), \\ C_{2,\Upsilon} = \mathcal{C}\Big(&\sum_{i\in\Upsilon} g^2_{mi}P_i + 2g_{mi}g_{ms}\delta_{is}\sqrt{P_iP_s} \\ &+g^2_{ms}P_s\big(1 - \sum_{j\in\overline{\Upsilon}}\delta^2_{js}\big)\Big), \end{aligned}$$

$$\text{Subject to} \quad \eta_N > 0, \quad \eta_N = 1 - \sum_{i=1}^{N}\delta_{is}, \tag{30}$$

for all subsets $\Upsilon \subseteq [1 : N]$ *where* $\overline{\Upsilon}$ *is the complement of* Υ. $\delta_{is} \in [-1, 1]$ *for* $i \in \{1, 2, \ldots, N\}$ *is the correlation factor between* X_i *and* X_s.

Proof: Obtained from Remark 6 and the proof of Corollary 3. Remark 6 shows that the cut-set bound for uplink HetNet with or without SMC is the same. Then, we obtain Lemma 5 by considering the $(N + 1) \times N$ cut-sets of N-UE uplink HetNet and following similar proof to that of Corollary 3. □

Corollary 4: The proposed scheme for N-UE uplink HetNet with SMC achieves capacity as $P_m \to \infty$ *by reaching the cut-set bound when the amplitude ratio of one UE-small to UE-macro cell link is equal to that of any other UE, i.e.,* $g_{mi}/g_{si} = g_{mj}/g_{sj}$ *for all* $(i, j) \in \{1, 2, \ldots, N\}$ *and* $i \neq j$.

Proof: Similar to the proof of Theorem 3 where as $P_m \to \infty$, the optimal $\acute{Q}^*_\Upsilon \to 0$ in (28). Then,

$$I_{1,\infty,\Upsilon} = I_{2,\infty,\Upsilon} \to \mathcal{C}\left(\sum_{i\in\Upsilon}(g^2_{si,k} + g^2_{mi,k})\rho_{ni}\right). \tag{31}$$

Therefore, from (31) and (29), $I_{1,\infty,\Upsilon} \triangleq C_{1,\Upsilon}$ since $\rho_{ni} \le P_i$ and $P_i\big(1 - \frac{\delta^2_{is}}{1-\sum_{j\in\overline{\Upsilon}}\delta^2_{js}}\big) \le P_i$ while the last term in $C_{1,\Upsilon}$ is zero when $g_{mi}/g_{si} = g_{mj}/g_{sj}$. In addition, from (31) and (29), $I_{3,\infty,\Upsilon} \triangleq C_{2,\Upsilon}$ since $\rho_{si} \le P_s$ and $P_s\big(1 - \sum_{j\in\overline{\Upsilon}}\delta^2_{js}\big) \le P_s$. □

B. Outage Performance

This section derives the common outage that pertains to the incorrect decoding of any or multiple UEs' messages. Following an analysis similar to that in Section VI-B, the common outage probability is given as follows:

Lemma 6: For a given target rate tuple $(R_{1t}, R_{2t}, ..., R_{Nt})$ *in the proposed uplink HetNet scheme with SMC, the average common* $(\bar{\mathbb{P}}_c)$ *is given as in (20) with*

$$\begin{aligned} \mathbb{P}_{sc} &= 1 - \mathbb{P}\left[(R_{1t}, R_{2t}, ..., R_{Nt}) \in \bigcup_{\Upsilon\subseteq[1:N]} \mathcal{G}(\acute{Q}^*_\Upsilon)\right], \\ \mathbb{P}_{mc} &= 1 - \mathbb{P}\big[R_\Upsilon < I_{3,\Upsilon}, \text{ for all } \Upsilon \subseteq [1 : N]\big], \end{aligned} \tag{32}$$

where $\mathcal{G}(\acute{Q}^*_\Upsilon)$ *is the spectral efficiency region achieved with* $\acute{Q}^*_\Upsilon$ *for any* $\Upsilon \subseteq [1 : N]$ *while* $\bigcup_{\Upsilon\subseteq[1:N]} \mathcal{G}(\acute{Q}^*_\Upsilon)$ *is the convex hull of all regions achieved with* $\acute{Q}^*_\Upsilon$ *for all* $\Upsilon \subseteq [1 : N]$.

Proof: For the small-cell, the correct probability $(\mathbb{P}_{sc}(CR))$ is the probability that the target rate tuple lies inside the convex hull region of all regions obtained with Q^*_Υ for every $\Upsilon \subseteq [1 : N]$. Then, $\mathbb{P}_{sc} = 1 - \mathbb{P}_{sc}(CR)$. Similarly, for the macro-cell, the correct probability $(\mathbb{P}_{mc}(CR))$ is the probability that the target rate tuple lies inside the rate region determined by rate constraints $I_{3,\Upsilon}$ for every $\Upsilon \subseteq [1 : N]$. Hence, $\mathbb{P}_{mc} = 1 - \mathbb{P}_{mc}(CR)$. □

VIII. Numerical Results

We now provide numerical results for the spectral efficiency regions of HetNet uplink transmission obtained by the proposed scheme with different values of the macro-cell power. We also compare our scheme with existing HetNet uplink schemes including DF relaying [4], [20] and NNC [41]. Figs. 7, 8 and 9 show the spectral efficiency while Figs. 10, 11 and 12 shows the outage performance. The channel parameters and transmit powers are given in each figure.

A. Spectral Efficiency

Fig. 7 shows how the proposed scheme with SMC enlarges the spectral efficiency region of the HetNet uplink transmission compared with DF relaying [4], [20] and NNC [41] schemes. SMC improves the spectral efficiency as the macro-cell power increases until it reaches the cut-set bound. This result implies

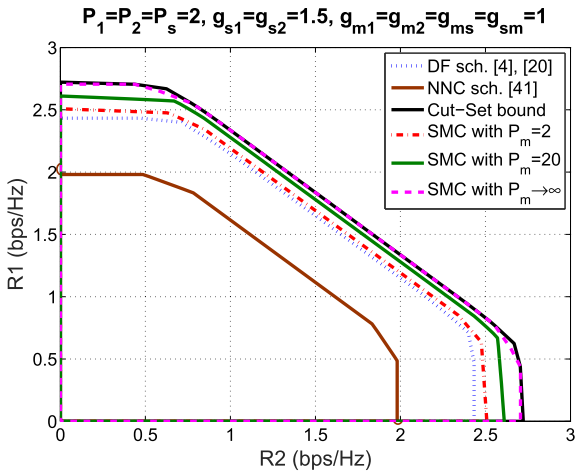

Fig. 7. Spectral efficiency regions for HetNet uplink transmission of the proposed scheme with SMC, DF relaying [4], [20] and NNC [41] schemes and the cut-set bound.

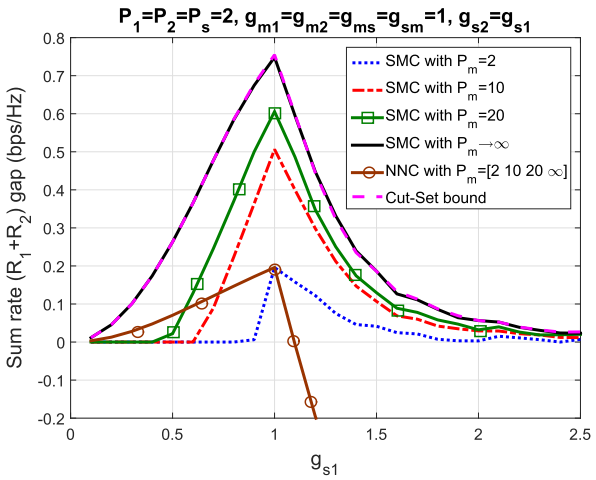

Fig. 9. Sum rate gaps of the MARC between the proposed SMC, NNC [41] and DF relaying [4], [20] schemes versus g_{s1} for the symmetric channel.

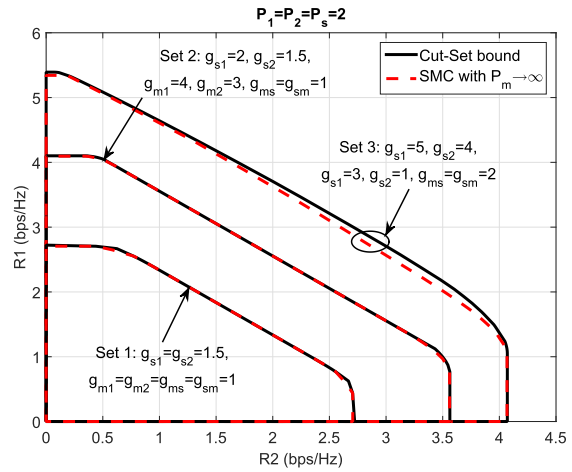

Fig. 8. Comparison between the spectral efficiency region of the proposed proposed scheme with SMC when $P_m \to \infty$ and the cut-set bound.

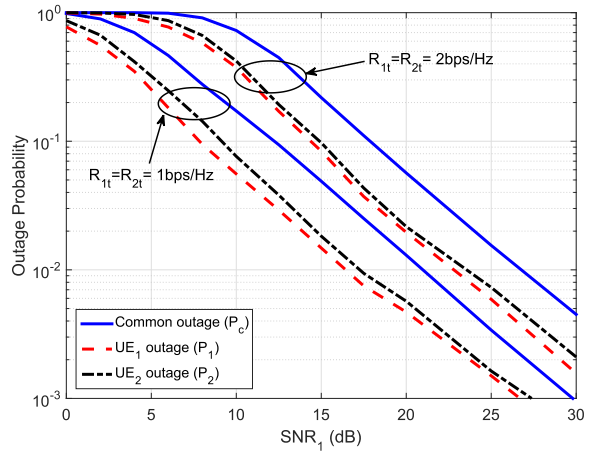

Fig. 10. Individual and common outage probabilities vs SNR1 for different target rates.

that UE1 and UE2 can improve their transmission rates not only by increasing their transmission powers but also by increasing the macro-cell power. Hence, in the uplink transmission for HetNet where the macro-cell is more powerful than UEs, the macro-cell can change its transmission power to adapt the spectral efficiency of both UEs based on a specific service requirement.

B. Asymptotic Capacity

Fig. 8 compares between the cut-set bound and the asymptotic spectral efficiency region at $P_m \to \infty$ for different sets of channel configurations. Results confirm Theorem 3 where the proposed scheme achieves capacity when $\frac{g_{m1}}{g_{s1}} = \frac{g_{m2}}{g_{s2}}$ as in sets 1 and 2 of channel parameters in Fig. 8. However, in set 3, the cut-set bound has a higher sum rate (throughput) than the proposed scheme. Moreover, for any channel configurations, the individual rates of the proposed scheme achieve the cut-set bound individual rates, which is also proved in [32].

Fig. 9 shows the sum rate gap between the proposed SMC scheme and the DF relaying scheme without SMC [4], [20] for symmetric channel. The rate improvement increases with P_m until it achieves the cut-set bound, and the maximum gap occurs when UE-small and UE-macro links have the same strength ($g_{s1} = g_{m1}$). This is because the DF relaying scheme [4], [20] starts improving the rate region over classical MAC (without the small-cell) when ($g_{s1} > g_{m1}$) and hence, reducing the gap with SMC scheme. Although NNC has the smallest spectral efficiency region for the channel configuration in Fig. 7, Fig. 9 shows that it can achieve a larger sum rate than the SMC scheme when g_{s1} is weaker than g_{m1}.

C. Outage Performance

We now provide numerical results for the outage probabilities formulated in Theorem 4. In these simulations, both UEs and the small-cell have the same power $P_1 = P_2 = P_s = P$ while the macro-cell power (P_m) is specified in each figure. Considering the small-cell radius of 20 to 200 meters [1], the inter-node distances (in meters) are $d_{r1} = d_{d1} = 14$,

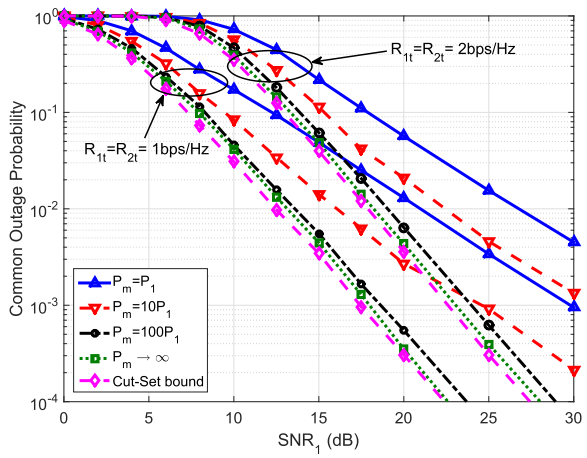

Fig. 11. Cut-set bound and common outage probabilities at $R_{1t} = R_{2t} = 2$ bps/Hz and different macro-cell powers.

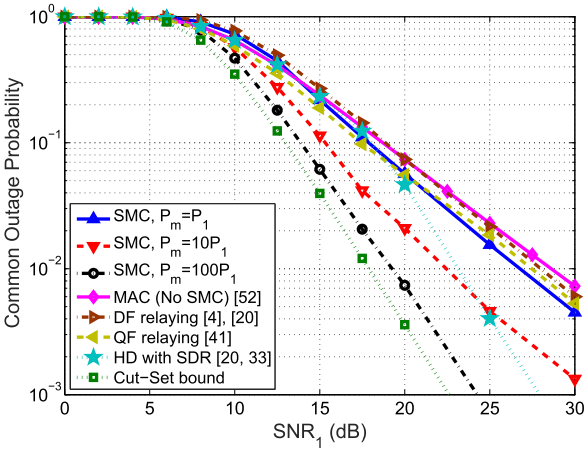

Fig. 12. Comparison between common outage probabilities of the proposed and existing schemes at $R_{1t} = R_{2t} = 2$ bps/Hz.

$d_{r2} = d_{d2} = 16$, and $d_{dr} = 18$. All the links are Rayleigh fading, and the average gain for each link from node j to i is given as $\mu_{ij} = \frac{1}{dij^{\alpha}}$ with path-loss factor $\alpha = 2.4$ [1]. For each fading channel, all the simulations are obtained using 10^6 samples, which are sufficient for outage probabilities above 10^{-4} and produce the same results over different simulation runs. With these settings, we define the average SNR at the macro-cell for signals received from UE1 (SNR1) and UE2 (SNR2) as follows:

$$\text{SNR1} = 10\log\left(\frac{P}{d_{d1}^{\alpha}}\right),$$
$$\text{SNR2} = 10\log\left(\frac{P}{d_{d2}^{\alpha}}\right) = \text{SNR1} + 10\log\left(\frac{d_{d1}^{\alpha}}{d_{d2}^{\alpha}}\right). \quad (33)$$

The average received SNRs at the small-cell are defined similarly. To minimize the outage, the simulations vary the power parameters at both UEs and the small-cell $(\rho_{n1}, \rho_{o1}, \rho_{n2}, \rho_{o2}, \rho_{s1}, \rho_{s2})$.

Fig. 10 shows the common and individual outage probabilities versus SNR1 for the proposed scheme with SMC where $P_m = P_1$. As expected, the common outage is higher than the individual outage and, at equal transmission rates, the individual outage for UE2 is higher than UE1 since UE2 has a weaker link to the small and macro cells.

In Fig. 10, the diversity order is one, as the small-cell always decodes both UEs' information and the quantized signal from the macro-cell is noisy as $P_m = P_s$. However, as P_m increases, the macro-cell quantizes its received signal with less quantization noise and sends a clear version of it to the small-cell as explained in Remark 8. Then, considering the common outage probability at different macro-cell powers, Fig. 11 shows that the proposed scheme starts achieving a full diversity order of two as P_m increases.

This full diversity order is achieved without requiring UEs to know their links to the small-cell as in [33] and [35] or even the order between their links to the small and macro cells as in [36]–[38]. Hence, the proposed scheme has a leaner design than existing schemes since it requires less transmission of non-information signals as specified in [2].

Fig. 11 further shows that even with $P_m \to \infty$, the cut-set bound outperforms the proposed scheme in contrast to the AWGN. This is because for the considered channel setting, statistically $E[g_{m1}g_{s2}] = E[g_{m2}g_{s1}]$ but they are different at each channel realization.

Figs. 7, 9 and 11 imply that the macro-cell can determine the uplink transmission performance by varying its power. Since the transmit power affects the cost of a service, this characteristic is very appealing for 5G networks, since with the same UE, different customers can have different service quality based on their paying plans.

Fig. 12 compares the common outage probabilities of the proposed SMC with existing DF [20], [33] and quantize-forward [41] relaying schemes and classical MAC (without the small-cell) [52]. Generally, as P_m increases, the proposed scheme outperforms all existing schemes and gets closer to the cut-set bound. When $P_m = P_1$, at high SNR, the proposed scheme outperforms the DF [20], QF [41] and classical MAC [52]. This is because the macro-cell receives better signals for UEs at high SNR than low SNR. Consequently, the macro-cell sends a better quantized signal to the small-cell to improve its decoding and reduces its outage probability. In contrast, at low SNR, the outage probability obtained from the decoding at the small-cell reduces outage performance. Hence, the QF relaying [41] and classical MAC [52] outperforms the proposed SMC and DF [20] relaying schemes.

Fig. 12 also considers the combined scheme of DF relaying [20] and successful DF relaying process [33] where the small-cell only forwards the correctly decoded message(s). This combined scheme can be applied to the HD channel of two phases where both UEs transmit their messages in phase 1 to the small and macro cells. Then, in phase 2, the small-cell forwards the successfully decoded message(s) coherently with its (their) corresponding UE(s). This scheme achieves a diversity order of two as shown in Fig. 12 but has lower performance than the proposed scheme as P_m increases.

IX. Conclusion

We have proposed a new HetNet uplink transmission scheme with joint small-macro cell cooperation (SMC).

The scheme uses superposition block Markov encoding at UEs, DF relaying with joint sliding window decoding at the small-cell, QF relaying and backward decoding at the macro-cell. We derived the spectral efficiency of this scheme and the optimal and sufficient quantization at the macro-cell for maximum spectral efficiency. We proved that when the ratios of UE-macro to UE-small cell links for both UEs are equal, the proposed scheme asymptotically achieves capacity by reaching the cut-set bound as the macro-cell power approaches infinity. We further formulated the outage probability, taking into account the outage events at the small and macro cells and the channel variation over different transmission blocks. We also generalized the proposed scheme to the N-UE scenario. Generally, results show that as the macro-cell power increases, spectral efficiency and reliability improve and outperform the existing schemes for HetNet uplink transmissions. These results hold 5G promise since, for each service having different spectral efficiency and reliability requirements, performance can be adjusted through the macro-cell with the same power from the UEs. Hence, this work encourages further analysis for the proposed scheme over cellular networks, including optimal power allocation from UEs and the small-cell, multi-antenna system performance and the impact of SMC on managing interference.

APPENDIX I
PROOF OF THEOREM 1

We prove Theorem 1 using the information theoretic analysis of a discrete memoryless MARC-RDC[5] (e.g. uplink HetNet with SMC) specified by a collection of pmf $p(y_m, y_s|x_1, x_2, x_s, x_m)$, where $x_i,\ i \in \{1, 2, s, m\}$ is the input signal of node i while $y_l,\ l \in \{s, m\}$ is the output signal of node l. We define a $(2^{nR_1}, 2^{nR_2}, n, Pe)$ code based on standard definitions as in [40].

We consider B independent transmission blocks each of length n. Two sequences of $B-2$ messages $w_{1,k}$ and $w_{2,k}$ for $k \in [1 : B-2]$ are to be sent over the channel in nB transmissions. Therefore, UE1 and UE2 do not send new information in the last two blocks, which reduces the achievable rates in (8) by a factor of $2/B$ (e.g. $R_1(1-2/B) \leq \min\{I_1, I_2, I_3\}$). This factor, however, becomes negligible as $B \to \infty$.

A. Codebook Generation

The codebook generation of the proposed coding scheme can be explained as follows. Fix the pmf $P^{\dagger} = p(u_1)p(x_1|u_1)p(u_2)p(x_2|u_2)p(x_m)p(\hat{y}_m|x_m)$. Then, for block $k \in \{1 : B\}$ and according to $P^{\dagger}$, randomly and independently generate 2^{nR_μ} codewords $u_\mu^n(w_{\mu,k-2})$ and 2^{nR_μ} codewords $x_\mu^n(w_{\mu,k}|w_{\mu,k-2})$ that encode $w_{\mu,k-2}$ and $w_{\mu,k}$, respectively where $\mu \in \{1, 2\}$. For each pair $(u_1^n(w_{1,k-2}), u_2^n(w_{2,k-2}))$, generate one sequences $X_s^n(w_{1,k-2}, w_{2,k-2})$. Similarly generate 2^{nR_m} codewords $x_m^n(l_{k-1})$ and 2^{nR_m} codewords $\hat{y}_m^n(l_k|l_{k-1})$ that encode the quantization indices l_{k-1} and l_k, respectively, where R_m is the transmission rate of the quantization index by the macro-cell.

[5]The channel is memoryless in the sense that given the current transmitted signals $(X_{1,k}, X_{2,k}, X_{s,k}, X_{m,k})$, the current received signals $(Y_{s,k}, Y_{m,k})$ are conditionally independent from the messages and the previous signals $(w_{1,k}, w_{2,k}, X_{1,j}, X_{2,j}, X_{s,j}, X_{m,j}, Y_{s,j}, Y_{m,j})$ for all $j \in \{1, 2, ..., k-1\}$ [40].

B. Encoding

Let $(w_{1,k}, w_{2,k})$ be the new messages to be sent in block k. Then, UE1 (UE2) transmits $x_1^n(w_{1,k}, w_{1,k-2})$ $(x_2^n(w_{2,k}, w_{2,k-2}))$.[6] The small (macro) cell has estimated $(\hat{w}_{1,k-2}, \hat{w}_{2,k-2})$ $(\hat{l}_{k-1})$ in block $k-1$. Then, the small (macro) cell transmits $x_s^n(\hat{w}_{1,k-2}, \hat{w}_{2,k-2})$ $(x_m^n(\hat{l}_{k-1}))$ in block k. Moreover, let $\hat{l}_{k-1} = L_{k-1}$ be correct quantization index estimated in block $k-1$ and sent in block k by the macro-cell, the macro-cell finds an index l_k in block k such that

$$\left(\hat{y}_m^n(l_k|L_{k-1}), x_m^n(L_{k-1}), y_m^n(k)\right) \in A_\epsilon^n, \tag{34}$$

where A_ϵ^n is a standard symbol for the set of jointly typical sequences of length n with respect to a joint distribution $p(x_m, y_m, \hat{y}_m)$ where the empirical entropies are ϵ-close to the true entropies (ϵ is a very small value, $\epsilon \to 0$) [53] . By covering lemma [40], such l_i exists if $R_m > I(\hat{Y}_m; Y_m|X_m)$.

C. Decoding

Without loss of generality, assume that all transmitted messages and the quantization indices are equal to 1. Then, the decoding can be described as follows.

1) At the Small-Cell: At the end of block $k+1$, the small-cell already knows $l_{k-1} = L_{k-1}$, $(w_{1,k-2}, w_{2,k-2}) = (1, 1)$ and $(w_{1,k-1}, w_{2,k-1}) = (1, 1)$ from the decoding in blocks $k-1$ and k. It then utilizes the signals received in blocks k and $k+1$ to find a unique triple $(\hat{w}_{1,k}, \hat{w}_{2,k}, \hat{l}_k)$ such that

$$\begin{aligned}&\left(x_1^n(\hat{w}_{1,k}, 1), u_1^n(1), x_2^n(\hat{w}_{2,k}, 1), u_2^n(1),\right.\\ &\qquad \left. x_m^n(L_{k-1}), \hat{y}_m^n(\hat{l}_k|L_{k-1}), y_s^n(k)\right) \in A_\epsilon^n\\ &\text{and } \left(u_1^n(1), u_2^n(1), x_m^n(\hat{l}_k) y_s^n(k+1)\right) \in A_\epsilon^n\end{aligned}$$

where A_ϵ^n is defined in similar way to that in (34). JT analysis [40] leads to the following rate constraints:

$$\begin{aligned}R_d &\leq \zeta_1 + \zeta_2,\\ R_1 &\leq I(X_1; \hat{Y}_m, Y_s|X_2, U_1, U_2, X_s, X_m),\\ R_1 + R_d &\leq \zeta_1 + \zeta_2 + I(X_1; Y_s|X_2, X_s, X_m, U_1, U_2),\\ R_2 &\leq I(X_2; \hat{Y}_m, Y_s|X_1, U_1, U_2, X_s, X_m),\\ R_2 + R_d &\leq \zeta_1 + \zeta_2 + I(X_2; Y_s|X_1, X_s, X_m, U_1, U_2),\\ R_1 + R_2 &\leq I(X_1, X_2; \hat{Y}_m, Y_s|U_1, U_2, X_s, X_m),\\ R_1 + R_2 + R_d &\leq I(X_1, X_2, X_m; Y_s|U_1, U_2, X_s) + \zeta_2,\end{aligned} \tag{35}$$

where $\zeta_1 = I(X_m; Y_s|U_1, U_2, X_s)$ and $\zeta_2 = I(\hat{Y}_m; X_1, U_1, X_2, U_2, X_s, Y_s|X_m)$. By combining $R_m > I(\hat{Y}_m; Y_m|X_m)$ and (35) and applying them to the Gaussian channel in (1) with the signaling in (2), we obtain the condition in (9) and the constraints $\acute{I}_1$, $\acute{I}_2$, $\acute{I}_4$, $\acute{I}_5$, $\acute{I}_7$ and $\acute{I}_8$ in (4).

[6]By convention, we assume that the old (new) information sent in blocks 1 and 2 ($B-1$ and B) are 1 ($w_{i,-1} = w_{i,0} = w_{i,B-1} = w_{i,B} = 1$ for $i \in \{1, 2\}$) [40, Ch. 16].

2) At the Macro-Cell: It employs backward decoding where in block k, it has already estimated $\tilde{w}_{1,k}$ and $\tilde{w}_{2,k}$ from the decoding in block $k+2$. Then, it looks for a unique pair $(\tilde{w}_{1,k-2}, \tilde{w}_{2,k-2})$ such that

$$\big(x_1^n(1, \tilde{w}_{1,k-2}), x_2^n(1, \tilde{w}_{2,k-2}), u_1^n(\tilde{w}_{1,k-2}), u_2^n(\tilde{w}_{2,k-2}), x_s^n(\tilde{w}_{1,k-2}, \tilde{w}_{2,k-2}), x_m^n(L_{k-1}), y_m^n(k)\big) \in A_\epsilon^n$$

JT analysis [40] leads to the following rate constraints:

$$\begin{aligned} R_1 &\le I(X_1, X_s; Y_m|X_m, X_2, U_2), \\ R_2 &\le I(X_2, X_s; Y_m|X_m, X_1, U_1), \\ R_1 + R_2 &\le I(X_1, X_2, X_s; Y_m|X_m), \end{aligned} \tag{36}$$

Applying these constraints to the Gaussian channel in (1), we obtain the constraints $\acute{I}_3$, $\acute{I}_6$, and $\acute{I}_9$ in (7).

APPENDIX II
PROOF OF THEOREM 2

Theorem 2 is proved by showing that for any quantization with variance $Q_T \notin \{Q_1^*, Q_2^*, Q_s^*\}$, the rate region at the small-cell obtained with Q_T is included inside the time-shared region between the regions of Q_1^*, Q_2^* and Q_s^*. The proof is obtained with help from Fig. 3 and the fact that I_1, I_4 and I_7 (I_2, I_5 and I_8) are decreasing (increasing) functions of Q.

A. For $Q_T < \min\{Q_1^*, Q_2^*, Q_s^*\}$ or $Q_T > \max\{Q_1^*, Q_2^*, Q_s^*\}$

First, assume that $Q_s^* = \min\{Q_1^*, Q_2^*, Q_s^*\}$. Then, for $Q_T < Q_s^*$, the rate constraints at the small-cell are given as follows:

$$\begin{aligned} R_1 &\le \min\{I_1(Q_T), I_2(Q_T)\} \overset{a}{=} I_2(Q_T), \\ R_2 &\le \min\{I_4(Q_T), I_5(Q_T)\} \overset{b}{=} I_5(Q_T), \\ R_1 + R_2 &\le \min\{I_7(Q_T), I_8(Q_T)\} \overset{c}{=} I_8(Q_T) \end{aligned} \tag{37}$$

where $I_l(x)$ for $l \in \{1, 2, 4, 5, 7, 8\}$ is the rate constraint in (8) with $Q = x$. In 37, (a) follows since $I_2(Q_T) < I_1(Q_T)$ as $Q_T < Q_s^* < Q_1^*$ and $I_2 < I_1$ for any $Q < Q_1^*$ since I_2 is an increasing function of Q. A similar analysis holds for (b) and (c).

Second, since $Q_T < Q_s^*$, $I_i(Q_T) < I_i(Q_s^*)$ for $i \in \{2, 5, 8\}$. Therefore, the rate region obtained with Q_T is included inside that obtained with Q_s^*.

Third, by performing a similar analysis with $Q_s^* = \max\{Q_1^*, Q_2^*, Q_s^*\}$, the rate constraints at the small-cell for $Q_T > Q_s^*$ are $R_1 < I_1(Q_T)$, $R_2 < I_4(Q_T)$, and $R_1 + R_2 < I_7(Q_T)$. Then, since $I_1 < I_2$, $I_4 < I_5$ and $I_7 < I_8$ for any $Q > Q_s^*$ since I_1, I_4, I_7 are decreasing functions of Q, the rate region obtained with Q_T is included inside that obtained with Q_s^*.

Last, a similar analysis holds when $Q_i^* = \min\{Q_1^*, Q_2^*, Q_s^*\}$ or $\max\{Q_1^*, Q_2^*, Q_s^*\}$ for $i \in \{1, 2\}$.

B. For $\min\{Q_1^*, Q_2^*, Q_s^*\} < Q_T < \max\{Q_1^*, Q_2^*, Q_s^*\}$

This case is proved with help from Fig. 3. Define SL_1 and SL_2 as the slopes of AB and CB lines, respectively. Then, we need to show when $|SL_1| >$ or $< |SL_2|$ where $|x|$ is the absolute value of x. If $|SL_1| < |SL_2|$, point C in Fig. 3 is outside the time-shared line (dashed black line).

First, for the two rate regions obtained with Q_1^* and Q_s^*, assume that $Q_1^* < Q_T < Q_s^*$. Then, at Q_T, $I_1(Q_T) < I_2(Q_T)$ since I_1 is a decreasing function of Q. Similarly, $I_8(Q_T) < I_7(Q_T)$ since I_8 is an increasing function of Q. Therefore, the coordinates of points A, B and C are

$$\begin{aligned} (R_1(A), R_2(A)) &= (I_1(Q_1^*), I_8(Q_1^*) - I_1(Q_1^*)), \\ (R_1(B), R_2(B)) &= (I_1(Q_s^*), I_8(Q_s^*) - I_1(Q_s^*)), \\ (R_1(C), R_2(C)) &= (I_1(Q_T), I_8(Q_T) - I_1(Q_T)), \end{aligned} \tag{38}$$

Second, point C in Figure 3 is outside the time-shared region for the range of Q_T that satisfies

$$\begin{aligned} &|SL_1| < |SL_2| \Leftrightarrow \frac{R_1(A) - R_1(B)}{R_2(B) - R_2(A)} < \frac{R_1(C) - R_1(B)}{R_2(B) - R_2(C)}, \\ &\Leftrightarrow R_1(A)(R_2(B) - R_2(C)) + R_1(B)R_2(C) \\ &\quad < R_1(C)(R_2(B) - R_2(A)) + R_1(B)R_2(A), \\ &\overset{a}{\Leftrightarrow} I_1(Q_s^*)\left[I_8(Q_T) - I_8(Q_1^*)\right] + I_1(Q_1^*)\left[I_8(Q_s^*) - I_8(Q_T)\right] \\ &\quad < I_1(Q_T)\left[I_8(Q_s^*) - I_8(Q_1^*)\right], \\ &\Leftrightarrow I_1(Q_s^*)\frac{I_8(Q_T) - I_8(Q_1^*)}{I_8(Q_s^*) - I_8(Q_1^*)} + I_1(Q_1^*)\frac{I_8(Q_s^*) - I_8(Q_T)}{I_8(Q_s^*) - I_8(Q_1^*)} \\ &\quad < I_1(Q_T), \\ &\Leftrightarrow I_1(Q_s^*)\delta_{18} + I_1(Q_1^*)(1 - \delta_{18}) - I_1(Q_T) < 0, \\ &\Leftrightarrow f_{1,1,8(Q_T)} < 0, \end{aligned} \tag{39}$$

where (a) is obtained by substituting (38) into (39) and some mathematical manipulations; $\delta_{T,18} = \frac{I_8(Q_T) - I_8(Q_1^*)}{I_8(Q_s^*) - I_8(Q_1^*)}$ and $0 < \delta_{T,18} < 1$ since I_8 is an increasing function of Q and $Q_1^* < Q_T < Q_s^*$.

Third, since I_1 is a decreasing function of Q, the formula (39) is satisfied for $Q_T \in (Q_1^*, Q_1^{1T})$ where Q_1^{1T} is obtained form solving $f_{1,1,8(Q_T)} = 0$.

Fourth, when $Q_s^* < Q_T < Q_1^*$, a similar analysis holds where the coordinates of points A, B and C are similar to those in (38) but replacing I_1 by I_2 and I_8 by I_7 since $I_2 < I_1$ for $Q_T < Q_1^*$ and $I_7 < I_8$ $Q_T > Q_s^*$. Then, considering the range of Q_T that satisfies $|SL_1| < |SL_2|$, we obtain

$$\begin{aligned} &I_2(Q_s^*)\delta_{27} + I_2(Q_1^*)(1 - \delta_{27}) - I_2(Q_T) < 0, \\ &\qquad \Leftrightarrow f_{2,1,7(Q_T)} < 0, \end{aligned} \tag{40}$$

where $\delta_{T,17} = \frac{I_7(Q_T) - I_7(Q_1^*)}{I_7(Q_s^*) - I_7(Q_1^*)}$. Then, since I_2 is an increasing function of Q, the formula (40) is satisfied for $Q_T \in (Q_1^{2T}, Q_1^*)$ where Q_1^{2T} is obtained from solving $f_{2,1,7(Q_T)} = 0$.

Last, a similar analysis holds for the two rate regions obtained with Q_2^* and Q_s^* when $Q_2^* < Q_T < Q_s^*$ and $Q_s^* < Q_T < Q_2^*$. Then, Theorem 2 is obtained by combining these optimal ranges for Q_T.

APPENDIX III
PROOF OF COROLLARY 3

Using the cut-set bound [40], the capacity for MARC-RDC (HetNet with SMC) is upper-bounded by

the maximum information flowing through the 6 cutsets in Fig. 4, which are expressed as follows:

$$\begin{aligned}
R_1 &\leq \min\{I(X_1; Y_s, Y_m|X_s, X_m, X_2), I(X_1, X_s; Y_m|X_m, X_2)\},\\
R_2 &\leq \min\{I(X_2; Y_s, Y_m|X_s, X_m, X_1), I(X_2, X_s; Y_m|X_m, X_1)\},\\
R_1 + R_2 &\leq \min\big\{I(X_1, X_2; Y_s, Y_m|X_s, X_m),\\
&\qquad I(X_1, X_2, X_s; Y_m|X_m)\big\},
\end{aligned} \tag{41}$$

for some joint distribution $p(x_1, x_2, x_s, x_m)$.

The optimal input distribution that maximizes the rate region in (41) is jointly Gaussian for the Gaussian channel in (1). This is because the distribution of (Z_1, Z_2, Z_s, Z_m) is $\mathcal{CN}(0, I_{4\times4})$ and by using entropy power inequality [40], it is easy to show that the rate region in (41) is maximized if (X_1, X_2, X_s, X_m) is jointly Gaussian, i.e., $(X_1, X_2, X_s, X_m) \sim \mathcal{N}(0, \Sigma)$ where Σ is the covariance matrix given as

$$\Sigma = \mathrm{cov}(X_1, X_2, X_s, X_m) = \begin{bmatrix} P_1 & 0 & \delta_{1s}\sqrt{P_1P_s} & \delta_{1m}\sqrt{P_1P_m} \\ 0 & P_2 & \delta_{2s}\sqrt{P_2P_s} & \delta_{2m}\sqrt{P_2P_m} \\ \delta_{1s}\sqrt{P_1P_s} & \delta_{2s}\sqrt{P_2P_s} & P_s & \delta_{sm}\sqrt{P_sP_m} \\ \delta_{1m}\sqrt{P_1P_m} & \delta_{2m}\sqrt{P_2P_m} & \delta_{sm}\sqrt{P_sP_m} & P_m \end{bmatrix} \tag{42}$$

where δ_{ji} is the correlation factor between X_j and X_i, $(\delta_{1s}, \delta_{1m}, \delta_{2s}, \delta_{2m}, \delta_{sm}) \in [-1, +1]$ and $\delta_{12} = 0$ since X_1 and X_2 are independent. The determinant $\det(\Sigma) \geq 0$ such that Σ is positive semi-definite and a valid covariance matrix. By applying the rate constraints in (41) to the Gaussian channel in (1) with Σ in (42), we obtain the cut-set bound in (15).

REFERENCES

[1] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos. (Feb. 2015). "A survey of millimeter wave (mmWave) communications for 5G: Opportunities and challenges." [Online]. Available: https://arxiv.org/abs/1502.07228

[2] "5G radio access: Technology and capabilities," Ericsson, Stockholm, Sweden, White Paper, Feb. 2015.

[3] S. Obeso, J. Luo, R. Halfmann, E. Schulz, and C. Hartman, "Intra-operator inter-mode spectrum sharing," in *Proc. IEEE 69th Veh. Technol. Conf. (VTC)*, Apr. 2009, pp. 1–5.

[4] G. Kramer and A. J. van Wijngaarden, "On the white Gaussian multiple-access relay channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Sep. 2000, p. 40.

[5] J. Korhonen, *Introduction to 4G Mobile Communications*. Norwood, MA, USA: Artech House, 2014.

[6] Z. Zhang, K. Long, A. V. Vasilakos, and L. Hanzo, "Full-duplex wireless communications: Challenges, solutions, and future research directions," *Proc. IEEE*, vol. 104, no. 7, pp. 1369–1409, Jul. 2016.

[7] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.

[8] Y.-S. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 5992–6010, Dec. 2013.

[9] D. Bharadia, E. Mcmilin, and S. Katti, "Full duplex radios," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 375–386, Oct. 2013.

[10] C. Galiotto, N. Marchetti, and L. Doyle, "Flexible spectrum sharing and interference coordination for low power nodes in heterogeneous networks," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, Sep. 2012, pp. 1–5.

[11] X. Chai, Y. Li, Y. Lv, and Z. Zhang, "Joint spectrum-sharing and base-station-sleep model for improving energy efficiency of HetNet," in *Proc. IEEE ICC*, Jun. 2015, pp. 1851–1856.

[12] Y. Lin, W. Bao, W. Yu, and B. Liang, "Optimizing user association and spectrum allocation in HetNets: A utility perspective," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1025–1039, Jun. 2015.

[13] M. Hong, R. Sun, H. Baligh, and Z. Q. Luo, "Joint base station clustering and beamformer design for partial coordinated transmission in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 226–240, Feb. 2013.

[14] G. Nigam, P. Minero, and M. Haenggi, "Coordinated multipoint joint transmission in heterogeneous networks," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4134–4146, Nov. 2014.

[15] R. Singh and C. S. R. Murthy, "Techniques for interference mitigation using cooperative resource partitioning in multitier LTE HetNets," *IEEE Syst. J.*, to be published.

[16] S. Iwelski, E. Majeed, Z. Bai, G. H. Bruck, and P. Jung, "Cooperative interference mitigation in heterogeneous LTE networks," in *Proc. IEEE WCNC*, Mar. 2015, pp. 135–140.

[17] C. Shen, J. Xu, and M. Van der Schaar, "Silence is gold: Strategic interference mitigation using tokens in heterogeneous small cell networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 6, pp. 1097–1111, Jun. 2015.

[18] A. Ghosh, L. Cottatellucci, and E. Altman, "Nash equilibrium for femtocell power allocation in HetNets with channel uncertainty," in *Proc. IEEE GLOBECOM*, Dec. 2015, pp. 1–7.

[19] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[20] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005.

[21] M. Osmani-Bojd and G. Hodtani, "On multiple-access relay channel with common message," *Trans. Emerg. Telecommun. Technol.*, vol. 27, no. 8, pp. 1030–1043, Aug. 2016, doi: 10.1002/ett.2821.

[22] A. Sahebalam and G. A. Hodtani, "General and new inner bound for multiple-access relay channel and two certain capacity theorems," *IET Commun.*, vol. 7, no. 13, pp. 1348–1359, Sep. 2013.

[23] G. Zeitler, R. Koetter, G. Bauch, and J. Widmer, "An adaptive compress-and-forward scheme for the orthogonal multiple-access relay channel," in *Proc. IEEE 20th Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2009, pp. 1838–1842.

[24] M. Lei and M. R. Soleymani, "Diversity–multiplexing tradeoff of the half-duplex slow fading multiple-access relay channel based on generalized quantize-and-forward scheme," *IEEE Wireless Commun. Lett.*, vol. 4, no. 1, pp. 74–77, Feb. 2015.

[25] G. Dietl, M. Sciora, G. Zeitler, G. Bauch, and J. Widmer, "A quantize-and-forward scheme for future wireless relay networks," in *Proc. IEEE VTC*, Sep. 2011, pp. 1–4.

[26] L. Liu, Y. Li, Y. Su, and Y. Sun, "Quantize-and-forward strategy for interleave-division multiple-access relay channel," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1808–1814, Mar. 2016.

[27] A. Winkelbauer, N. Goertz, and G. Matz, "Compress-and-forward in the multiple-access relay channel: With or without network coding?" in *Proc. 7th Int. Symp. Turbo Codes Iterative Inf. Process. (ISTC)*, Aug. 2012, pp. 131–135.

[28] L. Sankar, N. B. Mandayam, and H. V. Poor, "On the sum-capacity of degraded Gaussian multiple-access relay channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5394–5411, Dec. 2009.

[29] A. Abu Al Haija and M. Vu, "Efficient use of joint source-destination cooperation in the Gaussian multiple access channel," in *Proc. IEEE ICC*, Jun. 2013, pp. 3241–3246.

[30] C. Suh and D. N. C. Tse, "Feedback capacity of the Gaussian interference channel to within 2 bits," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2667–2685, May 2011.

[31] J. Jiang, Y. Xin, and H. V. Poor, "Achievable rates for discrete memoryless relay channels with generalised feedback," *Trans. Emerg. Telecommun. Technol.*, vol. 24, no. 2, pp. 212–231, Mar. 2013.

[32] A. Abu Al Haija and M. Vu, "An asymptotically capacity-achieving scheme for the Gaussian relay channel with relay-destination cooperation," in *Proc. 47th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2013, pp. 1–6.

[33] T. Elkourdi and O. Simeone, "Femtocell as a relay: An outage analysis," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4204–4213, Dec. 2011.

[34] P.-S. Lu, X. Zhou, and T. Matsumoto, "Outage probabilities of orthogonal multiple-access relaying techniques with imperfect source-relay links," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2269–2280, Apr. 2015.

[35] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
[36] A. Abu Al Haija and M. Vu, "Outage analysis for coherent decode-forward Relaying over rayleigh fading channels," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1162–1177, Apr. 2015.
[37] L. Pinals, A. Abu Al Haija, and M. Vu, "Link regime and power savings of decode-forward relaying in fading channels," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 931–946, Mar. 2016.
[38] A. Abu Al Haija and M. Vu, "Spectral efficiency and outage performance for hybrid D2D-infrastructure uplink cooperation," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1183–1198, Mar. 2015.
[39] A. Abu Al Haija and C. Tellambura, "An asymptotically capacity-achieving scheme for the Gaussian multiple-access relay channel," in *Proc. IEEE ICC*, May 2016, pp. 1–6.
[40] A. El Gamal and Y.-H. Kim, *Network Information Theory*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[41] S. H. Lim, Y.-H. Kim, A. El Gamal, and S.-Y. Chung, "Noisy network coding," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3132–3152, May 2011.
[42] R. Mudumbai, D. R. B. Iii, U. Madhow, and H. V. Poor, "Distributed transmit beamforming: Challenges and recent progress," *IEEE Commun. Mag.*, vol. 47, no. 2, pp. 102–110, Feb. 2009.
[43] M. H. M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 5, pp. 439–441, May 1983.
[44] D. Gesbert, S. Hanly, H. Huang, S. S. Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.
[45] K. Alexandris, A. Balatsoukas-Stimming, and A. Burg, "Measurement-based characterization of residual self-interference on a full-duplex MIMO testbed," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop (SAM)*, Jun. 2014, pp. 329–332.
[46] I.-H. Wang and D. N. C. Tse, "Interference mitigation through limited receiver cooperation," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2941–2965, May 2011.
[47] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
[48] A. Abu Al Haija and C. Tellambura, "Decoding delay and outage performance analysis of full-duplex decode-forward relaying: Backward or sliding window decoding," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4520–4533, Nov. 2016.
[49] P. Popovski, "Ultra-reliable communication in 5G wireless systems," in *Proc. IEEE Int. Conf. 5G Ubiquitous Connectivity*, Nov. 2014, pp. 146–151.
[50] *LTE Physical Layer Framework for Performance Verification*, 3rd Generation Partnership Project (3GPP), document TS R1-070674, Feb. 2007.
[51] A. Abu Al Haija, L. Pinals, and M. Vu, "Outage analysis and power savings for independent and coherent decode-forward relaying," in *Proc. IEEE Conf. Global Commun. (Globecom)*, Dec. 2015, pp. 1–6.
[52] R. Narasimhan, "Individual outage rate regions for fading multiple access channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jun. 2007, pp. 1571–1575.
[53] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.

Ahmad Abu Al Haija (S'10–M'15) received the B.Sc. and M.Sc. degrees from the Jordan University of Science and Technology, Irbid, Jordan, in 2006 and 2009, respectively, and the Ph.D. degree from McGill University, Montreal, QC, Canada, in 2015, all in electrical engineering. He is currently a Post-Doctoral Fellow with the Electrical and Computer Engineering Department, University of Alberta, Edmonton, AB, Canada. From 2013 to 2014, he was a Visiting Student with Tufts University, Medford, MA, USA. His research interest includes cooperation in multiuser channels, resource allocation for cooperative communication systems, wireless communications and performance analysis, and evaluation over fading channels.

Chintha Tellambura (F'11) received the B.Sc. degree (Hons.) from the University of Moratuwa, Sri Lanka, in 1986, the M.Sc. degree in electronics from the University of London, U.K., in 1988, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1993.

He was a Post-Doctoral Research Fellow with the University of Victoria from 1993 to 1994 and the University of Bradford from 1995 to 1996). He was with Monash University, Australia, from 1997 to 2002. He is currently a Professor with the Department of Electrical and Computer Engineering, University of Alberta. He has authored and coauthored over 490 journal and conference publications with total citations of over 11,000 and an h-index of 56. His current research interests include the design, modeling, and analysis of cognitive radio networks, heterogeneous cellular networks, and multiple-antenna wireless networks.

Prof. Tellambura and his co-authors received the Communication Theory Symposium best paper award in the 2012 IEEE International Conference on Communications, Ottawa, Canada. He is the winner of the prestigious McCalla Professorship and the Killam Annual Professorship from the University of Alberta. He served as an Editor for both the IEEE TRANSACTIONS ON COMMUNICATIONS and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and was the Area Editor for Wireless Communications Systems and Theory in the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS during 2007–2012.