QoS-Aware Downlink Cooperation for Cell-Edge and Handoff Users

Satyam Agarwal, *Student Member, IEEE*, Swades De, *Member, IEEE*, Satish Kumar, and Hari Mohan Gupta

***Abstract*—In this paper, we present a quality-of-service (QoS) aware cooperative downlink scheduling approach for cell-edge and handoff users that offers more reliability and a higher effective capacity. Cooperation (handoff) region is defined for active handoff users between two adjacent base stations (BSs) as a function of the user QoS requirements and network load. In addition, the proposed technique inherently acquires the inter-cell interference coordination by adjusting the position and size of the cooperation window suitably. Numerical results are presented showing reliability, user QoS and capacity gain performance, and region for cooperative scheduling in a coded communication scenario. Our analysis indicates that, cooperation provides a relatively less gain in effective capacity – up to about 40% with respect to non-cooperative handoff, when the QoS requirement is *loose*. On the other hand, when the QoS requirement is more *stringent*, the effective capacity gain can increase up to nearly 100%. Additionally we show that, while for applications with loose QoS requirement the cooperation window size is small, nearly 1% of the total area of the BSs participating in cooperation, it increases quite significantly, up to nearly 25%, for the applications with stringent QoS. Although the outage performance of the proposed approach is poorer than the joint transmission mode of cooperative multi-point scheme in lightly loaded networks, its effective capacity is significantly higher under varying network traffic load as well as QoS constraints.**

***Index Terms*—Cooperative handoff; cooperation region-of-interest; effective capacity; outage probability**

I. Introduction

The landmark for high data rate in 4th Generation (4G) wireless broadband access networks like LTE-A (Long Term Evolution-Advanced) and WiMAX-Mobile (World wide Interoperability for Microwave Access-Mobile) are set to be very high, around 1 Gbps in downlink and 300 Mbps in uplink as per the IMT-Advanced (International Mobile Telecommunications-Advanced) specifications [1], [2]. But these rates are not achievable at the cell-edge in current scenarios while maintaining the seamless mobility and quality of service (QoS). One of the major problems on the data rates at the cell-edge is imposed by inter-cell interference (ICI) (network effect). Dynamic sub-carrier and power allocation techniques are considered in [3] to mitigate the ICI and increase the data rate in orthogonal frequency division multiple access (OFDMA) based networks. Performance analysis of such techniques is presented in [4]. Fractional frequency reuse schemes [5] perform better at the cell-edge but limit the overall system capacity by not utilizing the full spectrum in each cell, i.e. without universal frequency reuse (UFR). Cooperative multi-point (CoMP) technique [6], [7] on the other hand is targeted to improve the cell edge user experience by creating virtual multiple-input multiple-output (MIMO) and utilizing beam-forming and precoding techniques.

The other problem on the cell edge users' performance is imposed by mobility of the users (user effect) called *handoff*. Handoff is performed on the basis of some threshold metric, which can be chosen as per the system requirements or the application constraints of an individual mobile station (MS) [8], [9]. The objective is to improve the system-wide performance as well as satisfy the user QoS requirements. Since the network resources are limited, a critical issue in inter-cell handoff is maintaining the user QoS. Many scheduling techniques have been proposed in the literature which take care of the user-level fairness as well as network capacity by taking the utility based approach [10]. However, maintaining fairness and capacity is more difficult at the cell edge when a MS decides to perform a handoff, or a macro-diversity approach is adopted to mitigate the disconnection time and loss of packets. Also, quantification of the network effect (ICI and co-channel interference) and user effect (mobility and user QoS) jointly to analyze the system performance at the cell-edge is a challenging task.

In this paper, we present *a new network-level inter-cell cooperation scheme in a UFR cellular network* (with frequency reuse factor $= 1$) *to increase the effective system capacity while improving the outage performance* at the cell-edge and coverage-overlap regions in a coded communication scenario. We provide an analytical model to capture the ICI and the user effect while evaluating the cell edge performance. The effective capacity concept in [11] for a single cell scenario is extended to the proposed network-level cooperation, and a simple macro-diversity model is considered to capture and optimize the user QoS performance. The key objectives and benefits of the proposed network-level cooperation scheme are as follows: 1) Outage probability analysis in a UFR, clustered ICI environment demonstrates significant improvement in outage performance in comparison with the conventional handoff and a competitive cooperative scheme (CoMP for single antenna receivers). 2) The effective capacity analysis – which jointly accounts the user QoS, network load, and physical layer bit error rate (BER) constraint or threshold signal to interference plus noise ratio (SINR) in optimum resource sharing – quantifies the capacity gain with respect

Preliminary finding of this work was presented at the IEEE WCNC 2012.

S. Agarwal, S. De, and H. M. Gupta are with the Department of Electrical Engineering and Bharti School of Telecommunication, IIT Delhi, New Delhi, India. S. Kumar was with the Bharti School of Telecommunication, IIT Delhi, New Delhi. Currently he is with Qualcomm India Pvt. Ltd., Hyderabad, India.
The work done by S. Kumar was during his academic time in IIT Delhi.

to the non-shared as well as compares the gain with respect to CoMP. 3) Multiple base station (BS) resource sharing is demonstrated to offer a quite high capacity gain at the cell edge and wide cooperation region, especially to the users with more stringent QoS. The impact of network load on capacity gain is shown to be not very significant, although the size and position of the gain window is dependent on network load.

In the next section related works are surveyed and the paper's contributions are highlighted. In Section III, network architecture and the proposed system model are presented. Section IV deals with the outage performance. In Section V, scheduling shared users and optimal resource allocation are analyzed. Section VI provides the numerical results on QoS and network load aware system capacity and cooperation performance. The paper is concluded in Section VII.

II. Related Work and Motivation

There is a rich literature on handoff ([8], [9], [12]–[14] and the references therein), which deal with the handoff parameter optimization, decision criterion, etc. But these studies do not deal with QoS and capacity related performance. These works use traditional approaches to optimize the system parameters, taking handoff as a discrete problem. As the next generation networks are evolving, the issues related to handoff are not treated individually as a physical layer problem. The protocol end points that are located in the BS will need to be moved from the source BS to the target BS. This relocation can be done in two different ways: 1) protocol status transfer; 2) protocol reinitialization after the handoff.

In LTE networks the relocation is done using protocol status transfer from source BS to target BS [15], [16]. A packet forwarding approach from the old BS to the new BS was also considered for the in-flight packets, whose impact on the user connection was studied by Bajzik et al. [16]. Their results showed that the extra load caused by the forwarding is not significant. The impact of forwarding on the end user connections can be reduced by using a scheduling architecture at the transport layer that is able to differentiate among different service classes while allocating the bandwidth. Tian et al. [17] proposed a seamless handoff scheme for high-speed trains via a dual-layer and dual-link architecture in LTE networks, where the MS (train) is equipped with two antennas – one at each end of its length. During the handoff, data transfer is continued from the serving BS via one antenna, while the other one is used for handoff processing with the target BS. To minimize the data loss, bi-casting to the two BSs is used with the data tunneling facility to the target BS.

At the cell edge ICI is a serious problem, particularly in OFDMA networks where a UFR plan is used. For the cell-edge users, the problems of ICI and handoff are considered together in the literature [18]. Some techniques to mitigate ICI (e.g., [18]) use ICI coordination (ICIC) among the BSs. To avoid radio link failure, ICIC can be used on top of the parameter optimization for handoff. It was shown that ICIC can overcome the radio link failure problem without affecting the handoff rates and with different selection of handoff parameters.

As suggested by Guedes and Yacoub [19], the cell overlap can extend up to 47% of the total area of two adjacent cells. Thus, while being in the coverage-overlap zone the signals from two or more BSs can be exploited [8], [9], [20]–[24] even if the MS is traveling at a high speed, because the coverage-overlap region is quite large. The physical layer cooperation studies by Lee et al. [22] and Chang et al. [23] further suggested that, in OFDM networks using one or two FFT modules in downlink, signals from two different BSs can improve the handoff performance in terms of cell-edge outage for high-speed multimedia services. The approach in [22] is called semi-soft handoff (SSHO), which uses a site selection transmit diversity (SSTD) for improved performance at the cell-edge. Another physical layer cooperation for downlink transmission at the cell-edge is the CoMP technique [6], which aims at interference control by using virtual MIMO concept, beam-forming downlink transmission through precoding. The coordinated scheduling and beam-forming from multiple transmit points require very precise synchronization, fast channel state feedback, and high backhaul signaling and processing overheads. As discussed in Garcia et al. [7], in the context of conventional receivers, multi-cell physical layer cooperation at the MS is equivalent to distributed multiple-input single-output (MISO) system. CoMP technology requirements were further discussed in [25] with respect to downlink performance in heterogeneous networks and in [26] regarding the 3G partnership project (3GPP) Release 11 [27] specification support.

Different physical layer cooperative schemes for downlink cell-edge users in a static environment was studied via simulation by Kumar et al. [28]. They suggested that, the rate supported to a cell-edge user with cooperation among BSs is not always better than the transmission from a single BS. In order to improve the performance of fractional frequency reuse networks, BS cooperation was proposed by Xu et al. [29] for cell-edge users to increase the system capacity. The authors showed that, adaptive reallocation of sub-channels can be used to further improve the performance. Their results also demonstrated that the spectral efficiency at the cell edge is higher for a high SNR at the cell edge, which is achieved by using cooperation as opposed to non-cooperation. On the other hand, the studies by Nosratinia et al. [30] and Laneman et al. [31] suggested that a user having a relatively bad channel will be benefited more by using cooperation, which is rather intuitive. So, the results presented in [29] are counter-intuitive with respect to the observations in [30] and [31]. We also note that the traffic load in the BSs and the individual user QoS should be considered while exploiting the signal from more than one BS, because the user QoS will define the need for multi-cell transmission, and the load on the BSs will decide the feasibility and limits of such transmissions.

To sum up, ICI and handoff are two critical problems for the cell-edge users. Cooperation among the BSs is a good way for the cell-edge users to mitigate ICI as well as radio link failure. However, *the prior inter-cell cooperation studies work at the physical layer which do not allow to jointly account for the user level parameters - such as QoS and threshold SINR, and the network level parameters - such as load distribution among the BSs.* To complement the prior studies, in this work we take a new approach, which we call *network-level cooperation*, to share the system resources from two or more BSs for

cell-edge user throughput and system capacity optimization. With respect to the studies in [22] and [23], we argue that the multi-BS transmission zone should be specified, which may depend on the user QoS, network load, and other user and system level parameters. To this end, we extend the outage performance to specify the boundaries of multi-BS transmissions, the gain in capacity, and the effect of user and system level parameters. First we present a typical clustered cellular network architecture for 4G networks to be consistent with the literature as considered by Nguyen and Sasase [32] and Mihovska et al. [33]. A new two-level queueing model is given to better utilize the resources and optimize the handoff performance of the cell-edge users in a UFR ICI environment. Unlike the studies by Lee et al. [22], Chang et al. [23], and Tian et al. [17], our queueing strategy does not require any packet forwarding approach from one BS to another for handoff users. In our approach we consider per-BS load and per-user QoS as the two parameters to decide whether and where to start resource sharing from more than one BS. We analytically capture the cell-edge effective capacity gain (user QoS benefit) and the corresponding optimum sharing zone as a function of network load and user QoS in which the cell-edge user can be scheduled from two or more BSs.

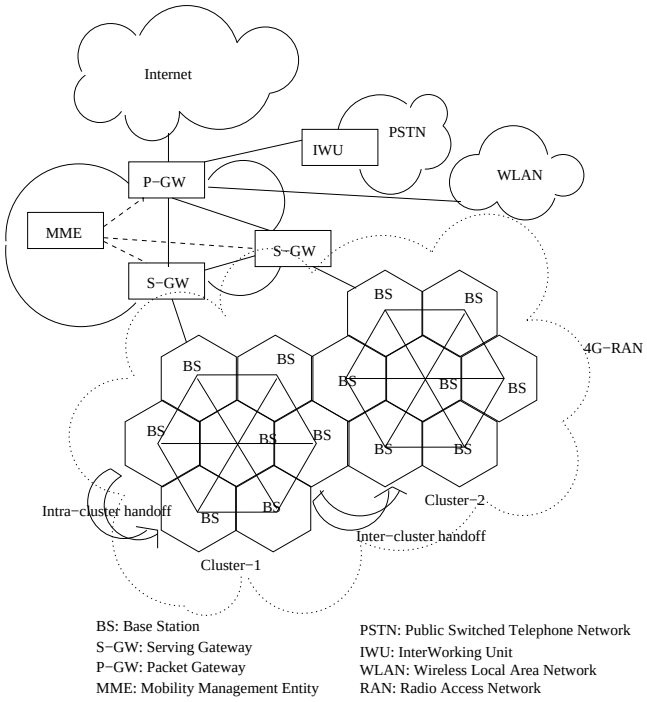


Fig. 1. Next generation mobile network architecture.

III. Network Architecture and System Model

A. Network Architecture

In our approach, Internet protocol has been considered as the transport technology for both core and access network in 4G networks [33], [34]. The IP-based core network connects the 4G mobile radio access network (RAN) with other access networks, such as wireless local area networks (WLAN), public switched telephone network (PSTN), and the Internet. In the 4G RAN, the BSs are connected to each other for local information sharing. This topology can be a mix of ring, tree, and mesh topologies. All BSs are equipped with the same functionalities and computation power. Several BSs are grouped into a cluster for a better localization of resource allocation, load balancing, and handoff processing within a cluster. One cluster consists of seven cells – one center cell and six cells around it, making it a two-tier network. We consider 4G mobile network architecture with a full mesh network topology, as shown in Fig. 1, where the BSs are inter-connected via high speed optical or wireless links. Each BS operates as a radio access router, performing functions in terms of dynamic wireless resource allocation (scheduling), handoff, and local database management. The core of the network consists of signaling gateway (S-GW), packet gateway (P-GW), and mobility management entity (MME). Each BS of a cluster is attached to a S-GW, where one S-GW can bear traffic for more than one cluster. P-GW is the interface for the Internet and the other switching networks like WLAN and PSTN. The main function of MME is to maintain a global database of all the users for authentication and authorization, location updates, and supervision of the network activities.

B. System Model

In the proposed network-level cooperation, the data traffic to/from a user in the coverage-overlap region is shared by two or more BSs, and the degree of sharing depends on the user and system parameters, such as QoS, cell loads, and ICI.

Simulation based studies by Bajzik et al. [16] showed that the impact of packet forwarding on handoff users can be reduced by using a transport layer scheduling that is able to differentiate among different service classes as well as cell-center and cell-edge users. Following the same principle, we present a new system queueing model as depicted in Fig. 2 for the users at the cell-edge and coverage-overlap regions. Queuing is applied at the *data link* and *transport layers*, to cooperate in transmission of user data while being in the coverage-overlap region. Two BSs, BS_i and BS_j are shown with a coverage overlap. N_i users are served only by BS_i, N_j users are served only by BS_j, and N_{ij} users in the coverage-overlap region are served by both BS_i and BS_j cooperatively. BS_i maintains queues for $N_i + N_{ij}$ users and serves them using scheduler $S_{downlink}$. Likewise, BS_j maintain queues for $N_j + N_{ij}$ users and serves them using $S_{downlink}$. *Note that, the scheduler at each BS, $S_{downlink}$ works on the same principle, but independently. Further, the resource allocation strategy at each BS is done via different resource blocks. Hence the SINRs experienced in reception from different cooperating BSs are independent.* A controller directs the flows from the classifier*, according to the routing table maintained at the controller. Based on the feedback from the BSs, the classifier is used to distinguish the incoming/outgoing flows if they are of a shared user – served by two BSs, or a non-shared user – served by only one BS. The decision, on the degree of cooperation, i.e., the sharing window size and its position, are taken based on the relative load of the BSs and a session's QoS requirement, which is determined at the call admission

*Classifier and controller are the integrated parts of the S-GW

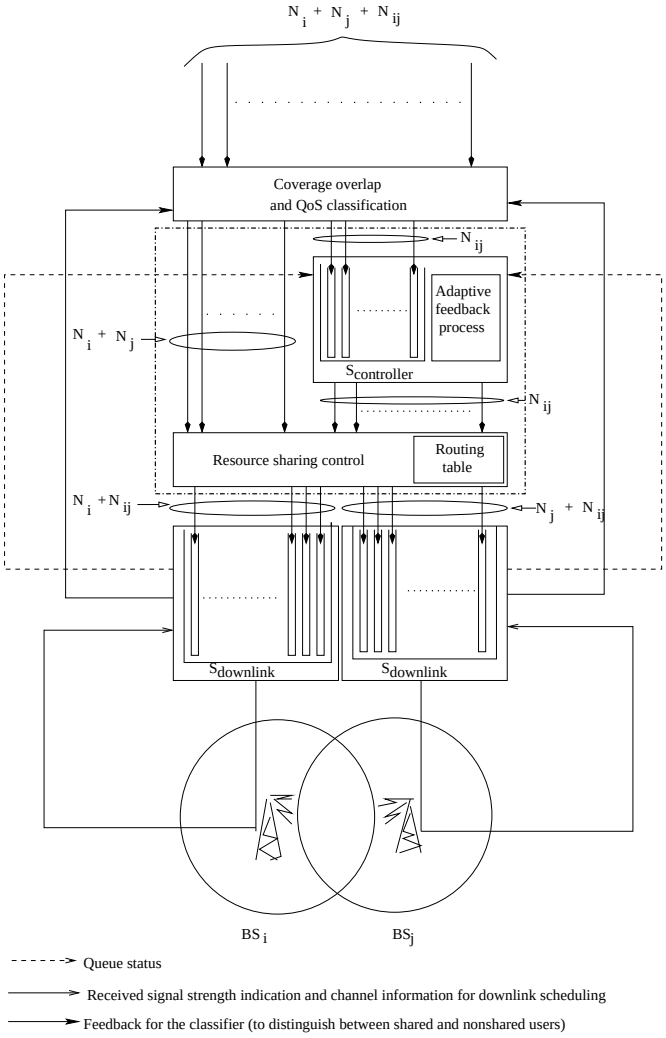


Fig. 2. System architecture and downlink queueing model for 2-cell network-level cooperation, with different queues for shared and non-shared users.

stage. The controller also maintains the queues for all users which can be served by both the BSs. Flow scheduling at the controller is according to the rule provided by $S_{controller}$.

C. Control and Data Traffic Loads

Before we present the performance analysis, the additional overheads of the proposed cooperative scheduling approach in comparison with HHO, SSHO, and CoMP for single antenna receivers (called, CoMP $n \times 1$) [7] are briefly evaluated.

Control signaling overhead of all the schemes are comparable, as they require feedback to the controller on downlink channel condition of the active set (the anchor BS and its neighbors whose signal strengths at the MS are acceptable). Also, except for HHO, all the schemes have downlink control signaling overhead for transmitting the resource block allocation information via multiple BSs. In particular, control overhead in the proposed approach and CoMP $n \times 1$ are comparable. In the proposed scheme, for the cell-edge users the resource blocks from the cooperating BSs are *non-overlapping*, whereas they are *common* in CoMP $n \times 1$ – thereby requiring more stringent coordination for synchronization.

The data load on the BS is the sum of data traffic requested by the users in the current cell and that of the neighbors that are in the cooperation region. For a cell of radius R, the area of cooperation is on average $2\pi R^2 - 3\sqrt{3}R^2$, which is about 34% of the total cell area. With uniform distribution of the users in a cell, on average half of the users in the cooperation region are anchored to the current BS, while the remaining half are anchored to the neighboring BSs. In the proposed scheme, since the cooperation is based on traffic sharing among the cooperating cells, the average traffic load in the system remains the same as in HHO and SSHO irrespective of the number of users in the cooperation region. Due to the same reason, the backhaul traffic load in the proposed scheme is comparable to that of HHO and SSHO. In contrast, since CoMP $n \times 1$ uses maximal ratio transmission based replication of data to the users who are in the cooperation region, the average data load on the BS is about 17% more than that in the proposed scheme. Also, since the replicated traffic has to be transmitted to all the N_c cooperating BSs, in this scheme there is a N_c fold increased backhaul data traffic due to the users in the cooperation region. We account for the cooperation-dependent BS traffic loads in the numerical results (Section VI).

IV. Outage Performance

Before studying the capacity gain, in this section we analyze the outage performance of the cell-edge users in a coded communication system and UFR clustered environment.

The normalized distance of a MS from BS_i is denoted as $d_{i,x}$, where x is its spatial position. The path loss between BS_i and MS is $L_{i,x} = d_{i,x}^{-l} 10^{\zeta_{i,x}/10}$, where l is the path loss exponent and $\zeta_{i,x}$ is a Gaussian random variable with zero mean and standard deviation σ, representing shadow fading. We note that, downlink power control is important to mitigate the effect of time-varying fading channel where same coding and modulation scheme is used, e.g. in 3GPP Release 99. In advanced communication systems, e.g., 3GPP LTE, adaptive modulation and coding schemes make use of fading channels at constant transmit power. Also, as noted in [35], in UFR scenario, same power allocation to all downlink users results in only negligible loss in system capacity. Accordingly, in UFR systems, constant power allocation to the users is considered [35], [36]. Thus, to simplify the outage analysis, we assume the downlink transmit power P is the same for all BSs. So, the SINR at position x in the ith cell can be written as:

$$\gamma_{i,x} = \frac{P \cdot L_{i,x}}{I^0_{i,x} + I^s_{i,x} + N_0 B}$$

where $I^0_{i,x}$ is the ICI, $I^s_{i,x}$ is the intra-cell interference, N_0 is the power spectral density of additive white Gaussian noise (AWGN). We assume, the intra-cell interference is negligible, i.e., $I^s_{i,x} \approx 0$, and $I^0_{i,x} \gg N_0 B$. Accordingly, the SINR formula can be approximated as:

$$\gamma_{i,x} = \frac{P \cdot L_{i,x}}{I^0_{i,x}} \quad (1)$$

ICI from BS_j affects only when the assigned frequency resource in BS_i overlaps with the assignment in BS_j [22].

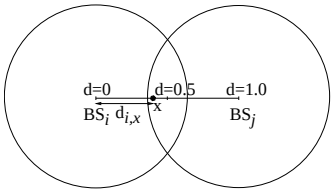


Fig. 3. Two BS cooperation model.

We define the collision probability per carrier from BS_j as:

$$P_{col,j} = P_{sel}(j|i) = \frac{P_{sel}(j) \cdot P_{sel}(i)}{P_{sel}(i)} = P_{sel}(j)$$

The second equality is a direct consequence of independent resource (carrier) allocation strategy (discussed in Section III-B). So, collision probability in BS_i is the resource selection probability in BS_j, and can be given as the load on BS_j.

$$P_{sel}(j) = \frac{\text{total used resources}}{\text{total available resources}} \triangleq \rho_j$$

Thus, the collision probability from BS_j is the load on BS_j, i.e., $P_{col,j} = \rho_j$. Hence, the ICI from all neighboring BSs is:

$$I^0_{i,x} = \sum_{j=1}^{N_{oc}} P \cdot L_{j,x} \cdot \rho_j \tag{2}$$

Here N_{oc} is the number of neighboring BSs (6 in first tier and 18 in first and second tier). Using (1) and (2) we get,

$$\gamma_{i,x} = \frac{L_{i,x}}{\sum\limits_{j=1}^{N_{oc}} L_{j,x} \cdot \rho_j} = \frac{10^{\zeta_{i,x}/10}}{\sum\limits_{j=1}^{N_{oc}} (d_{i,x}/d_{j,x})^l \cdot 10^{\zeta_{j,x}/10} \rho_j} \tag{3}$$

In our analysis, the first and second tier BSs are considered for ICI. For quantification of cooperative scheduling gain, a MS is assumed to be traveling on a straight line trajectory from BS_i towards BS_j at a constant speed (Fig. 3).

In the proposed system, M-QAM modulation is employed in each of the BSs, and coherent demodulation is assumed at the MS. Thus, the number of bits per symbol is $r = \log_2 M$. In an uncoded communication scenario, the outage probability in BS_i at position x is defined as the probability of received SINR being less than some predefined threshold SINR value:

$$P_{out,x}(i) = P[\gamma_{i,x} < \gamma_{th}]$$

γ_{th} is defined as per the downlink BER requirements.

Following the approximation approach by Schwartz and Yeh [37], we obtain $P_{out,x}(i)$ as:

$$P_{out,x}(i) = P\left[\ln \gamma_{i,x} < \ln \gamma_{th}\right] = 1 - Q\left(\frac{\ln \gamma_{i,x} - m_{(\ln \gamma_{i,x})}}{\sigma_{(\ln \gamma_{i,x})}}\right) \tag{4}$$

The statistical parameters $m_{(\ln \gamma_{i,x})}$ and $\sigma_{(\ln \gamma_{i,x})}$ in (4) are derived in Appendix A.

P_{out} in hard handoff (HHO) is defined as

$$P^{HHO}_{out,x} = \sum_{i=1}^{N_c} P_x(i) \cdot P_{out,x}(i)$$

N_c is the number of BSs in a cluster. The probability that a user at position x is connected to BS_i, $P_x(i)$ is defined as

$$P_x(i) = \text{P}[L_{i,x} - L_{j,x} > \delta_{ho}], \ \forall\ i \neq j \quad \text{and} \quad i,j \in N_c$$

where δ_{ho} is the handoff margin. We ignore the incoming/outgoing probability for MSs to/from BS_i. The reason being the consideration of MS with a defined trajectory pattern as shown in Fig. 3. This also ignores the ping-pong effects.

In SSHO [22], since SSTD is applied when a MS is in coverage-overlap region, P_{out} in an uncoded communication environment is defined as:

$$\begin{aligned} P^{SSHO}_{out,x} =& P_x(i) \cdot P_{out,x}(i) + \sum_{\forall\ j} P_x(j) \cdot P_{out,x}(j) \\ &+ \sum_{\forall\ j} P_x(i,j) \cdot P_{out,x}(i) + \sum_{\forall\ j} P_x(j,i) \cdot P_{out,x}(j) \\ &+ \sum_{\substack{\forall\ j,k \\ j \neq k}} P_x(j,k) \cdot P_{out,x}(j), \quad \forall\ \ i,j,k \in N_c \end{aligned}$$

where $P_x(i) + \sum\limits_{\forall\ j} (P_x(i) + P_x(i,j) + P_x(j,i)) + \sum\limits_{\substack{\forall\ j,k \\ j \neq k}} P_x(j,k) = 1$, and $P_x(i,j)$ is the probability that the MS is in the handoff region of BS_i and BS_j, defined as

$$\begin{aligned} P_x(i,j) =& \text{P}[0 < (L_{i,x} - L_{j,x}) < \delta_{ho} \quad \text{and} \quad L_{j,x} > L_{k,x}], \\ & \forall\ \ i \neq j \neq k \quad \text{and} \quad i,j,k \in N_c \end{aligned}$$

Note that, in SSHO $P_x(i,j)$ cannot be ignored because of the SSTD from multiple BSs in the handoff region.

In CoMP $n \times 1$ [7], cooperating BSs transmit the same content to the MS over the same resource blocks. As in maximal ratio transmission, the transmit power is optimally adjusted by the cooperating BSs to ensure maximum gain. So, from (3), the total SINR at the MS in CoMP $n \times 1$ is:

$$\gamma_{\text{CoMP}(n\times 1)} = \frac{\sum\limits_{i \in \mathcal{N}} d^{-l}_{i,x} \cdot 10^{\zeta_{i,x}/10}}{\sum\limits_{\substack{j=1 \\ j \notin \mathcal{N}}}^{N_{oc}} d^{-l}_{j,x} \cdot 10^{\zeta_{j,x}/10} \rho_j} \tag{5}$$

where $\mathcal{N}$ is the set of BSs transmitting to user u in CoMP $n \times 1$. From (5), the outage probability $P^{CoMP(n\times 1)}_{out,x}$ can be found as in HHO using (4), with $\gamma_{i,x}$ replaced by $\gamma_{\text{CoMP } n\times 1}$.

As highlighted in Section III-B, in the proposed network-level cooperation, transmission from the different BSs are done via different resource blocks. Therefore, the SINR experienced for reception from BS j is independent of the SINR experienced for reception from BS k ($k \neq j$). As a result, the total outage at point x when MS is anchored to BS i is the product of the outage probabilities from the individual cooperating BSs. Taking into account the fact that the MS at point x can be anchored to any BS in the cluster, we sum up the outage probability when the MS is connected to different BSs. Hence, in the proposed scheme, P_{out} in an uncoded communication environment at position x is the probability that the MS is in outage from all the BSs, which is obtained as:

$$P^{prop}_{out,x} = \sum_{i=1}^{N_c} P_x(i) \cdot P_{out,x}(i) \cdot \prod_{j=1}^{N_{oc}} P_{out,x}(j), \quad \forall\ \ i \neq j$$

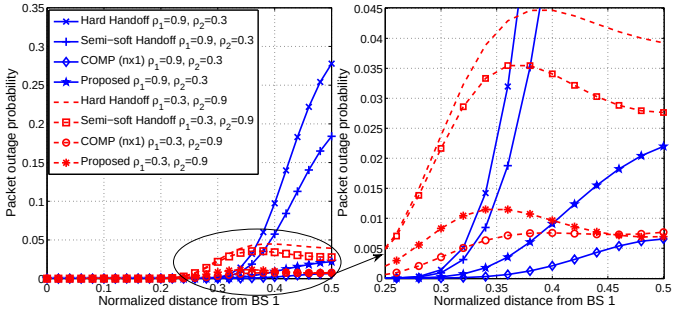


Fig. 4. Comparison of packet outage probability in HHO, SSHO, CoMP $n \times 1$, and the proposed scheme in different traffic loading conditions with $[336, 320, 16]_2$ linear coding and and 4-QAM. $\gamma_{th} = 3$ dB, path loss factor $l = 3$, shadow fading mean 0 and standard deviation 6 dB. 2-cell cooperation ($n = 2$) is considered with neighboring cell loads same as ρ_2.

Considering, for example, the downlink traffic, outage probability in HHO is to do solely with the signal received from the serving BS. In SSHO, this probability implies that, either the MS is in non-overlapped coverage region and the signal received is below γ_{th}, or it is in handoff region between two neighboring cells and the signal received from the BS with stronger signal is below γ_{th}. In the proposed cooperation scheme, the outage probability indicates that the MS is in some cell coverage region but the received signal quality at that location from any BS is not above γ_{th} – irrespective of the currently serving BS and the target BS in case of handoff.

In a coded communication scenario, the outage performance is obtained as packet outage probability or packet error probability (PEP). If $[n, k, d]_q$ linear block codes are used as a forward error correction (FEC) technique at the physical layer, where k is the message length in bits, n is the block length after coding, q is the alphabet size (binary in this case), and $d \leq n - k + 1$ is the Hamming distance, PEP for a particular scheme χ is defined as P^{χ}_{PEP} [38]:

$$P^{\chi}_{PEP} = \sum_{j=t+1}^{n} \binom{n}{j} \left(P^{\chi}_{out,x}\right)^{j} \cdot \left(1 - P^{\chi}_{out,x}\right)^{(n-j)}$$

where $P^{\chi}_{out,x}$ is the bit outage probability of a given scheme, i.e., $P^{HHO}_{out,x}$ for HHO, $P^{SSHO}_{out,x}$ for SSHO, $P^{CoMP(n\times1)}_{out,x}$ for CoMP $n \times 1$, and $P^{prop}_{out,x}$, for the proposed scheme.

The PEP of HHO, SSHO, CoMP $n \times 1$, and the proposed scheme with coded M-QAM are shown in Fig. 4 with $\gamma_{th} = 3$dB. Linear block code used is $[336, 320, 16]_2$. With the hamming distance $d = 16$, this coding can correct up to $\lfloor \frac{d-1}{2} \rfloor = 7$ errors. Though the outage probabilities are high, in practical systems it can be scaled down by considering adaptive M-QAM and FEC coding scheme, and the effect of high outage can be reduced by considering link-layer adaptive schemes, e.g., hybrid automatic repeat request (HARQ).

With $\rho_1 = 0.9$ and $\rho_2 = 0.3$, the ICI is less. Hence this case is *receive signal strength dominated*. The ICI being less prominent, the handoff is likely to happen at or beyond the midway between BS_1 and BS_2. So, in all the handoff/cooperation schemes, as the MS moves away from BS_1, the outage probability increases. In contrast, with $\rho_1 = 0.3$ and $\rho_2 = 0.9$, since the neighboring cells are now highly loaded, this case is *interference dominated*. Here, as the MS moves away from BS_1, the effect of high ICI causes a fast decrease of SINR, leading to the handoff occurrence at a distance closer to BS_1. As a consequence, in all the schemes the outage probability peaks at a distance closer to BS_1.

Both SSHO and the proposed scheme utilize macro-diversity and hence they provide more reliability to the user connection. But the proposed scheme considers network-level cooperation where the MS can utilize signals from multiple BSs in the same frame (via different resource blocks) instead of selection of one. Hence, the proposed scheme is expected to perform better than SSHO, as evident from Fig. 4. CoMP $n \times 1$ on the other hand operates by combining the physical signals with the same content transmitted over the common resource blocks from the cooperating BSs. This approach demands equal resource allocation from the cooperating BSs. At low neighboring cells load the cooperating BSs can provide the required additional resource, and as a result compared to the proposed approach, CoMP offers a better outage performance. At high neighboring cells load ICI dominates. Also, the additional resource available from the cooperating cells is insufficient in CoMP. But, in this case the proposed approach allows optimal load sharing among the unequally loaded cooperating cells. Therefore, although CoMP performs better at a distance closer to the BS_1, it performs increasingly poorly toward the cell-edge. We also note that, although under certain conditions the outage performance of CoMP may be better, with respect to capacity it is expected to have the disadvantage of occupying additional resource.

In the following analysis (in Section V) and numerical simulations (Section VI) we investigate the effective capacity gain with the proposed network-level cooperation with respect to no cooperation (when an MS is served by only one BS at a time) and contrast with the capacity gain of CoMP $n \times 1$ [7]. These comparisons are valid as far as the system level performance is concerned, because the architecture and system model we have presented in Sections III-A and III-B respectively are able to support these schemes.

V. Scheduling of Shared Users and Optimal Resource Allocation

To study shared resource scheduling, first we define the user/flow-specific QoS and review the effective capacity concept in *single cell scenario*, introduced by Wu and Negi [11].

A. Quality of Service (QoS)

QoS is defined by the value of maximum queue size Q_{max} for a user (traffic type) beyond which the delay violation probability exceeds a predefined threshold ϵ, i.e., $\sup_t \Pr\{Q(t) \geq Q_{max}\} \leq \epsilon$, where t is a scheduling window.

It was shown in [11] that, for a dynamic queueing system, where the arrival and service processes are stationary and ergodic, the probability that the queue length $Q(t)$ exceeds the threshold Q_{max} satisfies

$$\sup_t \Pr\{Q(t) \geq Q_{max}\} \sim e^{-\theta(\Omega) Q_{max}} \qquad (6)$$

where $\theta(\Omega)$ is a function of source rate Ω, Q_{max} is a large quantity. $f(Q_{max}) \sim g(Q_{max})$ means that $\lim_{Q_{max}\to\infty} \frac{f(Q_{max})}{g(Q_{max})} = 1$. At small values of Q_{max}, it was shown in [11] that the following approximation is more accurate:

$$\sup_t \Pr\{Q(t) \geq Q_{max}\} \approx \Upsilon(\Omega) e^{-\theta(\Omega) Q_{max}} \tag{7}$$

where $\Upsilon(\Omega)$ is the probability that the buffer is nonempty, i.e., $\Upsilon(\Omega) = \Pr\{Q(t) > 0\}$, at any chosen time instant t. Note that, the parameter $\theta(\Omega) \geq 0$ in (6) and (7), called QoS exponent [11], determines the probability of QoS violation. A large value of $\theta(\Omega)$ implies a stringent QoS requirement.

B. Effective Capacity

For a user, whose QoS requirement is indicated by the QoS exponent θ, the effective capacity at the link layer is a measure of the maximum constant arrival rate that can be supported by the system. With a service rate $\mu(n)$ at the nth block, the total channel service rate $R(m)$ is: $R(m) = \sum_{n=0}^{m} \mu(n)$, where m is the block length. Then, for a given θ, using the effective capacity concept in [11], the effective capacity for a discrete-time stationary and ergodic service process is given by,

$$E_C(\theta) \stackrel{\triangle}{=} -\lim_{m\to\infty} \frac{1}{\theta m} \ln E\{e^{-\theta R(m)}\}$$

For uncorrelated block fading channels, the service process $\{\mu(n), n = 0, 1, \cdots\}$ is uncorrelated, where the effective capacity expression simplifies to: $E_C(\theta) = -\lim_{m\to\infty} \frac{1}{\theta m} \ln\left[E\{e^{-\theta\mu(0)}\} \cdot E\{e^{-\theta\mu(1)}\} \cdots E\{e^{-\theta\mu(m)}\}\right] = -\frac{1}{\theta} \ln E\{e^{-\theta\mu(n)}\}$, for any $n = 0, 1, \cdots$. With a frame length T_f and system bandwidth B, $S \stackrel{\triangle}{=} T_f B$ is the total time-frequency resources available in one frame, and the above expression can be normalized as:

$$E_C(\theta) = -\frac{1}{\theta T_f B} \ln E\{e^{-\theta\mu}\} \tag{8}$$

C. Scheduling of Shared Users

The total resource allocated to the non-shared and shared users in cell BS_i is $\sum_{u=1}^{N_i+N_{ij}} S_i^u$, $j = 1, \cdots, N'$ $\left(1 \leq N' \leq N_{oc}\right)$, $j \neq i$, where S_i^u $(0 \leq S_i^u \leq S_i)$ is the resource allocated to user u in BS_i. The total resource available at the BS_i is S_i. For simplicity we assume $S_i = S$ $\forall\ i \in N_c$. From (8), for user u with QoS exponent θ^u scheduled from BS_i, the effective capacity is:

$$E_{C,i}^u(\theta^u) = -\frac{1}{\theta^u S} \ln E\{e^{-\theta^u \mu_i^u}\} \tag{9}$$

where $\mu_i^u = r_i^u S_i^u$ is the rate provided to user u from BS_i and r_i^u is its modulation index.

If the same user is scheduled from more than one BS, say BS_i and $\{BS_j\}$, for $j = 1, \cdots, N'$ $\left(1 \leq N' \leq N_{oc}\right)$, $j \neq i$, then the total effective capacity, which we term as *joint effective capacity* $E_{C,joint}^u(\theta^u)$, can be expressed as:

$$\begin{aligned} E_{C,joint}^u(\theta^u) &= \frac{-1}{\theta^u S} \ln E\{e^{-\theta^u \mu_i^u}\} - \frac{1}{\theta^u S} \sum_{j=1, j\neq i}^{N'} \ln E\{e^{-\theta^u \mu_j^u}\} \\ &= \frac{-1}{\theta^u S} \ln \left[E\{e^{-\theta^u \mu_i^u}\} \cdot \prod_{j=1, j\neq i}^{N'} \left\{E\left(e^{-\theta^u \mu_j^u}\right)\right\}\right] \end{aligned} \tag{10}$$

subject to the condition that the joint resources from $\{BS_j\}$ are the same as in (9) and the SINR is above the acceptable threshold. Mathematically, this condition can be stated as: $S^u = S_i^u + \sum_{j=1, j\neq i}^{N'} S_j^u$ with the conditions $\gamma_{i,x},\ \gamma_{j,x} > \gamma_{th}$, $\forall\ j = 1, \cdots, N',\ j \neq i$, where $\mu_i^u = r_i^u \cdot S_i^u$, and $\mu_j^u = r_j^u \cdot S_j^u$. Here S_b^u is the resource allocated from (BS_b) to user u, and μ_b^u is the corresponding rate achieved.

In (10), it is implicitly assumed that, all the BSs have their own resource allocation policies which work independently on the same principle for the whole network. Let us denote $P_i = \Pr\{\gamma_i^u \leq \gamma_{th}\}$, and $P_j = \Pr\{\gamma_j^u \leq \gamma_{th}\}$ $\forall\ j = 1, \cdots, N',\ j \neq i$. Then, $E\{e^{-\theta^u \mu_i^u}\} = e^{-\theta^u \mu_i^u}(1 - P_i) + P_i$, $E\{e^{-\theta^u \mu_j^u}\} = e^{-\theta^u \mu_j^u}(1 - P_j) + P_j$. Hence,

$$\begin{aligned} E_{C,joint}^u(\theta^u) = -\frac{1}{\theta^u S} \Big[&\ln \left[\{e^{-\theta^u \mu_i^u}(1 - P_i) + P_i\}\right. \\ &\left. \cdot \prod_{j=1, j\neq i}^{N'} \{e^{-\theta^u \mu_j^u}(1 - P_j) + P_j\}\right] \Big] \end{aligned} \tag{11}$$

The discussion in Section IV underlined the gain using network-level cooperation in terms of outage performance (Fig. 4), and hence (11) accommodates the same cooperation while scheduling a user from more than one BS.

In the context of CoMP (distributed MISO [7]), as there are $|\mathcal{N}| = n$ BSs transmitting the same content to the user u, the resource available to u from n BSs is n times the resources allocated by the single BS before cooperation. Hence the effective capacity in CoMP $n \times 1$ is:

$$E_{C,\text{CoMP}(n\times 1)}^u(\theta^u) = -\frac{1}{\theta^u n S} \ln E\{e^{-\theta^u \mu^u}\} \tag{12}$$

where $\mu^u = r^u S^u$ is the rate allocated to user u by one of the cooperating BSs, with S^u as the resource allocated by one of the cooperating BSs and r^u is its modulation index. It may be recalled here that, in CoMP $n \times 1$ all the cooperating BSs should have the required resources available for cooperation.

In network-level cooperation, the issues still to be addressed are: 1) how to divide the resources S_i^u and S_j^u among BS_i and $\{BS_j\}$, $j = 1, \cdots, N',\ j \neq i$, so that the effective capacity of the user can be increased, and 2) when to start joint scheduling of the users. We will address these problems and aim for optimal solutions in the following sub-sections.

D. Optimal Resource Allocation in the Proposed Approach

The first problem in hand is to maximize the joint effective capacity $E^u_{C,joint}(\theta^u)$ for user u in (10). Mathematically,

$$E^{u,opt}_{C,joint}(\theta^u) = \max_{\left\{\substack{S^u_i,\ S^u_j,\\ \forall\ j=1,\cdots,N^{'},\\ j\neq i}\right\}} \left\{ \frac{-1}{\theta^u S}\left[\ln E\{e^{-\theta^u \mu^u_i}\} + \sum_{\substack{j=1\\ j\neq i}}^{N^{'}} \ln E\{e^{-\theta^u \mu^u_j}\}\right]\right\} \tag{13}$$

$$\text{s.t.}\ \ S^u_i + \sum_{j=1, j\neq i}^{N^{'}} S^u_j = S^u \ \ , S(1-\rho_j) \geq S^u_j,$$
$$j = 1,\cdots,N^{'}, j\neq i, \text{and}\ \ S^u > S^u_i, S^u_j > 0$$

Using (11), the above optimization problem can be written as:

$$f^u_{\theta^u}\left(S^u_i, S^u_j \forall\ j=1,\cdots,N^{'}, j\neq i\right) = \max_{\left\{\substack{S^u_i,\ S^u_j,\\ \forall\ j=1,\cdots,N^{'}, j\neq i}\right\}} -\ln\left[\{e^{-\theta^u\mu^u_i}(1-P_i)+P_i\} \cdot \prod_{\substack{j=1\\ j\neq i}}^{N^{'}} \{e^{-\theta^u\mu^u_j}(1-P_j)+P_j\}\right] \tag{14}$$

By standard optimization technique the Lagrangian function of (14) can be written as:

$$\Lambda\left(S^u_i, S^u_j, \eta_j\ \ \forall\ j=1,\cdots,N^{'}, j\neq i,\ \lambda\right) = -\ln\left[\{e^{-\theta^u r S^u_i}(1-P_i)+P_i\}\cdot \prod_{\substack{j=1\\ j\neq i}}^{N^{'}} \{e^{-\theta^u r S^u_j}(1-P_j)+P_j\}\right] + \sum_{\substack{j=1\\ j\neq i}}^{N^{'}} \eta_j\left[S(1-\rho_j) - S^u_j\right] + \lambda\left[S^u_i + \sum_{\substack{j=1\\ j\neq i}}^{N^{'}} S^u_j - S^u\right] \tag{15}$$

where λ and $\eta_j\ \ \forall\ j=1,\cdots,N^{'}, j\neq i$, are the Lagrangian multipliers. We have taken $r^u_i = r^u_j = r$ for mathematical simplicity. Partial differentiation of (15) with respect to S^u_i, S^u_j, λ, and η_j will give the optimal solution for S^u_i and S^u_j. We set $\frac{\partial\Lambda(S^u_i,S^u_j,\eta_j,\lambda)}{\partial S^u_i} = 0$, $\frac{\partial\Lambda(S^u_i,S^u_j,\eta_j,\lambda)}{\partial S^u_j} = 0\ \forall j = 1,\cdots,N^{'}, j\neq i$, and $\frac{\partial\Lambda(S^u_i,S^u_j,\eta_j,\lambda)}{\partial\lambda} = 0$. After simplification,

$$\frac{(1-P_i)\theta^u r e^{-\theta^u r S^u_i}}{e^{-\theta^u r S^u_i}(1-P_i)+P_i} + \lambda = 0 \tag{16a}$$

$$\frac{(1-P_j)\theta^u r e^{-\theta^u r S^u_j}}{e^{-\theta^u r S^u_j}(1-P_j)+P_j} + \lambda - \eta_j = 0, \forall\ j=1,\cdots,N^{'}, j\neq i \tag{16b}$$

$$\text{and}\ \ S^u_i + \sum_{\substack{j=1\\ j\neq i}}^{N^{'}} S^u_j = S^u \tag{16c}$$

From the Karush-Kuhn-Tucker (KKT) conditions, we have

$$\eta_j(S(1-\rho_j) - S^u_j) = 0, \eta_j \geq 0, \forall\ j=1,\cdots,N^{'}, j\neq i \tag{17}$$

Solving for S^u_i and S^u_j, $\forall\ j=1,\cdots,N^{'}, j\neq i$, from equations (16a) and (16b), we obtain

$$S^u_i = -\frac{1}{\theta^u r}\ln\left(\frac{-\lambda P_i}{(\theta^u r+\lambda)(1-P_i)}\right) \tag{18a}$$

$$S^u_j = -\frac{1}{\theta^u r}\ln\left(\frac{P_j(\eta_j-\lambda)}{(\theta^u r+\lambda-\eta_j)(1-P_j)}\right) \tag{18b}$$

(16c), (17), (18a), and (18b) form a system of equations with Lagrange multipliers as the unknowns. In general, using numerical methods, we can solve this system of equations to obtain the values of Lagrange multiplier. Substituting Lagrange multiplier values back in (18a) and (18b), we obtain S^u_i and S^u_j, $\forall\ j=1,\cdots,N^{'}, j\neq i$, the optimal resource allocation from $N^{'}+1$ BSs while the user is in the cooperation region, so that the total effective capacity can be increased when the user is able to connect to these BSs.

The numerical methods do not give a closed form expression of the solution. As an alternative, an iterative procedure can be applied to solve the optimal resource allocation problem, which is as follows. From the KKT conditions in (17), we have either $\eta_j = 0$ or $S^u_j = S(1-\rho_j)$, $\forall\ j=1,\cdots,N^{'}, j\neq i$. Considering $\eta_j = 0\ \forall\ j$, the optimal solution can be computed from (16a) and (16b), which are obtained as:

$$S^u_i = \frac{S^u}{N^{'}+1} + \frac{1}{(N^{'}+1)\,\theta^u r}\ln\left[\frac{(1-P_i)^{N^{'}}\prod_{\substack{j=1\\ j\neq i}}^{N^{'}} P_j}{P_i^{N^{'}}\prod_{\substack{j=1\\ j\neq i}}^{N^{'}}(1-P_j)}\right] \tag{19a}$$

$$S^u_j = \frac{S^u}{N^{'}+1} + \frac{1}{(N^{'}+1)\,\theta^u r}\ln\left[\frac{P_i\,(1-P_j)^{N^{'}}\prod_{\substack{j_1=1\\ j_1\neq i,j}}^{N^{'}} P_{j_1}}{(1-P_i)\,P_j^{N^{'}}\prod_{\substack{j_1=1\\ j_1\neq i,j}}^{N^{'}}(1-P_{j_1})}\right] \tag{19b}$$

However, the above solution in (19) with $\eta_j = 0\ \forall\ j$ does not guarantee feasibility, because the constraint $S^u_k \leq S(1-\rho_k)$ for some $k\in\{j=1,\cdots,N^{'}, j\neq i\}$ may be violated. Let $\{\kappa\}$ is the set of all k for which the constraint is violated. To achieve a solution, at those $k\in\{\kappa\}$ we set the optimal value of S^u_k to $S(1-\rho_k)$. At this stage, the optimization problem in (13) is modified by replacing S^u with $S^u - \sum_{k\in\kappa} S^u_k$ and $j=1,\cdots,N^{'}, j\neq i, j\neq k, k\in\{\kappa\}$, to form a reduced optimization problem. The reduced problem is solved in the similar way as above, and the process is repeated until all the constraints are satisfied. The solution obtained through the above procedure satisfies the KKT conditions. The reader is referred to Appendix B for the proof.

The optimization in (13) is a maximization problem. The Hessian of the objective function in (15) can be computed as

$$\frac{\partial f^u_{\theta^u}\left(S^u_i, S^u_j\right)}{\partial S^u_i} = \frac{(1-P_i)\theta^u r e^{-\theta^u r S^u_i}}{e^{-\theta^u r S^u_i}(1-P_i)+P_i}$$
$$\frac{\partial^2 f^u_{\theta^u}\left(S^u_i, S^u_j\right)}{\partial^2 S^u_i} = \frac{-P_i(1-P_i)(\theta^u r)^2 e^{-\theta^u r S^u_i}}{(e^{-\theta^u r S^u_i}(1-P_i)+P_i)^2} \leq 0$$

We observe that the Hessian of the optimization problem with respect to S_i^u is always ≤ 0. Similar result holds when the Hessian is computed with respect to $S_j^u, \forall\ j = 1, \cdots, N', j \neq i$. The cross terms in Hessian are 0. Hence the Hessian of the optimization problem is negative semi-definite for all values of S_i^u and S_j^u. The obtained KKT point is optimal as the objective function is concave and the constraints are linear.

Numerical results on optimized resource allocation and effective capacity gain are presented in Section VI.

The *region of interest (RoI)* is where the gain in effective capacity by cooperation is positive. Though the above solution for resource allocation is optimized, it does not necessarily provide a positive gain for all the values of P_i and P_j. In the next subsection, we define the achievable positive effective capacity gain region, i.e., the RoI, and study its existence.

E. Region-of-Interest (RoI) and Its Existence

For simplicity of exposition, without loss of generality, we consider the cooperation between two neighboring BSs, BS_i and BS_j, with resource available from BS_j, $S(1-\rho_j)$ is greater than the resource S_j^u required by the user u.

Definition 1. *RoI $\Re^u(\theta^u)$ for user u with QoS exponent θ^u is defined as the region, where the gain in effective capacity is positive. So, $\forall\ \ i \neq j,\ i, j \in N_c$,*

$$\Re^u(\theta^u):\ \ E_{C,joint}^{u,opt}(\theta^u) - \max\{E_{C,i}^u(\theta^u), E_{C,j}^u(\theta^u)\} > 0 \tag{21}$$

Proposition 1. *A non-zero RoI $\Re^u(\theta^u)$ is achievable for a user irrespective of its QoS demand (QoS exponent), where $\Re^u(\theta^u)$ is defined in* (21).

Proof: For a straight line trajectory, as given in Fig. 3, there will be a point where $P_i = P_j\ \forall\ \ i, j \in N_c$. At this point the effective capacity from an individual BS is

$$E_{C,i}^u(\theta^u) = E_{C,j}^u(\theta^u) = -\frac{1}{\theta^u S}\ln\{F(S^u)\}$$

where $F(S^u) = e^{-\theta^u r S^u}(1-P_i)+P_i$ is a decreasing function of S^u for a fixed $P_i,\ 0 < P_i < 1$, such that $0 < F(S^u) < 1$. The optimized joint effective capacity in (11) is obtained as:

$$E_{C,joint}^{u,opt}(\theta^u) = -\frac{1}{\theta^u S}\ln\left[\left\{e^{-\theta^u r S_i^u}(1-P_i)+P_i\right\} \cdot \left\{e^{-\theta^u r S_j^u}(1-P_j)+P_j\right\}\right]$$

For $P_i = P_j$, from (19), $S_i^u = S_j^u = \frac{S^u}{2}$. So,

$$E_{C,joint}^{u,opt}(\theta^u) = -\frac{1}{\theta^u S}\ln\left\{F^2\left(\frac{S^u}{2}\right)\right\}$$

where $F(S^u)$ is defined as above. From the properties of the function $F(S^u)$, we obtain:

$$F^2\left(\frac{S^u}{2}\right) < F(S^u) \tag{22}$$

The proof of (22) is in Appendix C. From here we conclude:

$$E_{C,joint}^{u,opt}(\theta^u) > E_{C,i}^u(\theta^u),\ \text{i.e.,}\ E_{C,joint}^{u,opt}(\theta^u) - E_{C,i}^u(\theta^u) > 0$$

□

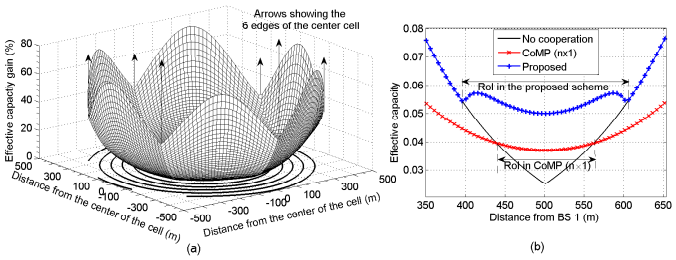


Fig. 5. (a) Effective capacity gain of the proposed scheme with respect to no cooperation, as given in (23), for $\theta = 0.9$, $\gamma_{th} = 3$ dB, and $\rho = 0.9$. (b) Capacity comparison with $\rho = 0.7$ and all other parameters same as in (a).

Hence it is proved that $\forall\ \ i \neq j;\ \ i, j \in N_c$, a non-zero RoI, $\Re^u(\theta^u)$ for user u of QoS exponent θ^u, is achievable while the resources are allocated optimally as per (19). A similar statement can be made for the case when the resource available from BS_j, $S(1-\rho_j)$ is less than S_j^u. Again, it is notable that the values of S_i^u and S_j^u are optimal only when the resources allocated from each of the BSs involved in the cooperation are strictly positive, i.e., $S_i^u > 0$ and $S_j^u > 0$.

VI. Numerical Results

The proposed network-level cooperation performance is evaluated in terms of effective capacity gain (compared to the case when the MSs are served from one BS only, i.e., no cooperation) and cooperation window, subject to the QoS constraint and network load. The results are contrasted with the signal combining approach (CoMP $n \times 1$) [7]. The effective capacity gain in the proposed approach is computed as:

$$E_{C,gain}^{u,opt}(\theta^u) \triangleq \frac{E_{C,joint}^{u,opt}(\theta^u) - E_C^u(\theta^u)}{E_C^u(\theta^u)} \cdot 100\% \tag{23}$$

The capacity gain in CoMP $n \times 1$ is similarly defined.

For simplicity we consider a 7 cell cluster. So, only the two tier cellular structure is considered for ICI, as in Fig. 1. The model contains one center cell and six interfering (second tier) cells with UFR plan [3]. As in the case of wireless systems with adaptive M-QAM modulation and also following the observations in [35], [36], the downlink power budget for each user is kept constant at 1 W. The radio signals are considered to have a path loss factor $l = 3$ and shadow fading with mean 0 dB and standard deviation 6 dB. The handoff margin δ_{ho} is taken 3 dB (from 3GPP spec. 25.331 v3.21.0). The user is assumed to travel with a constant speed on a straight line connecting two adjacent BSs where the distance between two adjacent BSs is 1000 m. The results on effective capacity gain is discussed first, and then the effect of user QoS exponent θ, threshold SINR value γ_{th}, and network loading factor ρ in a coded communication scenario are presented.

Fig. 5(a) shows the effective capacity gain due to network-level cooperation among more than one BS, where the QoS exponent value is taken as $\theta = 0.9$, $\gamma_{th} = 3$ dB (at $M = 4$), and the network loading in each cell is $\rho = 0.9$. The RoI defined in Section V-E is considered with respect to all neighboring cells from the center cell. Here, in each direction of movement of the MS, the number of cooperative BSs are

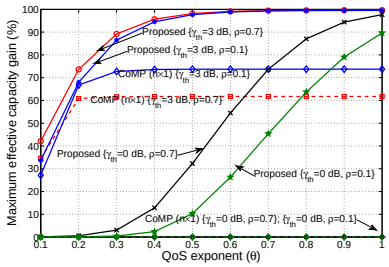


Fig. 6. Maximum effective capacity gain as defined in (24) versus θ for different SINR threshold γ_{th} and average network load ρ in 2-cell cooperation ($n = 2$). Center cell load $\rho_1 = 0.5$.

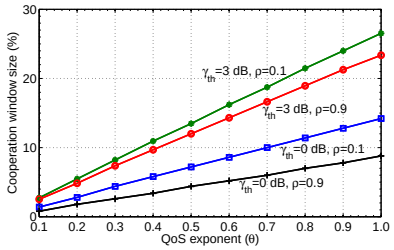


Fig. 7. Normalized cooperation window size in the proposed scheme versus θ for different γ_{th} and average network load ρ.

limited to two at any time and chosen such that the gain in effective capacity is maximum as per (21). The resource allocation is optimized as described in Section V-D. Here $E^u_C(\theta^u) = \max\{E^u_{C,i}(\theta^u), E^u_{C,j}(\theta^u)\} \;\; \forall \;\; i \neq j; \;\; i,j \in N_c$.

Relative effective capacity of the proposed scheme (2-cell cooperation) with respect to no cooperation and CoMP $n \times 1$ (with $n = 2$) is shown in Fig. 5(b), which indicates an appreciably better performance. The RoI in the proposed approach as well as in CoMP $n \times 1$, indicating positive gain with respect to no cooperation, are also highlighted.

It is difficult to show the relative performances in 3-D plots. So, we have compared the results on three aspects: (1) cooperation window size, (2) window position, and (2) maximum effective capacity gain. These metrics jointly capture the different user and network effects on the system performance.

A. User Effect

The effect of user QoS on effective capacity gain is studied here. The maximum effective capacity gain is defined as:

$$E^{u,opt}_{C,gain}(max)(\theta^u) \triangleq \max\left[E^{u,opt}_{C,gain}(\theta^u)\right] \qquad (24)$$

Fig. 6 shows maximum effective capacity gains of the two cooperation schemes (CoMP $n \times 1$ and the proposed one) versus QoS constraint θ for different SINR threshold γ_{th} and network load ρ. At a low ρ and a small value of θ, CoMP $n \times 1$ performs close to the proposed scheme, which is because with sufficient neighboring cell resource and at low ICI the capacity gain by signal combining in CoMP is close to the gain due to the proposed traffic sharing based cooperation. However, the maximum gain in CoMP $n \times 1$ is in general quite less and moreover it saturates soon. This is because, a higher value of θ demands a higher degree of cooperation. But in CoMP $n \times 1$ cooperative gain is limited because of the additional resource requirement, which increases with the degree of cooperation.

In contrast with the proposed approach, in CoMP $n \times 1$ gain is more at low system load ρ (see the plots with annotation 'CoMP $n \times 1$ $\{\gamma_{th} = 3$ dB, $\rho = 0.1\}$' and 'CoMP $n \times 1$ $\{\gamma_{th} = 3$ dB, $\rho = 0.7\}$'). This is because, at a higher system load CoMP finds it more difficult to have additional resources from the cooperating BSs (i.e., BS_2 in 2-cell cooperation), and also this scheme is oblivious to cell load asymmetry as compared to the proposed load sharing approach.

It can be observed that, for all values of γ_{th} and ρ, the gain is less with loose QoS requirement (less value of θ). As the θ is increased (i.e., toward stringent QoS), the gain in the proposed scheme reaches up to 100% at higher γ_{th} and ρ. Intuitively, in the cooperation region the packets to the MS can be delivered via multiple paths, thereby helping meet the delay guarantee easily. In delay-tolerant applications (with low value of θ), since the packets can be stored and eventually delivered reliably via only one BS, the benefit of cooperation is less.

At a low γ_{th}, the capacity gain is less because there is always a better BS to handle even a reasonably high QoS requirement. As a result, the cooperation requirement is less. But increasing the user QoS will increase the gain monotonically as the cooperation helps deal with the situation better. At a higher γ_{th}, the gain due to load sharing increases because neither of the BSs alone can guarantee the required data rate. Cooperation offers the macro-diversity gain by multiplexed transmission via multiple BSs. At less load the cooperation does not yield much gain because the resources available at one BS alone are likely to be sufficient to guarantee the user QoS. On the other hand, a highly loaded BS (with a higher ρ) will always look for cooperation to exploit the macro-diversity to counteract the self-congestion and hence provide more gain.

Notice that, at a high γ_{th} value (3 dB) the effect of ρ is not significant. This indifference is more prominent at higher θ values. At low γ_{th} the signal from a single BS (with a better signal quality) is sufficient to guarantee the QoS, and therefore the load is a more critical parameter for cooperation and hence it will have more impact on the effective capacity gain. In contrast, at a high γ_{th}, the MS is more likely in outage at the cell-edge, and hence to maintain the QoS the MS seeks cooperation and the effect of cell load gets a lesser precedence.

The cooperation window is the region in which a MS can gain by using resources from more than one BS. Fig. 7 shows the cooperation window size variation in the proposed approach versus QoS exponent θ at different γ_{th} and ρ. In general, for fixed γ_{th} and ρ, the window size increases with θ. This is because, for a strict QoS requirement the MS looks for more cooperation among the BSs by sharing the traffic even when it is not strictly in the coverage boundary of a cell.

It can be observed that, at a higher γ_{th} (3 dB) the user at the cell-edge will be more in outage as compared to a smaller γ_{th}, and hence to counteract the rate degradation because of the outage it will look for cooperation as early as possible. So, the window size for cooperation is large for a high γ_{th}. It is also observed from Fig. 7 that the cooperation region can be as small as about 1% of the total area at a small γ_{th} (0 dB), a high ρ (0.9), and a small θ (0.1), whereas it is as high as 25%

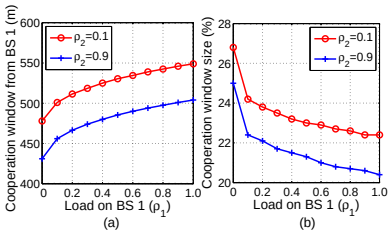


Fig. 8. Effect of load on BS_1 on (a) cooperation window position and (b) cooperation window size. $\theta = 0.9$, $\gamma_{th} = 3$ dB.

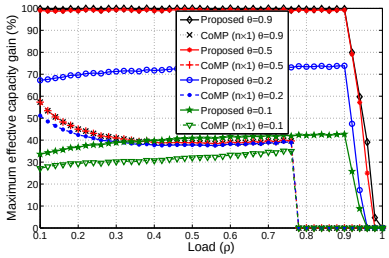


Fig. 9. Maximum effective capacity gain versus average network load in 2-cell cooperation ($n = 2$). Center cell load $\rho_1 = 0.5$, $\gamma_{th} = 3$ dB.

of the total area at $\gamma_{th} = 3$ dB, a low ρ (0.1), and a large θ (1.0). This is because, at a high ρ in both the BSs, the resources for cooperation are less and an attempt to cooperation leads to a higher congestion and hence cooperation region is reduced.

Note from Figs. 6 and 7 that, regardless of the values of ρ and γ_{th}, the maximum effective capacity gain as well as cooperation window size increase with θ. This indicates that the users with stricter QoS will achieve higher gain.

B. Network Parameter Effect

The results on user effect in Section VI-A showed the impact of QoS on the maximum capacity gain and cooperation window size. As noted in Fig. 4, the load on individual BS in a cluster also has an impact on outage and hence on cooperation. Denote, the two cooperating BSs as BS_1 and BS_2. Figs. 8(a) and 8(b) respectively show the effect of BS_1 load on cooperation window position and size in the proposed scheme. BS_2 load is kept constant at two extreme values, 0.1 and 0.9.

For a higher load on BS_2 ($\rho_2 = 0.9$), the window is situated near to BS_1, whereas the window is shifted towards BS_2 at a lower load ($\rho_2 = 0.1$). But, as the load on BS_1 increases, the window starts shifting towards BS_2 (Fig. 8(a)) and also the size reduces (Fig. 8(b)), thereby showing adaptability in traffic sharing and ICIC with cell load. This is because, when a cell carries more traffic than its neighboring cells, it creates more ICI to the users in the neighboring cells, which compels them to include in the sharing zone. Thus, the window will shift towards the other BS to coordinate the ICI. At the same time the heavily loaded BS cannot carry the traffic generated by these shared users. To counteract this additional traffic generation tendency, the window size also reduces such that a lesser number of users remain in the sharing zone.

Fig. 9 presents the effect of load ρ on BSs on the maximum effective capacity gain in CoMP $n \times 1$ and the proposed scheme. The plots in Fig. 9 also reconfirm the observation in Fig. 6 that, at low value of θ and low neighboring cell load ρ the maximum effective capacity gain in CoMP $n \times 1$ can be quite closer to that in the proposed approach. However, at higher values of θ, the relative gain in CoMP is quite small and it decreases with increased ρ due to replicated transmissions. The gain saturates soon with the increased network load, indicating that the increased ICI nullifies the attempt of capacity gain through signal combining (in CoMP) or load sharing (in the proposed scheme). At a high ρ the resources for cooperation is reduced and as a result the capacity gain reduces to zero. However, compared to CoMP, the proposed scheme can withstand a higher network load, because this approach does not increase the total data traffic.

VII. Conclusion

We have proposed a new network-level inter-cell cooperation strategy and analytically derived an optimal solution for resource allocation from the participating BSs. To this end, a next generation network architecture has been considered, and a two-stage queueing structure and its inter-working with the network entities have been presented. Outage performance of the proposed scheme is significantly better than the conventional handoff (HHO, SSHO) schemes. From the capacity gain results it is evident that, if the QoS requirement is high the cooperation window size is large as well as the effective capacity gain is high; at the same time the effect of network load is negligible. The numerical results demonstrate that, the effective capacity gain via network-level cooperation is maximized under high network load and when the acceptable SINR threshold is high. For loose QoS applications, the gain with respect to the non-cooperation case is up to 40%, whereas it is up to 100% for strict QoS applications. Correspondingly, the cooperation window size varies from 1% to as high as 25% of the total area of the cooperating BSs.

Compared to CoMP $n \times 1$ based joint transmission, outage performance of the proposed scheme is poorer when the neighboring cells are lightly loaded, whereas in an interference dominated scenario the proposed approach performs better at the cell-edge. However, on effective capacity the proposed approach significantly outperforms CoMP $n \times 1$ in terms of sensitivity to network traffic load as well as QoS constraint.

Appendix A
Approximate Statistics of SINR at Position x in the ith Cell in (4)

From (3) we have the approximate SINR,

$$\gamma_{i,x} = \frac{1}{\sum_{j=1}^{N_{oc}} \left(\frac{d_{j,x}}{d_{i,x}}\right)^{-l} \cdot 10^{(\zeta_{j,x}-\zeta_{i,x})/10} \cdot \rho_j} \tag{A.1}$$

where $d_{j,x}$ is the distance of the MS at position x in cell i from neighboring BS_j, given by $d_{j,x} = \left\{d_{i,x}^2 + 4R_0^2K^2 - 4R_0Kd_{i,x}\cos(\phi_{i,x} - \phi_{j,x})\right\}^{1/2}$ (Fig. A.1), with K indicating the tier of cells around BS_i; R_0 is the cell radius. $\phi_{j,x} = \frac{360j}{6K}$. Without loss of generality, only

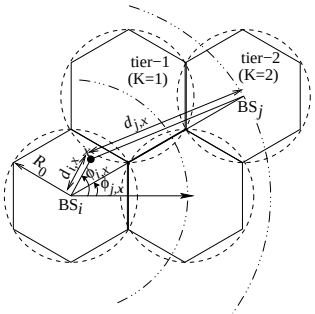


Fig. A.1. Distance measures for computation of approximate SINR.

the immediate neighbor cells are considered ($N_{oc} = 6$) for accounting the co-channel interference, which means $K = 1$.

Denoting $\eta_{ij} = \rho_j \left(\frac{d_{j,x}}{d_{i,x}}\right)^{-l}$ and $\lambda = \frac{\ln 10}{10}$, (A.1) becomes:

$$\gamma_{i,x} = \frac{1}{\sum_{j=1}^{N_{oc}} e^{\ln \eta_{ij} + \lambda(\zeta_{j,x} - \zeta_{i,x})}} \triangleq \frac{1}{\sum_{j=1}^{N_{oc}} e^{\tau_{ij}}} \tag{A.2}$$

We have, $\zeta_{i,x},\ \zeta_{j,x} \sim \mathcal{N}\left(0, \sigma^2\right)$. Denoting the correlation coefficient of $\zeta_{j,x}$ with $\zeta_{i,x}$ as ϱ_{ij},

$$\zeta_{i,x} - \zeta_{j,x} \sim \mathcal{N}\left(0, 2\sigma^2\left(1 - \varrho_{ij}\right)\right), \text{ and}$$
$$\tau_{ij} = \ln \eta_{ij} + \lambda\left(\zeta_{j,x} - \zeta_{i,x}\right) \sim \mathcal{N}\left(\ln \eta_{ij}, 2\sigma^2\lambda^2\left(1 - \varrho_{ij}\right)\right)$$

$$\text{Let} \quad z_{N_{oc}} = \ln \sum_{j=1}^{N_{oc}} e^{\tau_{ij}}, \quad \text{hence } z_{N_{oc}} = -\ln \gamma_{i,x} \tag{A.3}$$

$$\text{Further denote,} \qquad z_2 = \ln\left(e^{\tau_{i1}} + e^{\tau_{i2}}\right) \tag{A.4}$$

which implies $e^{z_{N_{oc}}} = e^{z_2} + \sum_{j=3}^{N_{oc}} e^{\tau_{ij}}$. The first and second moments of z_2 are obtained following the Schwartz and Yeh's approach [37]. Define $\omega = \tau_{i2} - \tau_{i1}$. The mean and variance of ω can be obtained as:

$$\begin{aligned} m_\omega &= m_{\tau_{i2}} - m_{\tau_i} = \ln \eta_{i2} - \ln \eta_{i2} \\ &= \ln\left(\frac{\rho_2}{\rho_1}\left(\frac{d_{2,x}}{d_{1,x}}\right)^{-l}\right) \end{aligned} \tag{A.5a}$$

$$\begin{aligned} \sigma_\omega^2 &= \mathbb{E}\left[\left(\omega - m_\omega\right)^2\right] = \lambda^2 \mathbb{E}\left[\left(\zeta_{2,x} - \zeta_{1,x}\right)^2\right] \\ &= 2\lambda^2\sigma^2\left(1 - \varrho_{21}\right) \end{aligned} \tag{A.5b}$$

$\mathbb{E}[\cdot]$ is the expectation operator. Mean of z_2 is obtained as:

$$m_{z_2} = \ln \eta_{i1} + \mathbb{E}\left[\ln\left(1 + e^{\omega}\right)\right]$$

(A.5a) and (A.5b) indicate that ω is a Gaussian random variable. Therefore,

$$\begin{aligned} \mathbb{E}\left[\ln\left(1 + e^{\omega}\right)\right] &= \int_{-\infty}^{\infty} \frac{\ln\left(1 + e^{\omega}\right)}{\sqrt{2\pi\sigma_\omega^2}} \exp\left[-\frac{\left(\omega - m_\omega\right)^2}{2\sigma_\omega^2}\right] d\omega \\ &= m_\omega Q\left(\frac{m_\omega}{\sigma_\omega}\right) - \frac{\sigma_\omega}{\sqrt{2\pi}} e^{-m_\omega^2/2\sigma_\omega^2} + \sum_{k=1}^{\infty} \frac{(-1)^k}{k} e^{-k^2\sigma_\omega^2/2} \\ &\cdot \left[e^{km_\omega} Q\left(\frac{-m_\omega - k\sigma_\omega^2}{\sigma_\omega}\right) + \sum_{k=1}^{\infty} e^{-km_\omega} Q\left(\frac{m_\omega - k\sigma_\omega^2}{\sigma_\omega}\right)\right] \\ &\triangleq G_1\left(m_\omega, \sigma_\omega\right) \end{aligned}$$

where $Q(\cdot)$ is the Q-function. Also, denoting $\psi_1 = -\frac{\sigma_{\tau_{i1}}}{\sigma_\omega}$,

$$\sigma_{z_2}^2 = \sigma_{\tau_{i1}}^2 - G_1^2\left(m_\omega, \sigma_\omega\right) + G_2\left(m_\omega, \sigma_\omega\right) - 2\psi_1^2 G_3\left(m_\omega, \sigma_\omega\right) \tag{A.6}$$

$$G_2\left(m_\omega, \sigma_\omega\right) = \int_{-\infty}^{\infty} \frac{\ln^2\left(1 + e^{\omega}\right)}{\sqrt{2\pi\sigma_\omega^2}} \exp\left[-\frac{\left(\omega - m_\omega\right)^2}{2\sigma_\omega^2}\right] d\omega$$

$$G_3\left(m_\omega, \sigma_\omega\right) = \int_{-\infty}^{\infty} \frac{\sigma_\omega e^{\omega}}{\left(1 + e^{\omega}\right)\sqrt{2\pi}} \exp\left[-\frac{\left(\omega - m_\omega\right)^2}{2\sigma_\omega^2}\right] d\omega$$

$$\begin{aligned} &= \sigma_\omega^2 \sum_{k=0}^{\infty} (-1)^k \left[e^{m_\omega(k+1) + (k+1)^2\sigma_\omega^2/2} Q\left(\frac{-m_\omega - \sigma_\omega^2(k+1)}{\sigma_\omega}\right)\right. \\ &\left. + e^{-km_\omega + k^2\sigma_\omega^2/2} Q\left(\frac{m_\omega - \sigma_\omega^2 k}{\sigma_\omega}\right)\right] \end{aligned}$$

In order to extend the process of obtaining the distribution of sum of more than two log-normally distributed random variables, from (A.3), let $\delta_{N_{oc}} = \tau_{iN_{oc}} - z_{N_{oc}-1}$, for $N_{oc} \geq 3$. Then, the mean and variance of $\delta_{N_{oc}}$ are obtained as:

$$m_{\delta_{N_{oc}}} = m_{\tau_{iN_{oc}}} - m_{z_{N_{oc}-1}} = \ln \eta_{iN_{oc}} - m_{z_{N_{oc}-1}}$$

$$\begin{aligned} \sigma_{\delta_{N_{oc}}}^2 =& \sigma_{\tau_{iN_{oc}}}^2 + \sigma_{z_{N_{oc}-1}}^2 - 2\varrho_{\tau_{iN_{oc}}, z_{N_{oc}-1}} \sigma_{\delta_{N_{oc}}} \sigma_{z_{N_{oc}-1}}, \\ =& \lambda^2\sigma^2\left(1 - \varrho_{iN_{oc}}\right) + \sigma_{z_{N_{oc}-1}}^2 \\ &- 2\lambda\left(\varrho_{\zeta_{N_{oc},x}, z_{N_{oc}}} + \varrho_{\zeta_{i,x}, z_{N_{oc}-1}}\right)\sigma\sigma_{z_{N_{oc}-1}} \end{aligned}$$

where $\sigma_{z_{N_{oc}-1}}$ is obtained recursively following the steps in (A.4) through (A.6), and $\varrho_{\zeta_{i,x}, z_j}$ is the cross-correlation coefficient between $\zeta_{i,x}$ and z_j.

Thus, with the notation $\psi_N = -\frac{\sigma_{z_{N_{oc}-1}}}{\sigma_{\delta_{N_{oc}}}}$, we have,

$$\begin{aligned} m_{z_{N_{oc}}} =& m_{z_{N_{oc}-1}} + G_1\left(m_{\delta_{N_{oc}}}, \sigma_{\delta_{N_{oc}}}\right) \\ \sigma_{z_{N_{oc}}}^2 =& \sigma_{z_{N_{oc}-1}}^2 - G_1^2\left(m_{\delta_{N_{oc}}}, \sigma_{\delta_{N_{oc}}}\right) + G_2\left(m_{\delta_{N_{oc}}}, \sigma_{\delta_{N_{oc}}}\right) \\ &- 2\psi_N^2 G_3\left(m_{\delta_{N_{oc}}}, \sigma_{\delta_{N_{oc}}}\right) \end{aligned}$$

Hence, recursively from:

$$m_{z_1} \equiv m_{\tau_{i1}} = \ln \eta_{i1} \quad \text{and} \quad \sigma_{z_1} \equiv \sigma_{\tau_{i1}} = 2\sigma^2\lambda^2\left(1 - \varrho_{i1}\right)$$

from the relationship in (A.3) we obtain the mean and variance of the approximately log-normally distributed $\gamma_{i,x}$ as:

$$m_{(\ln \gamma_{i,x})} = -m_{z_{N_{oc}}} \quad \text{and} \quad \sigma_{(\ln \gamma_{i,x})} = \sigma_{z_{N_{oc}}}$$

APPENDIX B

Here we show that the solution obtained from the iterative procedure satisfies the KKT conditions in (16) and (17). The optimal resource parameters $S_i^{u\#}, S_j^{u\#}$ are obtained in each iteration with $\eta_j = 0\ \forall\ j$. The relationship between $S_i^{u\#}$ and $S_k^{u\#}$ from (16a) and (16b) is

$$\begin{aligned} \frac{(1 - P_k)\theta^u r e^{-\theta^u r S_k^{u\#}}}{e^{-\theta^u r S_k^{u\#}}(1 - P_k) + P_k} &= \frac{(1 - P_i)\theta^u r e^{-\theta^u r S_i^{u\#}}}{e^{-\theta^u r S_i^{u\#}}(1 - P_i) + P_i} \\ \text{or, } P_i(1 - P_k)e^{-\theta^u r S_k^{u\#}} &= P_k(1 - P_i)e^{-\theta^u r S_i^{u\#}} \end{aligned} \tag{B.1}$$

For some $k \in \{j = 1, \cdots, N', j \neq i,\}$, having $S_k^{u\#} > S(1 - \rho_k)$, we set $S_k^{u*} = S(1 - \rho_k) = S_k^{u\#} - \beta_k, \beta_k > 0$. Let the optimal value of S_i^u at the final iteration be $S_i^{u*} = S_i^{u\#} +$

α. The value of η_k is computed from (16b) for the optimal resource allocation S_i^{u*} and S_k^{u*} as

$$\eta_k = \frac{(1-P_k)\theta^u r e^{-\theta^u r S_k^{u*}}}{e^{-\theta^u r S_k^{u*}}(1-P_k)+P_k} - \frac{(1-P_i)\theta^u r e^{-\theta^u r S_i^{u*}}}{e^{-\theta^u r S_i^{u*}}(1-P_i)+P_i}$$

Substituting $S_i^{u*} = S_i^{u\#} + \alpha$ and $S_k^{u*} = S_k^{u\#} - \beta_k$, in the above equation and using (B.1), we obtain

$$\eta_k = \frac{P_i(1-P_k)\theta^u r e^{-\theta^u r S_k^{u\#}}(e^{\theta^u r \beta_k} - e^{-\theta^u r \alpha})}{(e^{-\theta^u r S_i^{u*}}(1-P_i)+P_i)(e^{-\theta^u r S_k^{u*}}(1-P_k)+P_k)} \quad \text{(B.3)}$$

From (19a), we compute the value of S_i^u in the current iteration $S_i^{u\#}$, and in the next iteration $S_i^{u\dagger}$, with $\{\kappa\}$ set of k's violating the constraint in the current iteration as:

$$S_i^{u\#} = \frac{S^u}{N'+1} + \frac{1}{(N'+1)\,\theta^u r} \ln \left[\frac{(1-P_i)^{N'} \prod_{\substack{j=1 \\ j\neq i}}^{N'} P_j}{P_i^{N'} \prod_{\substack{j=1 \\ j\neq i}}^{N'} (1-P_j)} \right]$$

$$S_i^{u\dagger} = \frac{S^u - \sum_{k\in\kappa} S_k^{u*}}{N' - |\kappa| + 1} + \frac{1}{(N'-|\kappa|+1)\,\theta^u r} \ln \left[\frac{(1-P_i)^{(N'-|\kappa|)} \prod_{\substack{j=1 \\ j\neq i,k \\ k\in\kappa}}^{N'-|\kappa|} P_j}{P_i^{(N'-|\kappa|)} \prod_{\substack{j=1 \\ j\neq i,k \\ k\in\kappa}}^{N'-|\kappa|} (1-P_j)} \right]$$

Define $\alpha_1 = S_i^{u\dagger} - S_i^{u\#}$. Substituting $S_k^{u*} = S_k^{u\#} - \beta_k$, $\forall\ k \in \{\kappa\}$, and using value of $S_k^{u\#}$ from (19b), we get

$$\alpha_1 = \frac{\sum_{k\in\kappa} \beta_k}{(N' - |\kappa| + 1)} \quad \text{(B.4)}$$

We observe that α_1 is positive. From this, we infer that at each iteration the optimal value computed for S_i^u increases from the previous iteration, and hence the value in the final iteration $\alpha > 0$. As β and α are positive, from (B.3), η_k is always positive. Hence, the KKT conditions are satisfied for the final value obtained from the proposed iterative scheme.

Appendix C
Proof of Equation (22)

Assume $f(x) = e^{-ax}$, $a > 0$, and two functions $F(S^u) = f(x)w_1 + f(y)w_2$ and $F^2\left(\frac{S^u}{2}\right) = \left[f(\frac{x}{2})w_1 + f(\frac{y}{2})w_2\right]^2$, where $w_1 + w_2 = 1$ and $w_1, w_2 > 0$. The following observations are made: (i) $f(x)$ is a decreasing function of x; (ii) $0 \le f(x) \le 1 \ \ \forall \ \ x \ge 0$; (iii) $f(y) > f(x) \ \ \forall \ \ y < x$; (iv) $f(x) = f^2(\frac{x}{2})$. Let

$$\psi = F(S^u) - F^2\left(\frac{S^u}{2}\right), \quad \text{where} \quad \text{(C.1)}$$

$$\begin{aligned} F(S^u) &= f(x)w_1 + f(y)w_2 \\ F^2\left(\frac{S^u}{2}\right) &= \left[f\left(\frac{x}{2}\right)w_1 + f\left(\frac{y}{2}\right)w_2\right]^2 \end{aligned}$$

Substituting $w_1 = 1 - w_2$ and simplifying we have,

$$\begin{aligned} F(S^u) &= f^2\left(\frac{x}{2}\right) + \left\{f^2\left(\frac{y}{2}\right) - f^2\left(\frac{x}{2}\right)\right\} w_2 \\ F^2\left(\frac{S^u}{2}\right) &= \left[f\left(\frac{x}{2}\right) + \left\{f\left(\frac{y}{2}\right) - f\left(\frac{x}{2}\right)\right\} w_2\right]^2 \end{aligned}$$

Hence, from (C.1), after solving we have,

$$\psi = (1-w_2)\left[f^2\left(\frac{y}{2}\right) - f^2\left(\frac{x}{2}\right)\right]$$

In our case, $a = \theta^u r$, $w_2 = P_i$, $y = 0$, and $x = S^u$. Therefore,

$$\begin{aligned} F(S^u) &= e^{-\theta^u r S^u}(1-P_i) + P_i \\ F^2\left(\frac{S^u}{2}\right) &= \left[e^{-\theta^u r \frac{S^u}{2}}(1-P_i) + P_i\right]^2 \end{aligned}$$

So, $x > y$ and $1 > w_2 > 0$. Hence $f^2(\frac{y}{2}) > f^2(\frac{x}{2})$ and $\psi > 0$, and the proof of (22) follows.

References

[1] E. Dahlman, S. Parkvall, and J. Skold, *3G Evolution: HSPA and LTE for Mobile Broadband*, 2nd ed. Academic Press, 2008.

[2] I. F. Akyildiz, D. M. Gutierrez-Estevez, and E. C. Reyes, "The evolution to 4G cellular systems: LTE-Advanced," *Elsevier Phy. Commun.*, vol. 3, no. 4, pp. 217–244, Dec. 2010.

[3] D. Zhang and L. Cuthbert, "Dynamic subcarrier and power allocation in LTE networks," in *Proc. IEEE Wireless Commun., Networking, and Mob. Comput.*, Beijing, China, Sept. 2009.

[4] S.-E. Elayoubi, O. B. Haddada, and B. Fourestie, "Performance evaluation of frequency planning schemes in OFDMA-based networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1623–1633, May 2008.

[5] T. Novlan, J. G. Andrews, I. S. R. K. Ganti, and A. Ghosh, "Comparison of fractional frequency reuse approaches in the OFDMA cellular downlink," in *Proc. IEEE GLOBECOM*, Miami, FL, USA, Dec. 2010.

[6] R. Weber, A. Garavaglia, M. Schulist, S. Brueck, and A. Dekorsy, "Self-organizing adaptive clustering for cooperative multipoint transmission," in *Proc. IEEE VTC - Spring*, Budapest, Hungary, May 2011.

[7] V. Garcia, N. Lebedev, and J.-P. Gorce, "Capacity outage probability for multi-cell processing under Rayleigh fading," *IEEE Commun. Letters*, vol. 15, no. 8, pp. 801–803, Nov. 2011.

[8] N. D. Tripathi, J. H. Reed, and H. F. VanLandinoham, "Handoff in cellular systems," *IEEE Pers. Commun.*, vol. 5, no. 6, pp. 26–37, 1998.

[9] D. Wong and T. J. Lim, "Soft handoffs in CDMA mobile systems," *IEEE Pers. Commun.*, vol. 4, no. 6, pp. 6–17, 1997.

[10] B. G. Lee, D. Park, and H. Seo, *Wireless Communications Resource Management*, 1st ed. Wiley-IEEE Press, Dec. 2008.

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

[12] N. Zhang and J. M. Holtzman, "Analysis of handoff algorithms using both absolute and relative measurements," *IEEE Trans. Veh. Technol.*, vol. 45, no. 1, pp. 174–179, Feb. 1996.

[13] A. E. Leu and B. L. Mark, "A discrete-time approach to analyze hard handoff performance in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1721–1733, Oct. 2004.

[14] ——, "Discrete-time level-crossing analysis of soft handoff performance in cellular networks," *IEEE Trans. Inform. Th.*, vol. 52, no. 7, pp. 3283–3290, July 2006.

[15] "3GPP Long-Term Evolution (LTE)." [Online]. Available: http://www.3gpp.org/Highlights/LTE/LTE.htm

[16] L. Bajzik, P. Horvath, L. Korossy, and C. Vulkan, "Impact of intra-LTE handover with forwarding on the user connections," in *Proc. IEEE Mobile and Wireless Commun. Summit*, Budapest, Hungary, July 2007.

[17] L. Tian, J. Li, Y. Huang, J. Shi, and J. Zhou, "Seamless dual-link handover scheme in broadband wireless communication systems for high-speed rail," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 708–718, May 2012.

[18] D. Aziz and R. Sigle, "Improvement of LTE handover performance through interference coordination," in *Proc. IEEE VTC - Spring*, Barcelona, Spain, Apr. 2009.

[19] L. G. de R. Guedes and M. D. Yacoub, "Overlapping cell area in different fading conditions," in *Proc. IEEE Veh. Technol. Conf.*, Chicago, IL, USA, July 1995.

[20] T. K. Y. Lo, "Maximum ratio transmission," *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458–1461, Oct. 1999.
[21] A. Goldsmith, *Wireless Communications*. Cambridge Press, 2005.
[22] H. Lee, H. Son, and S. Lee, "Semisoft handover gain analysis over OFDM-based broadband systems," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1443–1453, Mar. 2009.
[23] J. Chang, Y. Li, S. Feng, H. Wang, C. Sun, and P. Zhang, "A fractional soft handover scheme for 3GPP LTE-advanced system," in *Proc. IEEE Intl. conf. Commun.*, Dresden, Germany, June 2009.
[24] A. F. Coskun and O. Kucur, "Performance analysis of hybrid transmit antenna selection/maximal-ratio transmission in Nakagami-m fading channels," *Wiley Wireless Commun. and Mob. Comput.*, vol. 13, pp. 1234–1245, 2013.
[25] D. Lee, H. Seo, B. Clerckx, E. Hardouin, D. Mazzarese, S. Nagata, and K. Sayana, "Coordinated multipoint transmission and reception in LTE-Advanced: Deployment scenarios and operational challenges," *IEEE Commun.*, vol. 50, no. 2, pp. 148–155, Feb. 2012.
[26] J. Lee, Y. Kim, H. Lee, B. L. Ng, D. Mazzarese, J. Liu, W. Xiao, and Y. Zhou, "Coordinated multipoint transmission and reception in LTE-Advanced systems," *IEEE Commun.*, vol. 50, no. 11, pp. 44–50, Nov. 2012.
[27] "3GPP TR 36.819: Coordinated multi-point operationfor LTE physical layer aspects." [Online]. Available: http://http://www.3gpp.org/DynaReport/36819.htm
[28] M. R. R. Kumar, S. Bhashyam, and D. Jalihal, "Throughput improvement for cell-edge users using selective cooperation in cellular networks," in *Proc. IEEE/IFIP Conf. Wireless and Optical Commun. Networks*, Surabaya, Indonesia, May 2008.
[29] L. Xu, K. Yamamoto, H. Murata, and S. Yoshida, "Adaptive base station cooperation and subchannel reallocation at cell edge in cellular networks with fractional frequency reuse," in *Proc. IEEE Conf. Pers. Indoor and Mobile Radio Commun.*, Tokyo, Japan, Sept. 2009.
[30] A. Nosratinia, T. E. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun.*, vol. 42, no. 10, pp. 74–80, Oct. 2004.
[31] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Th.*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.
[32] H. N. Nguyen and I. Sasase, "Downlink queuing model and packet scheduling for providing lossless handoff and QoS in 4G mobile networks," *IEEE Trans. Mob. Comput.*, vol. 5, no. 5, pp. 452–462, May 2006.
[33] A. Mihovska, C. Wijting, R. Prasad, S. Ponnekanti, Y. Awad, and M. Nakamura, "A novel flexible technology for intelligent base station architecture support for 4G systems," in *Proc. IEEE Conf. Wireless Pers. Multimedia Comm.*, Honolulu, HI, USA, Oct. 2002.
[34] B. G. Evans and K. Baughan, "Visions of 4G," *J. Electron. and Comm. Eng.*, vol. 12, no. 6, pp. 293–303, Dec. 2000.
[35] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, Jr., "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, Apr. 2009.
[36] D. Kim, J. Y. Ahn, and H. Kim, "Downlink transmit power allocation in soft fractional frequency reuse systems," *ETRI J.*, vol. 33, no. 1, pp. 1–12, Feb. 2011.
[37] S. C. Schwartz and Y. S. Yeh, "On the distribution function and moments of power sums with log-normal components," *Bell Sys. Tech. J.*, vol. 61, no. 7, pp. 1441–1462, Sept. 1982.
[38] P. Mary, M. Dohler, J.-M. Gorce, and G. Villemaud, "Packet error outage for coded systems experiencing fast fading and shadowing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 574–585, Feb. 2013.

ACKNOWLEDGMENT

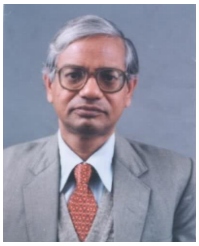
This work has been supported by the Department of Science and Technology (DST) under the grant no. SR/S3/EECE/0122/2010.

BIOGRAPHIES

Satyam Agarwal (S'13) received his B.Tech. in Electronics and Communication from Thapar University, India, in 2010 and M.Tech. in Wireless Communications and Networks, from the Department of Electrical Engineering at IIT Kanpur in 2012. He is currently working towards the Ph.D. degree with the Department of Electrical Engineering at IIT Delhi. His research interests include cooperative wireless communications and link layer protocol designs in cognitive radio networks. He is a student member of IEEE and IEEE Communications Society.

Swades De (S'02-M'04) received his PhD in Electrical Eng. from the State Univ. of New York at Buffalo in 2004. He is currently an Associate Professor of Electrical Eng. at IIT Delhi. He was an Assistant Professor of Electrical and Computer Eng. at New Jersey Institute of Technology (2004-2007). He worked as a post-doctoral researcher at ISTI-CNR, Pisa, Italy (2004), and has nearly 5 years industry experience in India on telecom hardware and software development (1993-1997, 1999). His research interests include performance study, resource efficiency in wireless networks, broadband wireless access, and communication and systems issues in optical networks. Dr. De currently serves as an Associate Editor of IEEE Communications Letters and Springer Photonic Network Communications journal. He is a member of IEEE, IEEE ComSoc and CompSoc, and IEICE.

Satish Kumar completed his M.S. from the Bharti School of Telecom, IIT Delhi, in 2011. He is currently associated with Qualcomm India Pvt. Ltd. in the field of wireless technology. His research interests include cooperative wireless communications, scheduling techniques in WCDMA and OFDMA networks, and LTE/WCDMA small cell optimization.



Hari Mohan Gupta received B.E. (Electronics and Communications) from University of Roorkee (now IIT Roorkee), M. Tech (UHF and Microwave Engineering) from IIT Kharagpur and Ph.D. (Electrical Engineering) from IIT Kanpur.

He joined the faculty of Electrical Engineering at Indian Institute of Technology, Delhi in 1973, where he is a professor since 1986. Prof. Gupta held positions of Head of the Department, Dean Undergraduate Studies, and Coordinator, Bharti School of Telecommunication at IIT Delhi. He held faculty positions at Mc Gill University, Montreal, Canada, and at Drexel University, Philadelphia, USA. He has been an academic visitor to University of Maryland, College Park, USA; Media Lab at Massachusetts Institute of Technology, Cambridge, USA; Swiss Federal Institute of Technology, Lausanne, Switzerland and several British Universities. He has been Vice-President of System Society of India, and is a founding member of Association for Security of Information Systems (ASIS). He is also President of Institution of Communication Engineers and Information Technologists (ICEIT). His research interests include mobile computing, satellite communications, multimedia information processing and photonic systems.