Spectrally-Efficient Cooperative Video Delivery in 5G Heterogeneous Wireless Networks

Antonios Argyriou*, Melike Erol-Kantarci†, Yanwei Liu‡

*Department of Electrical and Computer Engineering, University of Thessaly, Volos, 38221, Greece.
†School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, Canada.
‡Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100190, China.

***Abstract*—Heterogeneous cellular networks (HetNets), even in 5G systems, are plagued by the problem of intra-cell interference, i.e., a strong interfering transmission from a macro BS (MBS) interferes at all the users associated to the several small cells of the HetNet. This phenomenon may significantly reduce the spectral efficiency. With the dominating presence of bandwidth-hungry video in cellular systems, making the most out of the available spectrum resources is critical for the mobile operators. In this paper, we present a cooperative protocol that can ensure spectral efficient delivery of high quality video in HetNets. The key characteristic is that the transmitted information from a small cell base station is decoded at several users in the small cell when the MBS also transmits. The users that decode act subsequently as relays to the final user. Our novel protocol is supported by a new optimization framework for resource allocation in the HetNet and rate allocation at the video source. Performance results indicate that the presence of users, a previously unexploited dimension in HetNets, offers significant SE gains.**

***Index Terms*—Heterogeneous cellular networks, small cells, intra-cell interference, video streaming, video distribution, DASH, rate allocation, resource allocation.**

I. Introduction

Wireless bandwidth is a very valuable resource due to its scarcity. The lack of bandwidth in emerging 5G wireless cellular systems will be addressed with the deployment of heterogenous cellular networks (HetNets). The main reason that HetNets have such potential is that they improve the spatial reuse by deploying low power base stations (BS) like pico BS (PBS) and femto BS (FBS), that create around them small cells, picocells and femtocells respectively. In Fig. 1 the macro BS (MBS) and a PBS can transmit simultaneously and serve two users. If only the MBS was present in this setup it could serve only one user in the same time/frequency resource. This approach improves the spectral efficiency (SE) of the HetNet.

The need for higher bandwidth in 5G cellular systems is propelled by an explosive increase in the demand for high quality video [1]. Thus, the mobile network operators (MNOs) will have to face a situation where the small cells transmit a high volume of video data. Consequently it is of outmost importance to improve not only the SE but also the *video spectral efficiency* (VSE), that is the delivered video quality per wireless resource unit.

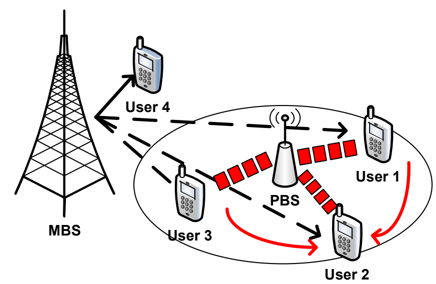


Fig. 1. System model. Red blocks indicate video packets originating from the PBS. Dashed lines indicate intra-cell interference from the MBS to the PBS users. Besides target user 2, users 1 and 3 also receive the interfering signal plus the PBS video packets. Then they forward to the desired user (red solid lines).

Even though the deployment of HetNets seems the most promising avenue for 5G system deployments, they suffer from a specific problem which is *intra-cell* interference. Intra-cell interference is generated from the MBS and impacts the low power BSs. Fig. 1 highlights this scenario where the MBS interferes with transmissions of the PBS to its users. It is clear from this simple example that intra-cell interference should be minimized to improve the SE of the complete HetNet. One technique for handling this type of interference in HetNets, is time-domain resource partitioning (TDRP) where the macro-cells shut off their transmission for a subset of the available time-domain resources (time slots) [2]. This technique was recently became a standard with the use of almost blank subframes (ABS) in 3GPP LTE under a generic scheme called enhanced ICIC (eICIC) [3]. With TDRP the objective is to identify the proper allocation of time-domain resources so that either interference is minimized at the users [4], or the overall HetNet throughput is maximized [2].

Video communication under this new HetNet TDRP framework was only recently investigated and the results showed that significant gains in terms of video quality can be obtained if the level of TDRP is carefully optimized [5]. However, existing research is not concerned with one feature of HetNets that may be proven to be critical for video delivery [2], [4], [5], and this is the *user dimension*. Small cell users can help by delivering video data through cooperative protocols. While video communication in cooperative networks has been studied extensively, in the context of HetNets that introduce their own unique problem of intracell interference, the problem is unexplored. Research works in small cells typically consider cooperation through contribution of the available storage space for improving video delivery [6].

In this paper our goal is to improve the quality of the delivered video in 5G HetNets. To this aim we consider a state-of-the-art HetNet that employs TDRP across the tiers together with state-of-the-art resource allocation [2], [4]. The HetNet allows device-to-device (D2D) communication, a feature that is already part of LTE Rel. 12 [3]. The idea we put forward can be explained with the help of Fig. 1 by considering the transmission of a single video packet. What we propose, is to allow the PBS to transmit video data to a user during a regular slot (RS) that the MBS also transmits. Then, to exercise successive interference cancellation (SIC) at other active users associated to a PBS in order to recover at several of them the video packet transmitted from the PBS. In our particular example in Fig. 1, the PBS users 2 and 3 will exercise SIC when the video packet is transmitted to user 1 from the PBS. Even though the packet may not be possible decode at the desired user, it may be possible to decode at other active small cell users. The reason is that SIC performance is sensitive to the instantaneous power of the two interfering signals [7]. During an ABS, where the MBS does not transmit, the cooperating users forward their received signals to the desired user by applying a distributed space-time code (DSTC). DSTC offers a *cooperative diversity* gain that improves the reliability of the transmission allowing thus a higher rate [8]. A flavor of the cooperative protocol we discuss in this work was presented in [9], while in this paper we improve its design and tailor it to video communication.

Based on the idea we explain explained before, the concrete contributions of this work are:

- We propose a cooperative protocol for spectral efficient video delivery in HetNets where intra-cell interference dominates.
- We also embed this new video-aware cooperative protocol into a new video-aware rate and resource allocation optimization framework that leads to a convex problem.

Our system extracts significant gains from the previously unlocked user dimension in HetNets. In particular video quality improvements in the order of more than 20% relative to classic cooperative schemes can be observed for a network of 8 PBS and 100 users. Also, as the user density is increased from a number of 100 to 200 users the performance of our scheme relative to non-cooperative schemes is improved from 25% in the first case to 37.5%. Hence, we have significant gain with respect to both non-cooperative and cooperative schemes.

II. System Model and Assumptions

Network Model: The HetNet that we study in this paper consists of a single macrocell with a MBS, the set $\mathcal{J}$ of PBSs, and the users. Each BS j communicates with the set of users $\mathcal{N}_j$. The MBS shuts off its transmissions for a fraction of the resources (ABS) that is denoted with η. During these time-domain resources the picocells in the network transmit and generate interference to every active user in the network (including all the users in all the other picocells). Hence, across BSs of the same tier (PBSs in our case) we consider that resources are reused which is the main benefit of small cells. The aggregate average interference power that a node receives *from all the small cells* during the ABS is $I_{\text{ABS},i}$ Watts (this value also accounts for the interference from relays that operate in the cooperative mode). For the non-blank resources (RS), the MBS and PBSs jointly transmit while the total level of the interference power that a node receives is $I_{\text{RS},i}$ Watts.

Channel Model: Every node has a single omni-directional antenna that can be used in half-duplex mode for transmission and reception. All the channels, are considered to be independent block-fading Rayleigh. The channel coefficients are quasi-stationary, that is they remain constant for the coherence period of the channel that is equal to the transmission length of the complete subframe (Fig. 2). We denote the channel from the j-th BS to the r-th user as $h_{j,r}$, and the channel from the r-th user to user i as $h_{r,i}$.[1] We adopt the LTE frame structure where each subframe is divided into two time slots during which an *information block* of L symbols is transmitted (Fig. 2). We also consider the path loss and shadowing effects according to the LTE channel model [3]: Distance-dependent path loss is given by $L(d) = I + 37.6\log_{10}(d)$, where d is the distance in Km, I=128.1 [3], and the shadowing standard deviation is 8 dB. Additive white Gaussian noise (AWGN) is assumed at every receiver with variance σ^2.

User Model: The users associate to a BS by using an SINR biasing rule [2], that is they associate to the BS that provides them the highest SINR. We also assume that users are static, and so we can calculate their average achievable rate under the protocols we define in the next section.

Scalable Video Delivery: We assume that the PBSs have the video files of each user cached in a scalable video coding (SVC) format that consists of many dependent layers. The delivery and caching of video layers entails costs but also reduces bandwidth consumption, energy, and user delay as shown in [10]. A more elaborate discussion is needed for the video delivery phase. Since video today is typically contained in units that have relatively large playback time duration, e.g. 10-second DASH segments [11] and not packets, these data can be delivered in a relatively large time period (relative to the variations of the fading channel) so that the complete unit is decodable (e.g., see this methodology proposed in [12]). This means that for the delivery of these units to user i what matters is the average data rate C_i experienced over several channel realizations. Furthermore, since the users are static their average data rate will be constant. To correlate the streaming rate and the channel rate we proceed as follows. If the different scalable descriptions for video i are indexed by l, the average rate experienced by user i must be at least

$$S_{i,l}/(T_i + B_i) \leq C_i \text{ bits/sec,} \tag{1}$$

[1] We distinguish the notation for the users that act as relays to facilitate the description of the proposed protocol.

where $S_{i,l}$ is the size of the l-th description from the scalable video file i, T_i is the total playback time of the file and B_i is the startup buffering delay [5]. As it will become clear in the next section, and also in our detailed problem formulation, we will relate C_i and the rate allocated to the user based on several practical constraints of our system. The second issue that has to be modeled is the differentiation of the videos. To allocate different resources and rate to different videos, we distinguish the videos based on the average mean-square error (MSE) reduction that is achieved per delivered video bit [5]. Hence, we define the video quality of video file i to be the ratio of the MSE reduction that is achieved when we receive all the frames from all the layers divided by its size in bytes. This ratio is denoted with parameter α_i [5], [12].

Power Consumption Model: To ensure that we compare systems fairly we do the following. The PBS uses in direct transmission mode a transmission power equal to P_{PBS}. This is set equal to the transmission power of the picocell BS and the transmission power of all the used relays in the cooperative mode. These are denoted as $P_{\text{PBS}}^{\text{COOP}}$ and $P_{\text{R}}^{\text{COOP}}$ respectively (i.e., $P_{\text{PBS}} = P_{\text{PBS}}^{\text{COOP}} + P_{\text{R}}^{\text{COOP}}$). We are are actually evem more strict regarding the efficiency of the cooperative modes. That is we also consider the power consumed at the receiver of a relay user when interfering signals are overheard.[2] For the remaining of this paper our assumption is that each relay user has available a power budget that is reduced by 50%. Thus, by allocating power equally across the PBS and the relays, we have that $P_{\text{PBS}}^{\text{COOP}} = P_{\text{PBS}}/2$ is *the total available power at all the participating relays*.

CSI: We also assume that users provide only average channel statistics to the base stations.

III. Proposed Transmission Mode

SIC during RS. With the proposed scheme named DSTC of interfering transmissions (DSTCIF), the transmissions of the PBS and MBS during the RS are overheard by the users in the picocell [9]. Hence the signal model during the first RS slot at relay user r is (similar for the second slot):

$$y_{\text{RS},r}(1) = h_{\text{PBS},r}x_1 + h_{\text{MBS},r}x_{\text{MBS}} + n_{\text{RS,r}} + w_r \tag{2}$$

In this expression w_r is the AWGN sample at the relay. The remaining aggregate interference power that a node r receives is denoted as $I_{\text{RS},r} = \mathbb{E}[|n_{\text{RS,r}}|^2]$. The aggregate interfering signals typically cannot be decoded. This is because it is impossible in practice to estimate the channel from mutiple sources located inside the macrocell and outside. Similar with other works [13], we consider that this interference is an aggregate term that constant can can be measured by the respective user.[3] The same expression holds for the transmission of the second information block x_2 from the PBS during the second slot of the RS. After the end of a regular subframe, a relay user attempts to decode the two blocks x_1, x_2 by employing

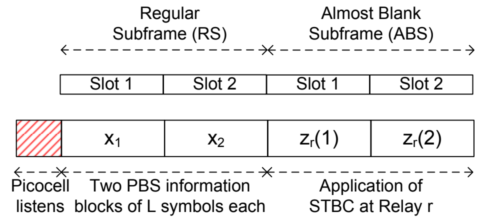


Fig. 2. Frame structure and behavior of the proposed transmission mode during a consecutive regular and a blank subframe. The frame structure is designed to be reminiscent of the LTE-A frame structure. Application of STBC in the proposed protocol occurs in the ABS. During the ABS the relays transmit simultaneously in the same time slot.

ordered SIC (OSIC) [7]. That is, the block with the highest energy/bit is decoded first, while the second interfering block is treated as noise. If there was no interference from the MBS the following condition must be true so that block x_1 from PBS is decoded:

$$\log_2(1 + \frac{P_{\text{PBS}}|h_{\text{PBS},r}|^2}{I_{\text{RS},r} + \sigma^2}) \geq m \Rightarrow \frac{P_{\text{PBS}}|h_{\text{PBS},r}|^2}{(I_{\text{RS},r} + \sigma^2)(2^m - 1)} \geq 1$$

The fractional term in the RHS of the last derivation is essentially the normalized SINR/bit that is required for decoding m bits/symbol. One can obtain a similar expression for the MBS data and by assuming $\mathbb{E}[|x_1|^2] = \mathbb{E}[|x_{\text{MBS}}|^2] = 1$, the following condition can be used so that x_1 is decoded first:

$$\frac{P_{\text{PBS}}|h_{\text{PBS},r}|^2}{2^m - 1} > \frac{P_{\text{MBS}}|h_{\text{MBS},r}|^2}{2^{m_{\text{MBS}}} - 1} \tag{3}$$

When x_1 is decoded successfully, the algorithm subtracts it from the aggregate signal $y_{\text{RS},r}(1)$.[4] This OSIC scheme is applied similarly for the signal $y_{\text{RS},r}(2)$. The implementation of the cancellation algorithm occurs at the level of *information blocks*. The verification of successful decoding of information block x_1 is done with an error correcting cyclic redundancy check (CRC) code.[5] Thus, upon the successful decoding, and with CSI at the relay (in this example $h_{\text{PBS},r}$), we can completely remove/cancel a complete block from the aggregate received signal $y_{\text{RS},r}(1)$.

The relay will transmit different blocks of information according to the decoding result. The relay will complete this action in the two slots of the ABS as presented in Fig. 2. The information blocks that the relays transmit as $q_{r,1}$, that takes the value $q_{r,1}$=x_1, or $q_{r,1}$=0 if the block is not decoded (similarly for $q_{r,2}$). This notation indicates that a relay user has either the decoded information available or nothing at all if SIC fails.

Randomized Distributed Space-Time Coding (R-DSTC). The DSTC operations at the relay are exercised for the block of information denoted as q_r. For easier understanding of the concept we select an orthogonal 2x2 Alamouti code for

[2]For typical modern LTE chips we notice that the transmitter (TX) and receiver (RX) power levels are similar.

[3]LTE Rel. 8 already implements the communication of the power of the local interference though the high interference indicator (HII).

[4]It is possible that different rules are used for selecting the symbol to be decoded first or even a completely different IC scheme. Our central concept is to cancel the interference of the MBS and extract the PBS data block.

[5]LTE uses a CRC for multiple resource blocks that is called the transport control block.

q. Hence, over the two consecutive slots 1,2 the r-th relay transmits:

$$z_r(1) = \boldsymbol{c}_r[g_{r,1}q_{r,1} \quad g_{r,2}q_{r,2}]^T = g_{r,1}c_{r,1}q_{r,1} + g_{r,2}c_{r,2}q_{r,2}$$
$$z_r(2) = \boldsymbol{c}_r[-g_{r,2}q_{r,2}^* \quad g_{r,1}q_{r,1}^*]^T = g_{r,1}c_{r,2}^*q_{r,1}^* - g_{r,2}c_{r,1}^*q_{r,2}^*$$

In this last equation $\boldsymbol{c}_r$ indicates the randomization vector that its purpose is to simplify the implementation of DSTC in a distributed form [8]. The parameters $g_{r,1}, g_{r,2}$ indicate power scaling for data blocks $q_{r,1}, q_{r,2}$. The power that is available for all the relays is split equally according to P_{R}/M, where M indicates the number of relays that have successfully decoded. The coefficient used for scaling is applied by the r-th relay for the first information block and is equal to $g_{r,1}$=$\sqrt{(P_{\text{R}}/2M)/\,\mathbb{E}[|q_{r,1}|^2]}$ (the same for the second block). Consequently, the available power is split equally as P_{R}/M between two STC blocks.

When the STC is used, the relays start to broadcast the blocks $z_r(1), z_r(2)$ during the ABS as shown in Fig. 2. Now we denote the channel that is formed between the relays and user i as the $1\times\ M$ vector $\boldsymbol{h}$. This remains constant for the entire duration of the ABS. Now we pack $z_r(1)$ for each involved relay into a $M\times\ 1$ vector denoted as $\boldsymbol{z}(1)$. Then, the received signal at the desired destination during the first slot is:

$$y(1) = \boldsymbol{h}^T\boldsymbol{z}(1) + n(1)$$

In the last equation $n(1)$ is the aggregate AWGN and interference[6] at the destination over the two consecutive slot transmissions. Similarly for $y(2)$. Next, we create a model for the equivalent channel by taking the complex conjugate in the second column of vector $\mathbf{y}$. The signal that results we denote it as $\tilde{\mathbf{y}}$. So we have:

$$\tilde{\mathbf{y}}_i = \underbrace{\begin{bmatrix} \sum_r g_{r,1}c_{r,1}h_r & \sum_r g_{r,2}c_{r,2}h_r \\ \sum_r g_{r,1}c_{r,2}^*h_r^* & -\sum_r g_{r,2}c_{r,1}^*h_r^* \end{bmatrix}}_{\mathbf{H}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\mathbf{x}} + \underbrace{\begin{bmatrix} n(1) \\ n^*(2) \end{bmatrix}}_{\mathbf{n}}$$

Decoding. From the received signal model in the last equation we calculate the covariance matrix of the noise vector that is the 2x2 matrix $\mathbf{\Sigma_n} = \text{diag}(I_{\text{ABS}}+\sigma^2, I_{\text{ABS}}+\sigma^2)$. For ensuring final decoding of the transmitted symbols, we us linear MMSE equalization for the received signal:

$$\hat{\mathbf{x}} = (\mathbf{H}^H\mathbf{\Sigma_n}^{-1}\mathbf{H} + \mathbf{I})^{-1}\mathbf{H}^H\mathbf{\Sigma_n}^{-1}\tilde{\mathbf{y}} \tag{4}$$

In the above $\hat{\mathbf{x}}$ is the decoding result of hard decision decoding and the channel is estimated from the preamble of the STC block [8].

SNR and Achievable Rate. For MMSE equalization we obtain the well-known result for the instantaneous SNR at the receiver after decoding block 1 is:

$$\gamma_i^{\text{D}} = \frac{1}{[(\mathbf{H}^H\mathbf{\Sigma_n}^{-1}\mathbf{H} + \mathbf{I})^{-1}]_1} - 1 \tag{5}$$

This is a quantity measured at the receiver when the DSCTIF mode is used. Each user i can calculate the average enjoyed data rate with the DSTCIF mode as

$$C_i^{\text{D}} = \mathbb{E}[\log(1 + \gamma_i^{\text{D}})] \tag{6}$$

[6]The total power of the interference signals during ABSs is $I_{\text{ABS},i}$.

after several measurements of γ_i^{D} user i communicates C_i^{D} to PBS j, similar to related work [2], [4].

The Implementation of DSTCIF. The video-aware resource allocation algorithm that is analyzed in the next section may allocate a non-zero portion of the ABS and RS resources for the DSTCIF mode. Then the PBS is responsible for ensuring that in practice the resources will be allocated as single RS and a single ABS subframe at the same time to the user. This can be realized by indicating during the *ABS period* the type of subframe (ABS or RS), and then set of users that will receive data blocks in specific subframes [4]. This mechanism has been part of the LTE-A specification and more specifically the 12th release [3]. Other actions that the PBS has to execute is to translate the resource allocation decisions (that we calculate in the next section) to discrete number of packet transmissions. Finally, one must note that this randomized version of DSTCIF is fully distributed and implementable in practical systems. The reader is referred to [9], [8] for details.

IV. Direct Transmission Mode

We also have to estimate the rate that is achievable during RS and ABS. We do that consistently with the recent works [2], [4]. The SINR between user i and the PBS and for ABS is:

$$\gamma_i^{\text{ABS}} = \frac{P_{\text{PBS}}|h_{\text{PBS},i}|^2}{I_{\text{ABS},i} + \sigma^2} \tag{7}$$

The instantaneous SINR during RS for user i is:

$$\gamma_i^{\text{RS}} = \frac{P_{\text{PBS}}h_{\text{PBS},i}^2}{I_{\text{RS},i} + \sigma^2} \tag{8}$$

Thus, the result for the direct transmission mode, are the values denoted as $C_i^{\text{RS}}, C_i^{\text{ABS}}$ that are calculated similarly with (6).

V. Problem Formulation and Solution

- For each user i associated to BS j the MNO that operates the HetNet must allocate a rate $r_i \in [0, R_i^{\text{MAX}}]$ that maximizes the video quality.
- The fraction of the ABS resources that the PBS j allocates to user i for transmitting with the direct mode is denoted as $z_i^{\text{ABS}} \in [0, 1]$. Similarly for the RS, i.e., $z_i^{\text{RS}} \in [0, 1]$, and also for the DSTCIF mode we define $z_i^{\text{D}} \in [0, 1]$.
- Hence, this is a *Joint Rate and Resource Allocation (JRRA)* problem.

summary, the decisions of each base station j are: (a) the *video rate allocation* vector for all the associated users i.e. $\boldsymbol{r}_j$=$\big(r_i \geq 0 : i \in \mathcal{N}_j, j \in \mathcal{J}\big)$ (b) the *resource allocation* vector for all users $\boldsymbol{z}_j = \big(z_i^{\text{ABS}}, z_i^{\text{RS}}, z_i^{\text{D}} \geq 0 : i \in \mathcal{N}_j, j \in \mathcal{J}\big)$.

Similarly the *resource allocation* vector for the regular slots, and DSTCIF. To minimize the notation we also define different concatenations of the variable vectors as follows: $\boldsymbol{z}$=$\big(\boldsymbol{z}_j \geq 0 : j \in \mathcal{J}\big)$, and similarly for $\boldsymbol{r}_j, \boldsymbol{r}$.

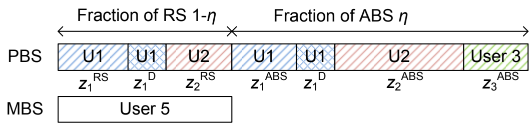


Fig. 3. The fraction of time-domain resources allocated to user i is indicated by $z_i^{\text{ABS}}, z_i^{\text{RS}}, z_i^{\text{D}}$ during the ABS and RS respectively (note that these are not packet transmissions).

For the first constraints we have to recall that the total fraction of the blank ABS resources that are available at the PBSs is η. This leads to:

$$\sum_{i \in \mathcal{N}_j} (z_i^{\text{ABS}} + z_i^{\text{D}}) \leq \eta, \forall j \in \mathcal{J} \quad (9)$$

$$\sum_{i \in \mathcal{N}_j} (z_i^{\text{RS}} + z_i^{\text{D}}) \leq 1 - \eta, \forall j \in \mathcal{J} \quad (10)$$

When a specific SVC layer is used, the average rate in bits/sec that must be achieved by a user i is less than the rate during both the ABS and RS subframes. The resources that are allocated during ABS and RS will also define what the maximum rate that we can stream the video. We can write the previous condition formally as:

$$r_i \leq \left(z_i^{\text{ABS}} C_i^{\text{ABS}} + z_i^{\text{RS}} C_i^{\text{RS}} + 2 z_i^{\text{D}} C_i^{\text{D}}\right), \forall i \in \mathcal{N}_j, j \in \mathcal{J} \quad (11)$$

The above can also support re-buffering constraints [5]. Furthermore, (11) ensures that resources are allocated to the proposed mode when it offers double rate when compared to the direct mode since it requires a double amount of time.

A. Objective

The objective of the HetNet operator is to maximize the total video quality by allocating the available resources. This means that the resources allocated to the small cells should be adjusted depending on the number of associated users and the specific content delivered to each one of them. Since during the ABSs higher spectral efficiency can be achieved because of lower interference, a balancing of the allocated resources across the video streams and across the ABS and RS is of outmost importance. Hence, we have the JRRA problem formulation with proportional-fair utility metric:

$$\max_{\boldsymbol{z}, \boldsymbol{r}} \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{N}_j} \alpha_i \log(r_i) \text{ s.t. } (9) - (11) \quad (12)$$

Recall that α_i is the average utility/bit of the video flow transmitted to user i. Thus, the objective in (12) is expressed in utility-per-bit delivered to the complete HetNet.

B. Solution Approach

The problem is convex since the objective and the constraints are linear. Furthermore it is decoupled across all the BSs since there is no coupling constraint. This means that it can be solved easily in polynomial time at each PBS. After the optimal decision vector $\boldsymbol{z}_j$ has been calculated, this decision is converted by the PBS to a discrete number of packet transmissions. On the other hand, the calculated optimal streaming rate r_i^*, is enforced by selecting SVC video description l that requires a streaming rate, calculated with (1) less than that of r_i^*.

VI. Performance Evaluation

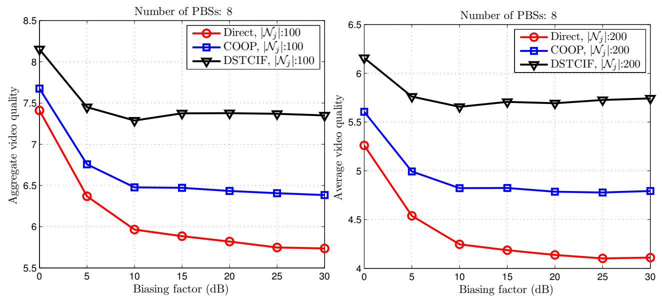
In this section, we present a thorough evaluation of the proposed algorithms that constitute our framework through simulations. The parameter settings for our simulations are set as follows. Downlink MBS and PBS transmit power are equal to 46dBm and 30dBm respectively. We also account for the distance-dependent path loss that is given by $L(d)$=128.1+37.6 $\log_{10}(d)$, where d is the distance in Km [3], and the shadowing standard deviation is 8 dB. The macrocell area is defined as a circle with a radius equal to 1 Km. The parameters of the wireless channel parameters are as follows. A wireless bandwidth of W=20 MHz, a noise power spectral density of σ^2=10^{-6} Watt/Hz, while the same Rayleigh fading model was used for all the channels. Our assumption in this case includes a frequency-flat fading wireless link that remains invariant per transmitted PHY frame, but it might vary between simulated frames. Channel estimation at the receivers is ideal. Thus, our channel model is comprehensive since it considers all the channel impairments. We have to also note that our system performs a precise PHY-level simulation of wireless packet transmissions.

The traffic model for every user is that of an full buffer that is infinite in size (i.e., available video content for all of them). The distribution of the users and picocell locations within the complete macrocell are random and uniform. Our tested systems include: (1) The Direct mode that uses only direct transmissions to the user but also video-aware resource allocation as in [2], [4], [5]. (2) COOP where the R-DSTC protocol in [8] is used for the ABSs so as to obtain extract a diversity gain and increase spectral efficiency (to represent how a typical cooperative system would be used). (3) The proposed DSTCIF. Also all systems use JRRA again for fairness. We average the results over a number of 100 topologies that are generated randomly.

The video content used in the experiments consists of the CIF sequences that were obtained from [14]. The video traces were already compressed in several layers using the SVC H.264 codec at different rates ranging from 128 Kbps and reaching values<7 Mbps. Each video frame was packetized in one slice. The selected sequences were encoded at a frame rate of 30 fps using the frame-type pattern of G16B7, i.e. IBBBBBBBP i.e., there are 7 B frames between every two P frames and a GOP size equal to 16 frames.

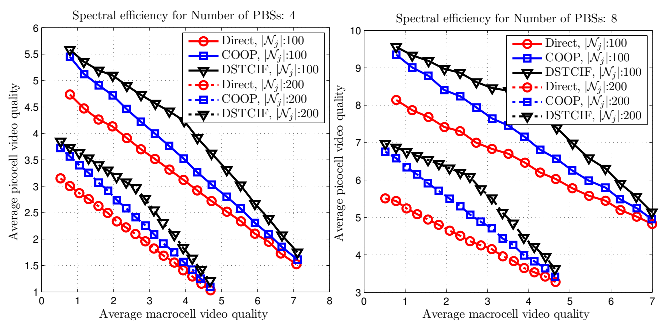
A. The Impact of HetNet Association through Biasing

Association through biasing in the paper is studied through simulations for a constant η=0.5. The related results can be seen in Fig. 4(a),4(b) for different values of the SINR bias. When the bias is increased then users that are at farther distances from the PBS join the picocell. For very high value of the bias effectively all the HetNet users are associated to the picocells. The well-known result for HetNets is that the optimal biasing threshold depends on the relative transmit



(a) (b)

Fig. 4. Aggregate macrocell vs. aggregate picocell video utility/sec that is delivered to the users.



(a) (b)

Fig. 5. Average macrocell vs. Average picocell video quality in terms of the discrete SVC layers delivered. As we observe the right of the x axis more resources are belong to the RS category.

powers of the MBS and PBSs [2]. Here, the optimal biasing is at 0 dB. With an increase in biasing (beyond the optimal) we observe worse performance because the users from larger distances are associated and they suffer lower rate and lower utility. However, benefits in this case are attributed to load balancing [2]. Nevertheless, the important result that we want to clarify with this first set of figures is that biasing can be optimally set for given MBS and PBS power.

B. Video Spectral Efficiency

In this set of experiments we configure the biasing threshold to 0 dB. For this set of results we present the average video quality of the users picocell versus the average quality for the users of the macrocell (only from the MBS to its associated users) for different fixed values of η. This allows us to demonstrate the effect of different TDRP on the use of our proposed DSTCIF mode. For example in these figures as we move in the x axis from left to right, η is reduced and the macrocell users enjoy high video quality contrary to the picocell users.

The results for all the tested systems can be seen in Fig. 5(a),5(b). DSTCIF is superior if it is compared to COOP and Direct for a high user density while the PBS density is low in Fig. 5(a). As expected for higher number of users the layers they receive are fewer but overall the performance improvement is still there. As we increase the density of the users this means that the JRRA algorithm is more important, since the rate of one PBS is shared across several users. On the other hand as we increase the number of the PBSs users in Fig. 5(b), all the tested systems can reach improved performance. The simple reason is that a smaller number of users are associated to the picocells and so a higher rate is possible for each user under any resource allocation scheme. Hence, more picocells leads to improved results due to the higher rate per user that is offered by DSTCIF and COOP. Another important result we see is that even for constant density of PBSs, we achieve better system performance as the user population grows. This is a key benefit of our scheme unlike non-cooperative transmissions.

VII. Conclusions

In this paper we presented a cooperative protocol and an optimization framework for spectral efficient video delivery in HetNets. The central concept is based on decoding the small cell BS data at a higher number of small cell users. Then a cooperative protocol is used to increase the receiver SNR and the achievable communication rate. Significant video quality improvements indicate the benefits of our scheme for increasing user densities. The proposed framework can be used even for non-video data transmission since it offers higher communication rate.

References

[1] Ericsson Mobility Report: On the Pulse of the Networked Society, June 2016.

[2] S. Singh and J. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 2, pp. 888–901, February 2014.

[3] 3GPP, "LTE-Advanced," http://www.3gpp.org/specifications/releases/68-release-12, 2013.

[4] S. Deb, P. Monogioudis, J. Miernik, and J. Seymour, "Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets," *IEEE/ACM Transactions on Networking*, vol. 22, no. 1, pp. 137–150, Feb 2014.

[5] A. Argyriou, D. Kosmanos, and L. Tassiulas, "Joint time-domain resource partitioning, rate allocation, and video quality adaptation in heterogeneous cellular networks," *IEEE Transactions on Multimedia*, vol. 17, no. 5, pp. 736–745, May 2015.

[6] M. Ji, G. Caire, and A. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *Selected Areas in Communications, IEEE Journal on*, vol. PP, no. 99, pp. 1–1, 2015.

[7] A. Argyriou, D. Kosmanos, and L. Tassiulas, "Optimizing interference cancellation in cooperative wireless networks with relay selection," in *CISS*, March 2014, pp. 1–5.

[8] B. Sirkeci-Mergen and A. Scaglione, "Randomized distributed space-time coding for cooperative communication in self-organized networks," in *IEEE SPAWC*, 2005.

[9] A. Argyriou, "User cooperation in heterogeneous cellular networks for improved spectral efficiency," in *Globecom*. IEEE, 2015.

[10] K. Poularakis, G. Iosifidis, A. Argyriou, and L. Tassiulas, "Video delivery over heterogeneous cellular networks: Optimizing cost and performance," in *IEEE Infocom*, 2014.

[11] T. Stockhammer, "Dynamic adaptive streaming over http: Standards and design principles," in *ACM Multimedia Systems*, 2011.

[12] A. Galanopoulos, G. Iosifidis, A. Argyriou, and L. Tassiulas, "Green video delivery in lte-based heterogeneous cellular networks," in *IEEE WoWMoM*, 2015.

[13] X. Zhang, Y. Sun, X. Chen, S. Zhou, J. Wang, and N. B. Shroff, "Distributed power allocation for coordinated multipoint transmissions in distributed antenna systems," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2281–2291, May 2013.

[14] Video Trace Library: http://trace.eas.asu.edu/.