Cache-Assisted Interference Cancellation in the Multi-User Downlink Fading Channel and its Application in HetNets

Antonios Argyriou

Abstract

Next generation wireless systems will be characterized by two things among others: the dense deployment of small cell base stations (SBS), and caching of popular data files at the edge. In this paper we study the downlink multi-user fading channel in a small cell when the interfering transmissions from another non-cooperative BS are files characterized by asynchronous reuse, that is the same file is transmitted several times depending on its popularity. For this system model we propose cache-assisted successive interference cancellation (CAIC): The small cell users cache popular files, and then for decoding the desired file from the SBS they use the cached data for canceling the interfering transmissions of the interfering BS. The result in this case is an interference-free point-to-point link. For this system we develop an analytical model for the ergodic capacity (fast fading) and outage probability (slow fading) of the multi-user channel from the SBS towards its users. Next, we validate our analytical tools with simulation and experiments, and finally we apply CAIC in HetNets where we show the significant performance improvements that it offers.

A. Argyriou is with the Department of Electrical and Computer Engineering, University of Thessaly, Greece, 38333 Greece e-mail: anargyr@uth.gr.

I. Introduction

Cell shrinking has been the most effective technique for achieving orders of magnitude better spectral efficiency in wireless communication systems. Maintaining a small cell is the advantage of wireless LANs that use a single low-power small cell base station (SBS). However, this trend has also given rise to the heterogeneous cellular network (HetNet) paradigm, a central component of future cellular systems, where a high power macro BS (MBS) is overlaid with tiers of SBSs that use short-range spectral-efficient communication. Cell shrinking allows the reuse of wireless resources within the same network, e.g. MBS and SBSs in HetNets [1].[1]

In cellular systems where resources are aggressively reused across cells, interference must be minimized with network-wide inter-cell interference coordination (ICIC) mechanisms. The options for handling the residual interference at small cells are limited: Either treat it as noise, or try to decode it with optimal successive interference cancellation (SIC) methods where the stronger signal is decoded first [2]. We believe that a third option, complementary to SIC, may be particularly suitable for a downlink multi-user channel, illustrated within the ellipsis in Fig. 1: If the transmitted data (file f_l in Fig. 1) from an interfering BS is available at an SBS user, then this user can use this prior information for canceling the interference without having to decode it. By using prior information, the SBS can achieve a capacity gain since interference is eliminated and the channel is converted to a single-input single-output (SISO) channel. The concept is applicable in several network topologies that include WLANs and cellular HetNets since it is a fundamental wireless communication topology, namely the downlink multi-user channel [2]. We argue that this optimization, that leverages prior information, may be even more promising in scenarios where the BS delivers files that feature asynchronous content reuse. That is, the same file is re-delivered over the

[1] Aggressive spectrum reuse can occur across networks, e.g. LTE Unlicensed operates in the unlicensed WiFi band, also neighboring WiFi SBSs reuse spectrum. The end result is that these smaller cells will suffer from interference.

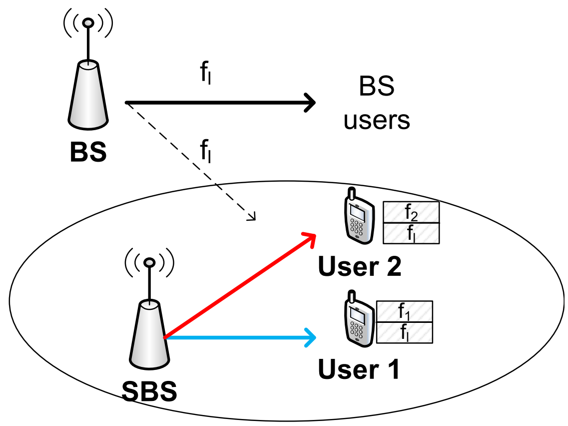


Fig. 1. An isolated small cell consists of a SBS that transmits its data to its associated users. The BS transmission of file f_l interferes (dashed line) at the SBS users. If this file is available at the users the capacity of the SBS multi-user downlink channel can be increased without any knowledge of the interference at the SBS.

wireless medium repeatedly depending on its popularity allowing for more opportunities of cache-assisted interference cancellation (CAIC). This type of traffic is characteristic of video-on-demand (VoD) applications that deliver popular files (e.g., viral Youtube videos), and constitutes the dominant traffic type in cellular networks and wireless LANs. This makes CAIC a perfect fit for modern wireless networks.

A. Related Work

Storage capacity at user devices has grown significantly the last few years. Caching content at the users has already been established as a technique that offers multiple benefits in small cells [3]–[11]. The concept of femtocaching, first proposed by Golrezai et al. [3], exploits content reuse by small-cell users by placing a cache at the SBS in Fig. 1. To improve efficiency, users in the small cell can also pool together their storage resources, forming thus a device-to-device (D2D) caching network and a single common virtual cache (CVC) [12]. The users store disjoint sets of files from a common library based on deterministic decisions that are controlled by the SBS. This allows optimal use of the storage resources in the

D2D network. Fully randomized caching decisions by the users can also be employed in a D2D caching network in order to populate the CVC [12]. This caching policy allows the throughput to scale linearly with the number of users and the cache space (per user) in the D2D network. The scaling behavior is similar with that of the deterministic D2D caching scheme. In the remaining of this paper we refer to these systems as classic caching systems since their purpose is to cache data for delivery to the primary user.

Moving one step beyond the baseline classic caching schemes, new concepts suggest the combination of caching with physical layer methods. Coded multicasting (or else index coding), is one of those ideas that combines caching of coded data at the users, with broadcast wireless transmissions from the SBS. The same coded wireless data transmission from the SBS is useful to more than one users [5], [6], [11], [13]. The authors that proposed this concept in [5] showed that when the collective cache size MU (M files in each of the U SBS users) is larger than the file library of F files, the number of required SBS transmissions scales as a function of these parameters. Shanmugam et al. [13] also considered the use of side information that is provided from the broadcast transmissions from the BS in Fig. 1, that is in full coordination with the SBS. Coded multicasting is again controlled from the SBS in order to increase index coding opportunities. In these works the SBS encodes the data over a finite field (similar to network coding). Beyond coded multicasting, coordinated and cooperative transmission from MIMO-enabled base stations can also be combined with caching in systems [14]. The problem of cooperative transmission from BSs and user caching, was studied in [15] under a randomized user and BS distribution (Poisson process model). In this last work the BSs transmit information that is superimposed at each BS, thus assuming full cooperation between two BSs through the backend. Furthermore, this coordinated BS transmission and user caching, can be optimized and fine-tuned for ensuring highly delay sensitive information like video [11], and even more secure the video delivery process [9], [10].

However, in our target setup the BS is non-cooperative entity: It transmits packets that interfere with all users associated to the SBS. This makes all the previous works not suitable for our setup. The problem is that the BS cannot be forced to use index coding [13], superposition coding [15], or apply some form of fully synchronized MIMO transmission [14]. Hence, in our scenario BS transmission is a source of uncontrolled intra-cell interference. In this case when the information is unknown the optimal strategy for the user is to use SIC [2]. There are however scenarios where the broadcasted information from the BS is already available at some users. Several works suggest the concept we presented in the introduction: Use a-priori information for decoding data that have been additively combined from the channel at the signal level [16]. In [16] the authors considered that in a three-hop chain network the node in the middle has available a-priori data packets because of their MAC-layer retransmission. A wireline backhaul can also provide a-priori data for decoding in network MIMO scenarios. Known-interference cancellation refers to the same concept of exploiting the availability of a-priori information but in a very practical setting where the two interfering transmissions may be asynchronous. The same concept can also be used in multi-hop wireless networks and the availability of a-priori data is because of multihop forwarding. In this work we focus in the most frequently used topology today that of an independent network that consists of a SBS and several users associated to it (the downlink multi-user channel).

B. This paper

As we discussed in the last subsection, content reuse is big part of the downlink traffic in modern wireless networks. To obtain a clear picture of the potential performance gains that may come from content reuse and a-priori data in real network deployments, key aspects of modern wireless systems must be taken into account. In this paper we first consider a realistic fading channel.[2] Second, we model practical wireless algorithms that are omnipresent and involve packet scheduling and adaptive modulation and coding (MCS) that aim to minimize the impact of fading. These algorithms perplex even more the packet-based transmission and the decoding algorithm that uses prior information. Third, we consider the very popular downlink multi-user channel, for which the sum-capacity with CSIT in fast fading is achieved by scheduling the user with the highest signal-to-interference plus noise ratio (SINR), i.e. one of the dominant wireless scheduling schemes (§6 [2]). Fourth, we do not consider any form of cooperation between the SBS and the BS. Hence, modeling the performance of the widely used and optimal strategy for the multi-user downlink fading channel in the presence of a non-cooperative interfering source that exhibits content reuse (as it is the case in real systems) is an open unstudied problem that has practical implications for modern wireless networks.

The specific contributions of this paper are the following: 1) We propose CAIC which is suitable for multi-user downlink channels when interfering signals are already cached locally. The numerical, simulation, and experimental evaluation of CAIC in a real wireless testbed, proves that the basic CAIC scheme offers capacity and outage benefits that grow linearly w.r.t. the space allocated for CAIC at the users. 2) A closed-form ergodic capacity model under CSIT and optimal multi-user scheduling for fast fading channels, and also an outage model for CAIC suitable for slow fading channels when only CSIR is available. 3) A practical algorithm that may serve as the basis of interfacing CAIC with classic caching schemes. 4) We propose the use of CAIC simultaneously with classic caching in 5G HetNets, and evaluate its performance under a variety of scenarios.

[2]Without channel state information at the transmitter (CSIT) the strict definition of capacity in a fading channel is not informative since it is zero.

C. Paper Organization

Our detailed system model is described in Section II. In Section III we develop an analytical model that characterizes the capacity of the multi-user fast fading channel with CAIC, and in Section IV the outage probability model under slow fading for a single user and the proposed scheme. The validation of the models is presented in Section V. Subsequently, we propose further system optimizations for interfacing our scheme with classic caching algorithms in Section VI. A thorough performance evaluation for the application of our scheme in HetNets is presented in Section VII, while we conclude this paper in Section VIII.

II. System Model

In this paper we consider the multi-user performance of a SBS that operates under the presence of an interfering BS when both deliver popular content. We consider that the BS is a partially-cooperative entity, that is it does not coordinate its transmissions with the SBS, but it only allows the SBS to know from overhearing what is the MCS that it transmits, and what is the id of the data. Hence, the BS is an autonomous entity that optimizes its transmissions at its own discretion (not the focus of this paper). Consequently, our network model consists of a single interfering BS, a SBS and its associated users contained in the set set $\mathcal{U}$=$\{0, 1, ..., U-1\}$ with $|\mathcal{U}|$=U (Fig. 1). As we will later demonstrate, this model can be used for a HetNet that consits of multiple SBSs. User mobility is not considered in this work. However, since our model does not depend on any type of global optimization, mobile users associating to a SBS can be easily accommodated. Informing the SBS regarding the contents of their cache will allow the SBS to include them immediately in the scheduling decisions regardless of the other users.

A. Signal and Interference Model

In our model each file is segmented into data chunks which is the smallest unit of the file. Clearly, a physical-layer (PHY) packet may contain more than one chunks but in this

TABLE I

System parameters and paper notation.

Symbol	Explanation	Symbol	Explanation
U	Users per small cell	$h_{x,y}$	Channel gain between node x and y
M	Cache size at a user	R_1	MCS used by the SBS
F	Size of file library	R_2	MCS used by the BS
P_{BS}	BS TX power	p_l	Popularity of the l-th file of the library
P_{SBS}	SBS TX power	γ	Zipf parameter
σ^2	Noise variance	f_l	the l-th file
C	Capacity	$P(i, f_l)$	Prob. that file f_l is located at user i
$C_{\mathrm{x,i}}^{\mathrm{scheme}}$	Capacity of "scheme" between node x and user i	$P_{\mathrm{CVC}}(f_l)$	Prob. that file f_l is located at the CVC

paper for exposition purposes we assume that it contains one chunk. In a time slot of the wireless system the transmitted PHY baseband data from the SBS and the BS are denoted as x_{SBS} and x_{BS} respectively. For the slow fading derivations the channels are assumed to be block-fading Rayleigh and quasi-stationary, that is they remain constant for the complete duration of a transmitted data block. The channel gain from the s-th to the i-th node is denoted as $h_{s,i}$. The fading coefficients are independent and $h_{s,i} \sim \mathcal{CN}(PL, 1)$, i.e., they are complex Gaussian random variables with unit variance and mean that depends on the path loss component (explained later). In both cellular and wireless LANs, scheduling and MCS adaptation are made possible because of Channel State Information at the Transmitter (CSIT), something that is also possible in modern systems [2]. Now assume that the BS encodes the data at a rate R_2 bits/s/Hz and the SBS at R_1 bits/s/Hz. With the use of OFDM, frequency selectivity is removed, and the baseband flat fading signal model at user i is:

$$y_i = \sqrt{P_{\mathrm{SBS}}} h_{\mathrm{SBS},i} x_{\mathrm{SBS}} + \sqrt{P_{\mathrm{BS}}} h_{\mathrm{BS},i} x_{\mathrm{BS}} + w_i \tag{1}$$

P_{SBS}, P_{BS} correspond to the transmit power of the SBS and the BS respectively and w_i is the noise sample. The average channel gain is affected by distance-dependent path loss according to $L(d)$=$128.1 + 37.6 \log_{10}(d)$ dB, where d is the distance in Km [17], and the shadowing standard deviation is 8 dB.[3] Additive white Gaussian noise (AWGN) is assumed at the receivers of the users with zero mean and variance equal to σ^2. Channel State Information at the Receiver (CSIR) and CSIT are available depending on our analysis. CSI from the BS to the user is not required at the SBS.

B. The Cache-Assisted Interference Canceling (CAIC) Receiver.

When a user receives the baseband signal in (1) that contains interference from the BS, it attempts to decode the desired data x_{SBS}. CAIC operates at the level of individual PHY packets: If the interfering data chunk x_{BS} is not locally available at the user, then it cannot remove it from (1) and improve decoding performance. Consequently, the receiver uses classic SIC and decodes the stronger of the two interfering signals [2]. Assuming for normalization that $\mathbb{E}[|x_{\mathrm{SBS}}|^2]$=$\mathbb{E}[|x_{\mathrm{BS}}|^2]$=1, to decode x_{SBS} first, the tested condition is:[4]

$$\frac{P_{\mathrm{SBS}}|h_{\mathrm{SBS},i}|^2}{2^{R_1}-1} > \frac{P_{\mathrm{BS}}|h_{\mathrm{BS},i}|^2}{2^{R_2}-1} \tag{2}$$

In the opposite case that the interfering BS signal x_{BS} has the highest energy/bit, it is decoded first. We do not assume perfect interference cancellation and the signal might not be decoded. But in case the signal from the BS is correctly decoded, with CSIR at the user (the estimated channel $h_{\mathrm{BS},i}$) the receiver can remove x_{BS} from the aggregate received signal y_i. Note that the decoding algorithm at the user must also know R_2 in order to recreate

[3]Inter-cell interference from other macrocells may also be taken into account in an aggregate form.

[4]It is possible that different rules are used for selecting the symbol to be decoded first or even a completely different SIC scheme. In this case we add the division so that we compare not just the signal power but the energy/bit: If a signal for example has lower power but uses low order modulation e.g. BPSK then it may have higher chance of being decoded when compared to a signal that has higher power but lower energy/bit due to a high order modulation..

locally the modulated signal x_{BS} (by modulating and coding the cached data). This allows us to decode at the next stage x_{SBS} by using the following baseband signal that does not contain interference:

$$y_i = \sqrt{P_{\mathrm{SBS}}} h_{\mathrm{SBS},i} x_{\mathrm{SBS}} + w_i \tag{3}$$

However, in the case of incorrect decoding of x_{BS} the decoding of x_{SBS} has to treat x_{BS} as noise leading to worse performance, i.e., (1) is applicable. CAIC is enabled when the data chunk x_{BS} is found in the local cache of user i. Since this data chunk is locally available in a correct decoded form, the previous algorithm is simplified: In particular (3) is valid since x_{BS} can be subtracted easily from the composite signal in (1) because it is known (it does not have to be decoded).

C. File Library & Content Re-use.

The file library that can be delivered in this network is the set $\mathcal{F}=\{f_0, f_1, ..., f_{F-1}\}$ with $|\mathcal{F}|=F$. It is important to clarify the use of the caching space. The SBS and the users each have a storage space of M files each. For the purpose of our model and our later analysis this space of M files is separated into two parts virtually. This means that separation is not physical, that is for a file that should exist in both caches we maintain a single copy in the hard disk. The first part is the CAIC cache that has a size M_{CAIC} files. This space is dedicated only for CAIC and this is the only part of the cache that concerns the performance analysis in Sections III and IV. We also allow the existence of a CVC that stores cached data that are required for the consumption of the small cell users and is part of our discussion in Section VI. Hence, the available space of the CVC is UM_{CVC} files and is only dedicated to delivering the primary information requested by a user. The specific subset of files that is cached at node i is denoted as $\mathcal{H}_i$, while the union of the caches of all the U small cell users is $\mathcal{H}=\bigcup_{i=1}^{U}\mathcal{H}_i$. We assume that the caches have been populated during the earlier system operation (typically the caches are re-populated with a period of a few days [3], [5], [8]). The demand for the file library within a certain time period (e.g., a few hours or days) is assumed to be fixed and known in advance [3]. For example, the demand of future requests can be estimated by analyzing past user requests [18]. The popularity of the F files is modeled with a Zipf distribution: The popularity of file f_k is indicated by p_k and so $\sum_{k=0}^{F-1} p_k = 1$. The file popularity is decreasing in the index, i.e., $p_k \geq p_l$ if $k \leq l$. The popularity of the k-th file, denoted by p_k, is inversely proportional to its rank:

$$p_k = 1 \Big/ (k+1)^{\gamma} \sum_{l=1}^{F} \frac{1}{l^{\gamma}}, \ 0 \leq k \leq F-1.$$

III. Ergodic Capacity with CSIT in Fast Fading

A. Capacity-Achieving Multi-User Scheduling with CAIC at the SBS.

In practical wireless systems packet schedulers at the BS and SBS are typically channel-aware and dynamic, that is they consider the instantaneous fading levels. For maximizing the sum-capacity in the downlink direction, when the available transmit power is P, the user that has the best channel must be scheduled [2]. Assuming Gaussian inputs and a capacity achieving channel code, we can apply Shannon's formula for user i and we have an additional case with CAIC: Without CAIC it is C^{NoCAIC}=$\log(1+\frac{P|h_{\text{SBS,i}}|^2}{P|h_{\text{BS,i}}|^2+\sigma^2})$ bits/s/Hz, while if CAIC can be applied the capacity is C^{CAIC}=$\log(1+\frac{P|h_{\text{SBS,i}}|^2}{\sigma^2})$ bits/s/Hz. Hence, in our system the user that achieves the highest instantaneous SINR is still scheduled by the SBS but with the additional consideration of CAIC on this SINR. This is a cache-assisted sum-capacity-maximizing scheduling policy that enhances the well-known optimal opportunistic scheduling scheme in the downlink multi-user channel [2]. The next natural question is what is the performance of such a joint scheduling/CAIC scheme.

B. Capacity

For a given set of transmission parameters at the BS, that include the data file f_l and R_2, the capacity of the SBS downlink multi-user fast-fading fading channel with CSIT and under optimal scheduling is

$$C^{\text{CAIC}}(f_l, R_2) = \mathbb{E}\left[\log_2(1 + \max_{i \in \mathcal{U}}\{S_i\})|f_l, R_2\right], \tag{4}$$

where the expectation is with respect to the random variable S_i that is the SINR of each user. Reformulating the capacity expression to account for a single random variable $S{=}\max_{i\in\mathcal{U}}\{S_i\}$, which is the maximum instantaneous SINR of any user, we have:

$$C^{\text{CAIC}}(f_l, R_2) = \mathbb{E}[\log_2(1 + S)|f_l, R_2] = \int_0^\infty \log_2(1 + s) f_S(s) ds \tag{5}$$

To derive an expression for C^{CAIC} we focus on calculating the cumulative distribution function (CDF) of S. The probability density function (PDF) $f_S(s)$ is calculated the by differentiation.

C. CDF of the SINR random variable with a SIC Receiver

For minimizing the complexity of our analysis, we define the random variables $Y{=}P_{\text{SBS}}|h_{\text{SBS},i}|^2$, $X{=}P_{\text{BS}}|h_{\text{BS},i}|^2$, and their expectation: $\mathbb{E}[Y]{=}\mu_i{=}P_{\text{SBS}}\,\mathbb{E}[|h_{\text{SBS}}|^2]$, and $\mathbb{E}[X]{=}\lambda_i{=}P_{\text{BS}}\,\mathbb{E}[|h_{\text{BS}}|^2]$, where μ_i is the average channel gain from the SBS to user i, and λ_i is the average channel gain from the BS to user i. To calculate the CDF of the SINR that the SBS experiences towards the optimal user, we must note that it depends on the maximum SINR of any user. Furthermore, since we have two decoding options, that is classic SIC and CAIC, the instantaneous SINR is different and denoted as S_i^{SIC}, and S_i^{CAIC} respectively. If we denote with $P(i, f_l)$ the probability that file f_l is located at user i, the CDF is:

$$F_S(s) = \Pr\{S \leq s\} = \Pr\{\max_{i\in\mathcal{U}}\{S_i\} < s|f_l, R_2\} = \Pr\left\{\max_{i\in\mathcal{U}}\left\{S_i^{\text{SIC}} 1_{f_l \notin \mathcal{H}_i}, S_i^{\text{CAIC}} 1_{f_l \in \mathcal{H}_i}\right\} < s\right\} \tag{6}$$

$$= \prod_{i\in\mathcal{U}}\left(\Pr\{S_i^{\text{SIC}} < s, f_l \notin \mathcal{H}_i\} + \Pr\{S_i^{\text{CAIC}} < s, f_l \in \mathcal{H}_i\}\right) \tag{7}$$

$$= \prod_{i\in\mathcal{U}}\left(\Pr\{S_i^{\text{SIC}} < s\}(1 - P(i, f_l)) + \Pr\{S_i^{\text{CAIC}} < s\}P(i, f_l)\right) \tag{8}$$

Formula (7) is because the availability or not of f_l defines two mutually exclusive events. Finally, (8) is because the probability of file f_l residing (w.p. $P(i, f_l)$) or not (w.p. $1-P(i, f_l)$) at a user is independent of the SINR random variable.

Now the optimal caching policy for f_l with CAIC is straightforward to derive: Each user caches the most popular files (least recently used (LRU)) until the available space of M_{CAIC} files for CAIC is consumed. This allows us to calculate $P(i, f_l)$ for a given available space M_{CAIC} and file popularity.

1) Analysis of Classic SIC: To calculate (8) and later (5) numerically, the two CDFs in (8) must be calculated. Let us denote with $C_{\text{BS},i}$ the instantaneous capacity of the channel that is formed between the BS and user i. Even without CAIC, the SINR of the SBS transmission depends on whether the BS data can be decoded or not for the given R_2 that the BS selected. Hence, the CDF is conditioned on the two outcomes of a single event and is written as:

$$F_i^{\text{SIC}}(s) = \Pr\{S_i^{\text{SIC}} < s\} = \Pr\{S_i < s, C_{\text{BS},i} < R_2\} + \Pr\{S_i < s, C_{\text{BS},i} \geq R_2\} \tag{9}$$

We calculate the CDF as the probability of the two mutually exclusive events above.

The first event: This is the case that the BS transmission is not decoded at the user. It is defined as $\Pr\{S_i < s, C_{\text{BS},i} < R_2\}$. The important detail is that if we apply SIC, the SINR S_i will be different depending which signal has the highest energy/bit since this will be the one that will be decoded first. In particular if the signal from the SBS has higher energy/bit than the signal from the BS, i.e., if $\{Y > \frac{2^{R_1}-1}{2^{R_2}-1}X\}$ (this is condition (2)), it will be $S_i = \frac{Y}{X+\sigma^2}$. In the opposite case that the signal from the BS has higher energy/bit than the signal from the SBS, then SIC will attempt to decode first the BS, and if it is successful, it can remove it from the aggregate signal. Thus, it will be $S_i = \frac{Y}{\sigma^2}$.

This discussion leads to the decomposition of the event $\{S_i < s, C_{\text{BS},i} < R_2\}$ to two mutually exclusive events as

$$\Pr\{S_i < s, C_{\text{BS},i} < R_2\} = \Pr\{S_i < s, C_{\text{BS},i} < R_2, Y > \frac{2^{R_1}-1}{2^{R_2}-1}X\}$$
$$+ \Pr\{S_i < s, C_{\text{BS},i} < R_2, Y < \frac{2^{R_1}-1}{2^{R_2}-1}X\}, \tag{10}$$

depending on which block is decoded first. Elaborating further on (10) we obtain:

$$\Pr\{S_i < s, C_{\text{BS},i} < R_2\} = \Pr\{Y - sX < s\sigma^2, Y > \frac{2^{R_1}-1}{2^{R_2}-1}X\}$$
$$+ \Pr\{X - (2^{R_2}-1)Y < (2^{R_2}-1)\sigma^2, Y < \frac{2^{R_1}-1)}{2^{R_2}-1}X\} \tag{11}$$

The event $\{Y > \frac{s}{2^{R_2}-1}X\}$ considers all the cases where the block that originates from the SBS, has the highest energy/bit. If this event is true, then $\{Y - sX < s\sigma^2\}$ is true for the SBS SINR. An important detail is that this event includes the case that the BS cannot be decoded if $\{Y > \frac{s}{2^{R_2}-1}X\}$ is true. The reason is simply that the energy/bit is lower for the BS and so if the SBS cannot be decoded, we can definitely not decode the BS. On the other hand, when $\{Y < \frac{s}{2^{R_2}-1}X\}$ is true in the second of the two independent events in (11), then similarly with before we only need to consider the probability that the BS will not be decoded.

X and Y are independent exponential random variables their joint PDF is separable. Thus, for the first event:

$$\Pr\{Y - sX < s\sigma^2, Y > \frac{2^{R_1}-1)}{2^{R_2}-1}X\} = \frac{\lambda_i}{\lambda_i + \mu_i \frac{P_{\text{BS}}}{P_{\text{SBS}}} \frac{2^{R_1}-1)}{2^{R_2}-1}} - \frac{\lambda_i \exp(-\frac{\mu_i s\sigma^2}{P_{\text{SBS}}})}{\lambda_i + \mu_i \frac{P_{\text{BS}}}{P_{\text{SBS}}} s} \tag{12}$$

Similarly we calculate the probability of the second event in (11) and then by adding the two results we obtain:

$$\Pr\{S_i < s, C_{\text{BS},i} < R_2\} = \frac{\lambda_i}{\lambda_i + \mu_i \frac{P_{\text{BS}}}{P_{\text{SBS}}} \frac{2^{R_1}-1)}{2^{R_2}-1}} - \frac{\lambda_i \exp(-\frac{\mu_i s\sigma^2}{P_{\text{SBS}}})}{\lambda_i + \mu_i \frac{P_{\text{BS}}}{P_{\text{SBS}}} s} + \frac{\mu_i}{\lambda_i \frac{P_{\text{SBS}}}{P_{\text{BS}}} \frac{2^{R_2}-1}{2^{R_1}-1)} + \mu_i} - \frac{\mu_i \exp(-\frac{\lambda_i (2^{R_2}-1)\sigma^2}{P_{\text{BS}}})}{\mu_i + \lambda_i \frac{P_{\text{SBS}}}{P_{\text{BS}}}(2^{R_2}-1)} \tag{13}$$

Second Event: The second event in (9) considers the case that the BS is decoded:

$$\Pr\{S_i < s, C_{\text{BS},i} \geq R_2\}$$
$$= \Pr\{S_i < s, C_{\text{BS},i} \geq R_2, Y > \frac{2^{R_1}-1}{2^{R_2}-1}X\} + \Pr\{S_i < s, C_{\text{BS},i} \geq R_2, Y < \frac{2^{R_1}-1}{2^{R_2}-1}X\} \quad (14)$$

In (14) we followed the same approach with before, only now the probability of the first of the two disjoint events is zero. This is a result of SIC that selects to decode first the data block with the highest energy/bit: If the signal from the SBS is stronger (i.e., $\{Y > \frac{2^{R_1}-1}{2^{R_2}-1}X\}$), and the user fails to decode it (i.e., $\{C_{\text{BS},i} < R_2\}$), then it is impossible to decode the BS since the SBS was not canceled. This leads to

$$\Pr\{S_i < s, C_2 \geq R_2, Y < \frac{2^{R_1}-1}{2^{R_2}-1}X\} \quad (15)$$
$$= \Pr\{Y < s\sigma^2, X - (2^{R_2}-1)Y > (2^{R_2}-1)\sigma^2, Y < \frac{2^{R_1}-1}{2^{R_2}-1}X\}$$
$$= \frac{\mu_i \exp(\frac{-\lambda_i(2^{R_2}-1)\sigma^2}{P_{\text{BS}}})}{\mu_i + \lambda_i \frac{P_{\text{SBS}}}{P_{\text{BS}}}(2^{R_2}-1)}\Big(1 - \exp(-(\frac{\mu_i}{P_{\text{SBS}}} + \frac{\lambda_i}{P_{\text{BS}}}(2^{R_2}-1))s\sigma^2)\Big),$$

where the third line is obtained after integration of the pdf's of the SINR X and Y, in the limits specified by R_1, R_2. In this result note that when the decoding of the BS succeeds, then this data chunk will be successfully removed from the composite signal. This means that the CDF of the SBS SINR is $\{Y < s\sigma^2\}$ because the decoder only needs to combat AWGN. To understand the result consider the case $\lambda_i \to \infty$. This corresponds to a very low value for the average channel gain, i.e., $\mathbb{E}[|h_{\text{BS},i}|^2] \to 0$ and so the power of the received signal from the BS is very low. But one notes that we now consider the event $\{Y < \frac{2^{R_1}-1}{2^{R_2}-1}X\}$, i.e., the event that the signal from the SBS will be received at an even lower power level than the BS. Thus, the probability of the event in (15) becomes zero simply because this event occurs infrequently as $\lambda_i \to \infty$.

Final SIC-aware CDF: By combining the results in (13), (15), we have that (9) becomes

$$F_i^{\text{SIC}}(s) = \Pr\{S_i^{\text{SIC}} \leq s\} = \Big(\frac{\lambda_i}{\lambda_i + \mu_i \frac{P_{\text{BS}}}{P_{\text{SBS}}}\frac{2^{R_1}-1}{2^{R_2}-1}} - \frac{\lambda_i \exp(-\frac{\mu_i s\sigma^2}{P_{\text{SBS}}})}{\lambda_i + \mu_i \frac{P_{\text{BS}}}{P_{\text{SBS}}}s} + \frac{\mu_i}{\lambda_i \frac{P_{\text{SBS}}}{P_{\text{BS}}}\frac{(2^{R_2}-1)}{2^{R_1}-1} + \mu_i} - \frac{\mu_i \exp(-\frac{\lambda_i(2^{R_2}-1)\sigma^2}{P_{\text{BS}}})}{\mu_i + \lambda_i \frac{P_{\text{SBS}}}{P_{\text{BS}}}(2^{R_2}-1)}\Big)$$
$$- \frac{\mu_i \exp(\frac{-\lambda_i(2^{R_2}-1)\sigma^2}{P_{\text{BS}}})}{\mu_i + \lambda_i \frac{P_{\text{SBS}}}{P_{\text{BS}}}(2^{R_2}-1)}\Big(1 - \exp(-(\frac{\mu_i}{P_{\text{SBS}}} + \frac{\lambda_i}{P_{\text{BS}}}(2^{R_2}-1))s\sigma^2)\Big) \quad (16)$$

2) Analysis with CAIC: In the closed-form outage results in (6) there is also the expression for CAIC when the user has knowledge of the BS data. By following the methodology in our previous derivations it is straightforward to derive this as

$$F_i^{\mathrm{CAIC}}(s) = \Pr\{S_i^{\mathrm{CAIC}} \leq s\} = 1 - \exp(-\mu_i s \sigma^2), \tag{17}$$

since it corresponds to a point-to-point SISO channel.

D. BS Model

Our final objective is to include the impact of the particular decision made by the BS. Recall that in our model the BS acts in such a way that it serves its associated users in the best possible way without any consideration for the impact of the selected parameters of the transmission on the small cell users. This means it will adapt R_2 as it considers best, while the selection of the file f_l is also independent from what takes place in the small cells.

The SBS capacity with CAIC depends on how frequently the particular file f_l is transmitted by the BS. In our model that incorporates reusable content, p_l indicates the probability that the user that is associated to the BS requests f_l. Hence the capacity under all the potential file requests is:

$$C(R_2) = \sum_{f_l \in \mathcal{F}} C(f_l, R_2) p_l \tag{18}$$

With an adaptive MCS system at the BS, the selection of R_2 is conditioned on the instantaneous SINR γ of the BS-associated user. If we index MCSs with k then for an SINR $\gamma_{k-1} < \gamma < \gamma_k$ the BS will use the k-th MCS [19]. However, this will affect the performance of SIC for the user scheduled by the SBS according to (4),(16). Based on this discussion we can calculate the capacity as:

$$C = \sum_k \int_{\gamma_{k-1}}^{\gamma_k} C(R_2 = k) f_\gamma(\gamma) d\gamma = \sum_k C(R_2 = k) \int_{\gamma_{k-1}}^{\gamma_k} f_\gamma(\gamma) d\gamma$$

This is calculated fairly easily since we are not interested in the PDF $f_\gamma(\gamma)$ of the SINR between the BS and its users, but only how frequently each allowed MCS is used [19].

For calculating numerically the capacity one has to use the CDF in (16) and (17) to obtain the complete CDF in (8). Basides our own caching policy for the space of M_{CAIC} files, any new caching policy for CAIC can be defined leading to the calculation of $P(i, f_l)$. Then, the capacity for a given R_2 and file f_l can be caluclated with (5). Finally, any desired file popularity rule or MCS rule can be inserted into expressions (18) and (19) to calculate the final ergodic capacity of the small cell.

IV. Outage Analysis for Slow Fading with CSIR

Similar analysis can be performed when CSIR is only available under slow fading. That is, we can only calculate the probability that the SBS can communicate with a user at a given rate R_1 since the capacity is zero under slow fading. Multi-user selection is not possible in this case ((6) is not applicable).

The outage expression is obtained by starting from its definition, and by re-using the CDF calculated previously:

$$\begin{aligned}
&\text{Pout}^{\text{CAIC}}(i, f_l, R_1, R_2) = \Pr\{C_{\text{SBS},i}^{\text{SIC}} < R_1\}(1 - P(i, f_l)) + \Pr\{C_{\text{SBS},i}^{\text{CAIC}} < R_1\}P(i, f_l) \\
&= \Pr\{\log_2(1 + S_i^{\text{SIC}}) < R_1\}(1 - P(i, f_l)) + \Pr\{\log_2(1 + S_i^{\text{CAIC}}) < R_1\}P(i, f_l) \\
&= \Pr\{S_i^{\text{SIC}} < 2^{R_1} - 1\}(1 - P(i, f_l)) + \Pr\{S_{1i}^{\text{CAIC}} < 2^{R_1} - 1\}P(i, f_l) \\
&= F_i^{\text{SIC}}(2^{R_1} - 1)(1 - P(i, f_l)) + F_i^{\text{CAIC}}(2^{R_1} - 1)P(i, f_l) \qquad (19)
\end{aligned}$$

From the definition of an outage event without CAIC and with CAIC we obtain the first expression, and the remaining derivations are obtained by using the Shannon capacity formula, and replacing it in the two SINR CDFs in (16) and (17).

V. Model Validation

Our next objective is to validate the analytical models. For the validation step we present Monte Carlo simulations and experimental results from an actual testbed.

A. Simulation Setup

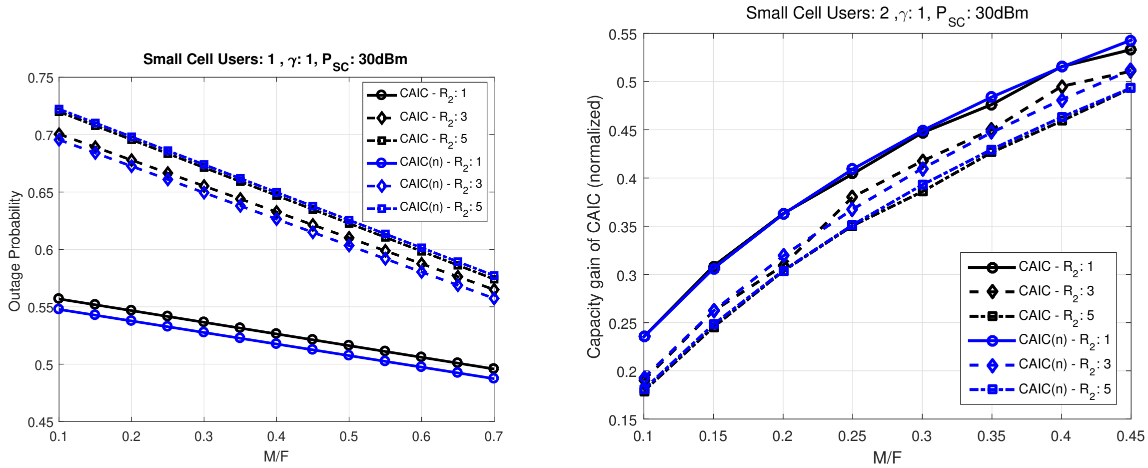
The simulation results are meant to validate the analytical outage formula in (19) for a single user. We set γ=1 in the Zipf distribution, a constant value of R_1=1, and explore a different fraction of the file library M/F that each user stores and also R_2. We also set λ=μ=1, i.e., equal distance from the BS and SBS to the user. We considered a library of F=1000 synthetically generated files where each one had a size of 1Mbyte. Each file was split to 8000 chunks of 1000 bits each. This file size and number of chunks ensure sufficient averaging and larger file sizes did not affect our results. For this configuration we needed a 23-bit identifier for each chunk that was appended before the payload. Four bits were also added to simulate the transmission of the information R_2. Other generic protocol/packet headers were not considered.

B. Outage Model Validation with Simulations

The outage probability of the CAIC receiver for a single user, i.e., expression (19) is plotted in Fig. 2(a) together with the PHY packet-level simulation results. The numerical analysis that provides results is based on a closed-form equation, and can accurately match the packet-level simulation for the Rayleigh channel. This is true for different parameters of the file caching syb-system (M/F), and that of the wireless system (MCS R_2). We observed that numerical and simulation results agree very well because our analysis also considers the impact of SIC error propagation.

C. Experimental Setup & Hardware Experiments

Next we present numerical results for the ergodic capacity model developed in Section III, and compare it with real experimental results. Several research works have implemented practical systems for SIC based on software-defined radio (SDR) platforms [20]. For this paper we implemented the CAIC decoder and the SBS scheduler. We implemented these on a GNURadio testbed that had available four nodes acting as the BS, SBS, and two users. Each



(a) Simulation (CAIC) and numerical (CAIC(n)) outage probability results of a single user v.s. the fraction of the file library that each user stores (M/F) for different R_2.

(b) Experimental (CAIC) and numerical (CAIC(n)) ergodic capacity results with two users v.s. M/F for different R_2.

Fig. 2. Simulation and experimental results

node is an off-the-shelf PC connected to a Universal Software Radio Peripheral 2 (USRP2) with the RFX2400 daughterboard. The RFX2400 operates at the 2.4GHz frequency range. All PCs are installed Debian 8 and GNURadio. We used a standard GNURadio configuration for the wireless link, that is on the transmitter side, the DAC sampling rate is set to $400{\times}10^6$ samples/sec, while the interpolation rate is set to 200%(4 interpolation rate in the DAC chip itself and 50 interpolation rate controlled by GNURadio), and the number of samples per symbol is 2. On the receiver side, the ADC rate is set to $100{\times}10^6$ samples/sec while the decimation rate is 50. To simplify the SIC algorithm, we used only BPSK modulation, and so the resulting bit rate is 1Mbps. The signal to be subtracted is recreated by using the information of the BPSK modulation in our implementation. The channel is estimated with an MMSE estimator from a fixed preamble in each ACK packet.

D. Ergodic Capacity Model Validation & Throughput Gains

To validate the model with the real experiments, we tested a small number of F=10 synthetically generated files with γ=1, and a different fraction of the file library M/F that each user stores. We configured λ=μ=1, i.e., equal average distance from the BS and SBS to the user. The capacity under the CAIC for a single user is plotted together with the measurements for the PHY throughput gain from USRP2. The numerical analysis that provides results is based on a closed-form equation, and can accurately match the experiment. This is true for different parameters of the file caching subsystem (M/F), and MCSs of the BS (R_2). The results also indicate a linear scaling behavior for the user performance as M/F is increased.

VI. Classic Caching and CAIC

The previous models and analyses focused on the ergodic capacity and outage probability of CAIC for different CSI assumptions. The results, obtained for a fundamental communication channel, have general applicability in several different scenarios. In addition, we stress again that CAIC can operate independently from the classic caching algorithms since it is concerned with data that an interfering neighbor transmits. Nevertheless, the space for caching at a user may need to be used by both CAIC and a classic caching algorithm. The later manages the CVC in the small cell (an example in Fig. 3). In this section we model the performance of this type of system and we also propose a first potential algorithm that manages the cache for joint CAIC and classic caching.

A. Outage Analysis with for Joint Classic Caching and CAIC

To model the performance of such a joint system, and eventually evaluate its performance in a HetNet, we investigate a joint metric that considers the possibility that the file f_i that the SBS user wants, may not be located in the CVC (a cache miss or outage event), but also that the wireless transmission fails in case it is found. In this work we extend the definition of an outage event in fading channels with cache-enabled transmitters as follows.[5]

Definition 1: Outage is the event of a failed communication of a data unit either due to wireless transmission failure, or lack of the data unit in the local caching system of the transmitter.

Definition 2: Let $P_{CVC}(f_i)$ denote the probability that file f_i resides in the cache of the transmitter. The outage probability for a cache-enabled communication system is equal to:

$$\text{Pout}(i) = (1 - P_{\text{CVC}}(f_i)) + P_{\text{CVC}}(f_i)\text{Pout}^{\text{wireless}} \tag{20}$$

We can now elaborate on the outage expression in (20) by adding the impact of classic caching and CAIC. In our initial expression we include the indicator function so that a user is in outage if the required file is not in the CVC, or if the file is located in the CVC but at the same time the instantaneous capacity of the channel from the SBS to user i (denoted as $C_{\text{SBS},i}$) is lower than the used MCS R_1. The outage probability is given by:

$$\begin{aligned}\text{Pout}(i) = \Pr\{C_{\text{SBS},i}1_{f_i \in \mathcal{H}} < R_1 | R_2, f_l\} = (1 - P_{\text{CVC}}(f_i)) + P_{\text{CVC}}(f_i)\big(\Pr\{C^{\text{SIC}}_{\text{SBS},i} < R_1\}(1 - P(i, f_l))\\ + \Pr\{C^{\text{CAIC}}_{\text{SBS},i} < R_1\}P(i, f_l)\big)\end{aligned} \tag{21}$$

Besides the case that the desired file f_i for user i is not found in the CVC, the second important case is if the SBS or the users that form a D2D network have file f_i w.p. $P_{\text{CVC}}(f_i)$. Now if user i does not have locally the simultaneously transmitted file f_l from the BS, then the probability that it does not receive f_i is equal to the probability that classic SIC does not decode it (w.p. $\Pr\{C^{\text{SIC}}_{\text{SBS},i} \leq R_1\}$). This case is similar to the baseline system. If f_l is

[5]Even though an outage event has the same impact on the user (not receive the file), for the transmitter the implications are different in the previous two cases: It must retransmit the data in the first case, or it must fetch them from another network node in the second [3]–[8].

locally available (w.p. $P(i, f_l)$), user i can employ CAIC and remove f_l completely leading to an effective point-to-point interference-free channel within the small cell.[6]

B. Implementation of CAIC and a Deterministic Classic D2D Caching Algorithm (D2DSIC)

The plethora of classic caching algorithms that have been presented in the literature [3]–[8], manage the CVC in the D2D network and affect $P(i, f_l)$, i.e., the probability that a certain file is located in specific node. These algorithms try to use efficiently the available space from all users associated to the SBS. Here, we propose one algorithm for interfacing our scheme with these class of classic caching schemes. For demonstration purposes we focus on a class of D2D classic caching systems where caching decisions are deterministic. With deterministic caching the SBS knows which specific user, or network node in general, has a specific file. According to the caching algorithm we will describe shortly, the probability that node i has in its local cache file f_l is:

$$P(i, f_l) = \begin{cases} 1 & \text{if } l = k\%U + 1, 0 \leq k \leq UM, UM \leq F \\ 0 & \text{if } UM < k, UM \leq F \\ 1 & \text{if } 0 \leq k \leq f_{\min}, UM > F \\ 1 & \text{if } l = (F+t)\%U + 1, k = f_{\min} + 1, UM > F \\ & (t = [0, ..., UM - F - f_{\min}(U-1) - 1]) \\ 1 & \text{if } l = (k\%U) + 1, k > f_{\min} + 1, UM > F \\ 0 & \text{if } l \neq (k\%U) + 1, k > f_{\min} + 1, UM > F \end{cases} \tag{22}$$

This function specifies whether file f_l is cached at node i depending on different conditions that involve the id of a file, the size of the file library, the available caching capacity. Note the deterministic nature of this function, i.e., it is effectively an indicator function since a file will either be located or not at a user depending on its id. To explain the D2D caching algorithm captured in (22) clearly, we distinguish two cases. If $UM \leq F$, and if the file

[6]If we need to account for self-requests a term $P_{\text{CVC}}(f_i)P(i, f_i)$ has to be added.

has index k less than UM, this means that it is placed in the CVC with a deterministic modulo operation in the first line above. Since there is a file-f_l-to-user-i assignment policy based on the modulo operation, the probability that a user has the file will be 1 or 0. If the index k is larger than UM then the probability of any user having this file is zero (second line in the above).

Caching is more complicated when condition $UM > F$ is true, i.e., there is extra caching space beyond what the file library requires. The extra space is allocated for CAIC. Data in the virtual CAIC space are only used by the CAIC algorithm. The extra space used for CAIC can be measured with parameter $f_{\min} = \lfloor \frac{UM-F}{U} \rfloor$ that indicates the average extra caching space available per user in terms of files. For example if $f_{\min}$ is equal to 2, it means that every user has two file positions for storing CAIC files which means that it can replicate the 2 most popular files in addition to the files it has in the CVC.

Example 1: To explain (22) more clearly, we present an example with MU=4×6=24, and F=10 in Fig. 3. The extra space is 14 files and $f_{\min}$=2, which means that the first 2 most popular files (with id k=0,1) can be fully replicated at every user leading to a probability of 1 (third line of (22)). This caching decision consumes $f_{\min} \times (U-1)$=10 files. The remaining space of 4 files will be used for starting the replication of file with id k=2 (fourth line of (22) and case k=$f_{\min}$+1). This results in the partial replication of f_2. Hence, with our algorithm popular files enjoy a high degree of replication to improve CAIC opportunities.

Finally, it is easy to calculate the probability that a file f_i is in the CVC:

$$P_{\mathrm{CVC}}(f_i) = \begin{cases} 1 \text{ if } 1 \leq k \leq UM \\ 0 \text{ if } UM < k \leq F \end{cases} \tag{23}$$

VII. Application of Classic D2D Caching and CAIC in HetNets & Performance Evaluation

For cellular systems the highly-efficient short-range communication of SBSs must be complemented with MBSs to ensure umbrella coverage. As we explained this gives rise to the HetNet paradigm where a high power MBS is overlaid with tiers of low-power small

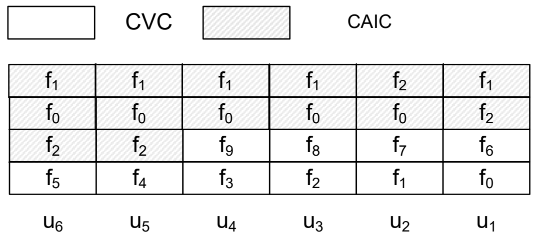


Fig. 3. Interface of classic caching algorithms with CAIC. Contents of the cache at each user according to the proposed caching algorithm for different M.

SBSs that use short-range spectral-efficient communication. Despite the several benefits of the HetNet architecture, it suffers from a major drawback: Intra-cell interference is caused from the MBS to the users communicating with the low power SBSs, when they operate in the same physical time and frequency resources (co-channel deployments). One strategy for handling this type of interference in HetNets, is time-domain resource partitioning where the macrocells shut off their transmission for a subset of the available resources [21]. This technique was standardized in 3GPP LTE Rel. 10 under the more general scheme of enhanced inter-cell interference coordination (eICIC). Users associated to the small cells can achieve higher data rates in almost blank subframes (ABS) since interference from the MBS is limited to the bare minimum [17]. However, even with resource blanking during the MBS transmissions (needed to ensure umbrella coverage for the complete network) in the regular subframes (RS), the small cells experience poor performance. This observation motivated the development of eICIC guidelines that suggest the allocation of ABS to cell edge users and RSs to users closer to the SBS so as to improve efficiency. Still, when the MBS transmits all the users will suffer from lower rate.

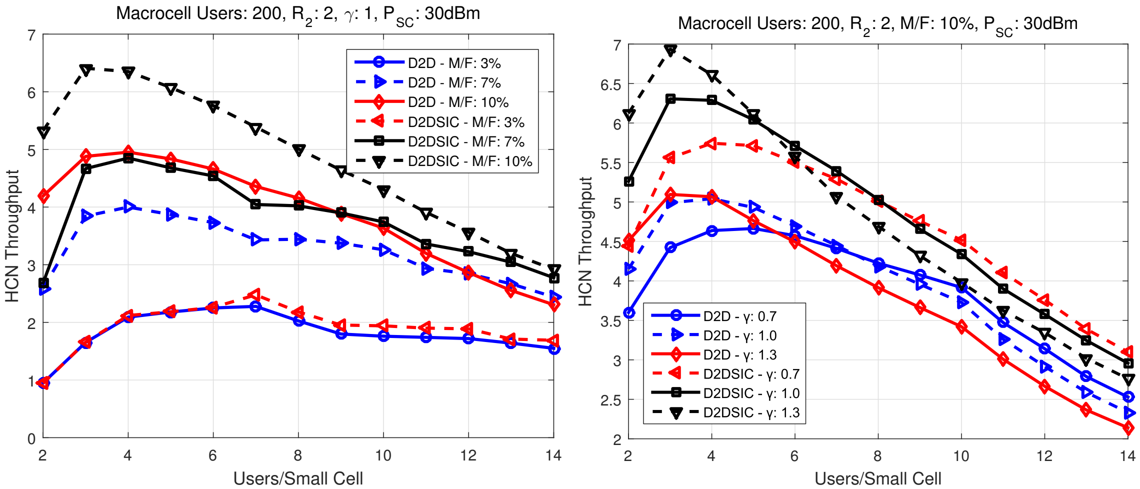
The HetNet scenario is ideal for exploring the performance benefits of CAIC in a modern topology and this is what we do next. More specifically, the role of the BS in the CAIC analysis we presented in the previous sections, is assumed by the MBS, while several small cells exist and deploy independently the joint CAIC and classic D2D caching framework described in Section VI.

A. Simulation Setup for a HetNet

The U users per SBS are assumed to be deployed randomly and uniformly within the small cell they are associated, which is the dominant model in industry and academia [22]. The users are static as we already explained during the description of our channel model. To evaluate the performance under the most challenging HetNet conditions we consider no resource blanking, that is the MBS always transmits. The transmission power of the MBSs is equal to P_{MBS}=46dBm while P_{SC} varies depending on the experiment (based on realistic settings [21]). Regarding the used MCSs, R_2 was set to a constant value for a given experiment. Also R_1 was set to the value that maximizes the throughput. We present results for the throughput of the complete HetNet or the throughput per-user. Each plotted data point was obtained by averaging the results of 2×10^4 topologies. In all the figures our scheme is denoted as D2DSIC since it combines classic D2D caching and CAIC.

B. Result for Different User Cache Size M

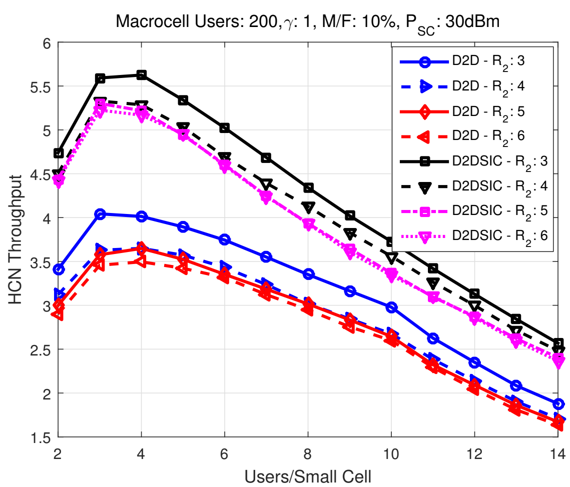
In the figures we compare with state-of-the-art D2D caching systems [3]. Recall that with the classic D2D caching system if a file is locally available it is delivered through the highly-efficient short-range communication. Without any from of cooperation between the MBS and the SBS, this is the optimal scheme [3], [4], [12]. Results for the HetNet throughput v.s. the number U of the users that are associated to each small cell can be seen in Fig. 4(a). For a small number of users/small cell we have many deployed SBSs which means higher spatial reuse but poor caching performance (CVC misses). Regarding CAIC, for higher number of small cells, or low number of users per small cell, the probability that a user has in its respective part of the CVC file f_l is high (given that the file is in the CVC in the first place). Since the CVC stores files based on their popularity, this strategy means



(a) Different M/F. Constant user population of 200 users.

(b) Different γ and M/F =0.1. Constant user population of 200 users.

Fig. 4. Throughput v.s. the small cell users per small cell U.



(a) γ=1, M/F=0.1, and different MBS R_2. Constant user population of 200 users.

(b) γ=1, M/F=0.1, and different SC transmit power $P_{\rm SC}$. Constant user population of 200 users.

Fig. 5. Throughput v.s. the small cell users per small cell U.

that in this low user-per-small-cell regime, users have stored in their own cache the most popular files that can also coincide frequently with the file f_l transmitted from the MBS. For slightly higher number of users/small cell caching performance improves resulting in the observed peaks. However, the overall HetNet performance is then reduced because of lower spatial reuse for both systems since fewer small cells are deployed.

The performance increase is improved as the ratio M/F of the file library that each user stores is increased. This motivates two performance improvement approaches: First the increase of the CVC space to improve the classic caching gain, and second the increase of the CAIC cache per user to improve the CAIC gain by lowering U and increasing M. In our network this effectively means that even though we can keep MU constant and obtain the classic caching gain, we can create cells of smaller radius but with higher storage requirements for the individual users in order to increase the CAIC gain.

In the right part of the x axis in Fig. 4(a), for low number of small cells or high number of users per small cell, the opportunity for having file f_l locally (term $P(i, f_l)$ for CAIC is lower because even though the CVC space MU is higher overall, f_l can be at any of the U users of the CVC reducing thus the opportunities for CAIC. Hence, in this regime there is poor use of the caching resources at the users. This means that CAIC performance converges to that of the D2D system. Still, our scheme offers improved robustness to the presence of the MBS since it can offer an interference-free channel from the MBS more frequently as M is increased.

C. Results for Different γ

The previous results were obtained for γ=1 while next we evaluate different popularity distributions and the results are plotted in Fig. 4(b). When γ is increased, i.e., there is higher degree of content reuse, the performance gains of our scheme are higher especially for small U. Baseline D2D achieves higher performance because the file is found more frequently in the CVC. However, in our system a higher γ means that popular files are transmitted more frequently from the MBS and so CAIC can be used more frequently.

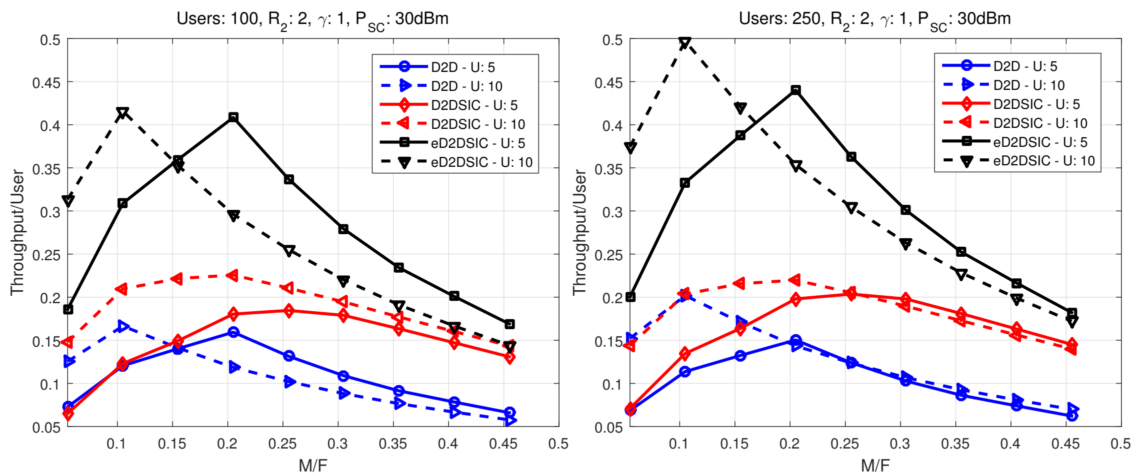
The performance of every scheme worsens for higher values of γ as the number of small cells is decreased. To understand this behavior recall that we do not consider self-requests, i.e., this is the throughput excluding self-requests, which means that for higher γ there are more self requests. One final observation, that is distinguishable for γ=0.7, is that after U=10 the performance drop has a steeper slope which is again attributed to the fact after that point every file is present in the CVC leading to lower gains.

D. Results for Different MBS MCS R_2

Next we study a subset of the parameters related to wireless communication like the MCS and the transmit power of the MBS. This will allow us to evaluate the robustness of CAIC to the high-power source of interference. Regarding the impact of the MCS R_2 used by the MBS, we notice in Fig. 5(a) that the peak performance gain of our scheme is increased as R_2 is decreased. However, the important detail is that for D2D higher R_2 leads to worse SIC performance and in particular the term $\Pr\{C_{\text{SBS},i}^{\text{SIC}} > R_1\}$ as we can see from (16) and (19). That is the classic SIC fails because higher order MCSs have higher probability of a decoding failure. CAIC does not experience the same level of performance degradation because it is based on cache-assisted SIC: The term $\Pr\{C_{\text{SBS},i}^{\text{CAIC}} > R_1\}$ in (19) is a quantity independent of R_2 since it does not have to decode the MBS.

E. Small Cell Transmission Power P_{SC}

Reducing the transmission power of the small cells reduces ICI in the complete HetNet, which is something positive, but also reduces the communication rate in each small cell. Hence, there is a well-known tradeoff involved in this case (the classic power control problem for ICI minimization). The related results for our setup can be seen in Fig. 5(b). For high small cell density, that occurs in the left part of this figure, the performance increase with increased transmission power in the small cells is negated by the reduction in spatial reuse. Hence, all configurations of the CAIC system converge to a HetNet throughput equal with a normalized value of 5. Switching from 30dBm to 33dBm offers minor benefit in this regime. On the other hand, when the density of the small cells is reduced, as we move to right of



(a) γ=1 and constant user population of 100 users. (b) γ=1 and constant user population of 100 users.

Fig. 6. Throughput v.s. M/F and different number of users per small cell U.

the x axis, then higher transmission power leads to higher performance as expected. It is also important to note that the relative gain of CAIC in a very dense network (e.g. for U=14) is higher regardless of the transmit power.

F. The Case of $MU > F$

One of the most interesting results is obtained when the storage capacity of the users is increased. Now we present the throughput per user to emphasize the result more clearly. For 100 macrocell users in Fig. 6(a) we see that our scheme is more robust to variations in U for higher values of M/F. The peak performance for the classic D2D occurs at M/F=0.1 and M/F=0.2 because the CVC contains the entire file library for U=10 and U=5, respectively. Our baseline scheme CAIC experiences a peak for higher M/F since the additional space is used by the CAIC system. For a higher number of 250 users in the complete macrocell in Fig. 6(b) the CAIC throughput with is double than that of D2D for a storage space equal to M/F=0.3. Hence, with higher user density the benefit is that the CVC can be distributed to more users allowing thus the CAIC cache to be increased. This means higher performance.

VIII. Conclusion

In this paper we proposed cache-assisted interference cancellation (CAIC) for wireless communication in small cells that operate under an interfering BS that transmits re-usable files. Cached popular files at the users are used for canceling the transmissions of the BS when it transmits these files. The result is an interference-free channel which means that performance gains in terms of capacity or outage probability come in addition to the well-known benefits of classic caching. We provide capacity and outage analyses that are corroborated with simulation and experimental results. The applicability of our concept is also demonstrated in HetNets with significant performance improvements.

References

[1] "NGMN 5G white paper," 2015.

[2] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge University Press, 2005.

[3] N. Golrezaei, K. Shanmugam, A. Dimakis, A. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in IEEE Infocom, 2012.

[4] N. Golrezaei, P. Mansourifard, A. Molisch, and A. Dimakis, "Base-station assisted device-to-device communications for high-throughput wireless video networks," Wireless Communications, IEEE Transactions on, vol. 13, no. 7, pp. 3665–3676, July 2014.

[5] M. Maddah-Ali and U. Niesen, "Fundamental limits of caching," in ISIT, July 2013, pp. 1077–1081.

[6] M. Ji, A. Tulino, J. Llorca, and G. Caire, "On the average performance of caching and coded multicasting with random demands," in Wireless Communications Systems (ISWCS), 2014 11th International Symposium on, Aug 2014, pp. 922–926.

[7] X. W. Jinbei Zhang, Xiaojun Lin, "Coded caching under arbitrary popularity distributions," in ITA, 2015.

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

[9] L. Xiang, D. W. K. Ng, R. Schober, and V. W. S. Wong, "Cache-enabled physical layer security for video streaming in backhaul-limited cellular networks," IEEE Transactions on Wireless Communications, vol. 17, no. 2, pp. 736–751, Feb 2018.

[10] ——, “Secure video streaming in heterogeneous small cell networks with untrusted cache helpers,” IEEE Transactions on Wireless Communications, vol. 17, no. 4, pp. 2645–2661, April 2018.

[11] M. Mahdian, N. Prakash, M. Médard, and E. Yeh, “Updating content in cache-aided coded multicast,” IEEE Journal on Selected Areas in Communications, pp. 1–1, 2018.

[12] 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.

[13] K. Shanmugam, A. Dimakis, and G. Caire, “Index coding problem with side information repositories,” in Allerton, Oct 2013, pp. 1525–1530.

[14] A. Liu and V. K. N. Lau, “Exploiting base station caching in mimo cellular networks: Opportunistic cooperation for video streaming,” IEEE Transactions on Signal Processing, vol. 63, no. 1, pp. 57–69, Jan 2015.

[15] C. Yang, X. Wang, B. Xia, and H. Ding, “Joint interference cancellation in cache- and sic-enabled networks,” in GLOBECOM 2017 - 2017 IEEE Global Communications Conference, Dec 2017, pp. 1–6.

[16] A. Argyriou, “Wireless network coding with improved opportunistic listening,” IEEE Transactions on Wireless Communications, vol. 8, no. 4, pp. 2014–2023, April 2009.

[17] “3GPP LTE-Advanced,” http://www.3gpp.org/article/lte-advanced, 2010.

[18] E. Bastug, M. Bennis, and M. Debbah, “A transfer learning approach for cache-enabled wireless networks,” in Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015 13th International Symposium on, May 2015, pp. 161–166.

[19] A. Goldsmith, Wireless Communications. Cambridge Univ. Press, 2005.

[20] S. Sen, N. Santhapuri, R. Choudhury, and S. Nelakuditi, “Successive interference cancellation: Carving out mac layer opportunities,” Mobile Computing, IEEE Transactions on, vol. 12, no. 2, pp. 346–357, 2013.

[21] J. Andrews, S. , Q. Ye, X. Lin, and H. Dhillon, “An overview of load balancing in hetnets: old myths and open problems,” IEEE Wireless Communications, vol. 21, no. 2, pp. 18–25, April 2014.

[22] Qualcomm, “D2D deployment scenarios (r1-131412),” April 2013.