Benefits of Cache Assignment on Degraded Broadcast Channels

Shirin Saeedi Bidokhti, Michèle Wigger, and Aylin Yener

Abstract

Degraded K-user broadcast channels (BC) are studied when receivers are facilitated with cache memories. Lower and upper bounds are derived on the capacity-memory tradeoff, i.e., on the largest rate of reliable communication over the BC as a function of the receivers' cache sizes, and the bounds are shown to match for some special cases. The lower bounds are achieved by two new coding schemes that benefit from *non-uniform cache assignment*. Lower and upper bounds are also established on the global capacity-memory tradeoff, i.e., on the largest capacity-memory tradeoff that can be attained by optimizing the receivers' cache sizes subject to a total cache memory budget. The bounds coincide when the total cache memory budget is sufficiently small or sufficiently large, characterized in terms of the BC statistics. For small cache memories, it is optimal to assign all the cache memory to the weakest receiver. In this regime, the global capacity-memory tradeoff grows as the total cache memory budget divided by the number of files in the system. In other words, a *perfect global caching gain* is achievable in this regime and the performance corresponds to a system where all cache contents in the network are available to all receivers. For large cache memories, it is optimal to assign a positive cache memory to every receiver such that the weaker receivers are assigned larger cache memories compared to the stronger receivers. In this regime, the growth rate of the global capacity-memory tradeoff is further divided by the number of users, which corresponds to a local caching gain. Numerical indicate suggest that a uniform cache-assignment of the total cache memory is suboptimal in all regimes unless the BC is completely symmetric. For erasure BCs, this claim is proved analytically in the regime of small cache-sizes.

I. INTRODUCTION

Storing popular contents at or close to the end users improves the network performance during peak-traffic time. The main challenge is that the contents have to be cached before knowing which files the users will request in the peak-traffic period. A conventional approach is to store the same popular contents in the cache memories of the users. This allows the receivers to locally retrieve the contents without burdening the network. However, further caching gains, i.e., the so called *global caching gains*, are possible if different contents are stored at different
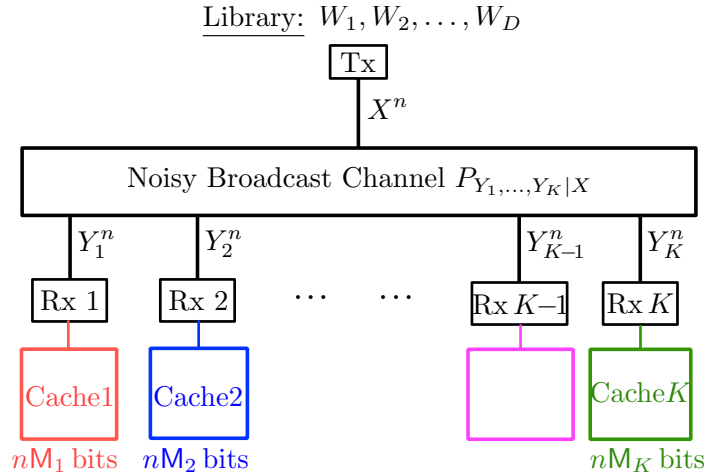
Fig. 1. Noisy broadcast channel with cache memories at the receivers.

users [1]. Specifically, a careful design of the cache contents creates coding opportunities to simultaneously serve multiple users during the peak-traffic periods, henceforth called the *delivery phase*.

In this paper, we focus on the scenario depicted in Figure 1. A transmitter communicates with receivers $1, \ldots, K$ which are equipped with cache memories. The delivery-phase communication takes place over a noisy broadcast channel (BC) where the receivers have access to cache memories. The BC-model has previously been studied in [25]–[42]. The simplified version where the BC is a common noise-free bit-pipe to all users was analyzed in [1]–[16, 21]–[24] under the assumption that all receivers have equal cache sizes, and in [15, 16] under the assumption that various receivers have different cache sizes. Caching was studied for many other scenarios, e.g., for interference networks [44]–[46], hierarchical networks [54]–[56], and cellular networks [46]–[53].

In [25]–[27] the gains of caching in noisy broadcast networks are investigated. Specifically, we have proposed a joint cache-channel coding scheme and focused on erasure BCs with two sets of receivers: a set of cache-aided weak receivers (where each channel has the same erasure probability) and a set of strong receivers without cache memories (where each channel has the same erasure probability). Previous works have adapted a separate cache-channel coding architecture where the encoders (resp. decoders) consist of a cache encoder (resp. decoder) that only exploits the cache contents and a channel encoder (resp. decoder) that only exploits the channel statistics; see Figure 2. By contrast, in a joint cache-channel coding scheme, the encoders and decoders simultaneously exploit the knowledge of the channel statistics and the cache contents, leading to improved performance.

The joint cache-channel coding scheme in [25]–[27] loads (piggybacks) the information that is intended for the strong receivers, but is already cached at the weaker receivers, onto the information that is communicated to
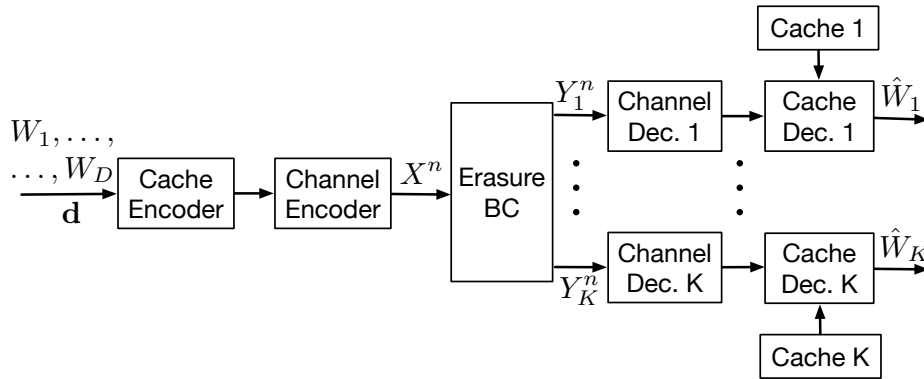
Fig. 2. Separate cache-channel coding architecture.

the weak receivers[1]. When the rate of the piggybacked information is modest, this can be done without harming the decoding performance at the strong receivers. In some sense, piggyback coding provides the stronger receivers virtual access to the weaker receivers' cache-memories as if these cache contents were locally present at the stronger receivers.

The previous works [25]–[27] have shown that when different receivers have different channel statistics, then assigning larger cache memories to the weaker receivers significantly improves the performance compared to the traditional uniform cache assignment. In addition to mitigating the rate-bottleneck at the weaker receivers, non-uniform cache assignment allows to achieve new global caching gains by the means of *joint cache-channel coding* [25]–[27, 32].

Motivated by the new gains of caching in noisy broadcast networks, in this work, we address the problem of efficient cache assignment in broadcast networks and devise two joint cache-channel coding schemes by using piggyback coding, superposition coding, and coded caching.

A. Main Contributions and Implications

The main contributions of the paper are as follows:

- *Superposition-Piggyback Coding*: We generalize the piggyback-coding scheme of [26], that is specific for erasure BCs, to arbitrary BCs with a cache memory only at the weakest receiver and account for different channel qualities in the network by employing superposition coding. We show that this scheme is optimal for small cache memory sizes.

- *Generalized Coded-Caching*: The coded-caching scheme in [1] is generalized to *noisy BCs* with unequal cache sizes. The scheme is optimal for a particular cache assignment.

- *A New Converse Result:* A general converse result is provided for degraded BCs with arbitrary cache sizes at the receivers. It strictly improves over the existing converse results for degraded BCs in [26, 28] and for the

---

[1] The proposed piggyback coding can be seen as a simplified version without binning etc. of "Slepian-Wolf coding over broadcast channels" in [58], which applies to more general scenarios.

noise-free bit-pipe model in [1, 6, 9, 10, 19].

- *Global Capacity-Memory Tradeoff:* Lower and upper bounds are derived on the global capacity-memory tradeoff. They are shown to match when the total available cache memory is small or large. Suboptimality of the popular approach of assigning equal cache memory to all receivers is proved analytically for erasure BCs in the small cache size regime and shown numerically for erasure and Gaussian BCs in all regimes of cache sizes.

More specifically, we first propose a coding scheme that we call *superposition piggyback-coding* by assuming that only the weakest receiver has a cache memory. Using this scheme all receivers gain virtual access to the weakest receiver's cache memory as if the cache contents were locally present at each of these receivers.

The second scheme generalizes the coded caching in [1] to account for different channel statistics and different cache sizes at the receivers. We assign larger cache sizes to the weaker receivers and use piggyback coding to transmit higher rates of information to the stronger receivers without harming the communication to the weaker receivers. As a consequence, the amount of the virtual cache memory that is provided to the stronger receivers increases compared to the original coded-caching scheme, resulting in an improved performance.

The performance criterion of interest in this paper is the *capacity-memory tradeoff.* That is, the largest rate, as a function of the available cache memories, so that the transmitter can reliably send the messages demanded by the receivers over the noisy BC.

We present a new upper bound on the capacity-memory tradeoff of degraded BCs[2] that improves the previous upper bound in [26, 28]. Using the upper bound, we show the optimality of the superposition piggyback-coding scheme when only the weakest receiver has a cache memory and its size is below a certain threshold that depends on the BC statistics. Moreover, we show that the generalized coded-caching scheme is optimal for a particular cache assignment.

When the BC is a noise-free bit-pipe, the upper bound on the capacity-memory tradeoff leads to a lower bound on the delivery rate-memory tradeoff that improves the previous lower bounds in [1, 6, 9, 10, 19].

The upper bound is asymmetric in the cache sizes: the cache memories at the weaker receivers increase the upper bound more than the cache memories at the stronger receivers. In this sense, the upper bound reinforces the intuition obtained from the lower bounds that the capacity-memory tradeoff increases when larger cache memories are assigned to the weaker receivers as compared to the stronger receivers. To make this statement more precise, we derive upper and lower bounds on the *global capacity-memory tradeoff*, where one is allowed to optimize over the cache assignment subject to a global cache constraint. The lower bound is obtained using the following cache-assignment strategy and coding schemes:

- For a small total cache-size M, all of it is assigned to the weakest receiver, and superposition piggyback-coding is applied. This strategy is optimal in the small total cache-size regime and achieves a global capacity-memory tradeoff that grows as $\frac{M}{D}$, where D denotes the total number of files. Thus, in this regime a *perfect global*

---

[2]Since for our purposes only the conditional marginal distributions matter, it suffices that the BC is *stochastically degraded* [61].)

caching gain is achieved, i.e., the same performance as in a systems where *all* cache memories in the network are accessible by *all* the receivers.

- For moderate total cache-size M, generalized coded-caching with parameters $t = 1, \ldots, K - 1$ and the corresponding cache-assignments are employed. The larger the total cache-size, the larger the parameter t needs to be chosen. However, the larger t, the smaller the global caching gain, since with increasing t the overlap of the different cache contents increases as well, and duplicated cache contents cannot provide global caching gain.

- When the total cache-size M equals the total cache memory of generalized coded-caching with parameter $t = K - 1$, then generalized coded-caching is optimal. For total cache memories exceeding this threshold, it is optimal to uniformly assign the *additional* cache memory across the K receivers. This additional cache memory can only bring *local caching gain* and the same content can be stored at all the K receivers. In other words, for total cache memory exceeding a threshold, the global capacity-memory tradeoff grows as $\frac{1}{K} \cdot \frac{\mathsf{M}}{D}$.

Finally, this paper proves analytically that for erasure BCs a uniform cache allocation is *strictly suboptimal* in the regime of small cache memories, unless all receivers have same channel statistics. Numerical simulations show that the same holds for all regimes of cache memory and also for Gaussian BCs.

B. Notation

Random variables are denoted by uppercase letters, e.g. A, their alphabets by matching calligraphic font, e.g. $\mathcal{A}$, and elements of an alphabet by lowercase letters, e.g. $a \in \mathcal{A}$. We also use uppercase letters for deterministic quantities like rate R, capacity C, number of users K, cache size M, and number of files in the library D. Vectors are identified by bold font symbols, e.g., $\mathbf{a}$, and matrices by the font A. We use the shorthand notation A^n for $(A_1, \ldots, A_n)$. The Cartesian product of $\mathcal{A}$ and $\mathcal{A}'$ is $\mathcal{A} \times \mathcal{A}'$, and the n-fold Cartesian product of $\mathcal{A}$ is $\mathcal{A}^n$. $|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$.

Finally, for indices w_1 and w_2 taking value in $\left\{ 1, \ldots, \lfloor 2^{\ell_1} \rfloor \right\}$ and $\left\{ 1, \ldots, \lfloor 2^{\ell_2} \rfloor \right\}$, respectively, we denote by

$$w_1 \bigoplus w_2$$

the index in $\{1, \ldots, \lfloor 2^{\ell_{\max}} \rfloor\}$ that corresponds to the XOR of the length-$\ell_{\max}$ binary representations of w_1 and w_2, where $\ell_{\max} := \max\{\ell_1, \ell_2\}$.

We will be using the abbreviation i.i.d. for *independent and identically distributed.*

C. Outline

The remainder of the paper is organized as follows. Section II describes the problem setup. Section III recalls known results for the scenario without cache memories. The main results of this paper are described in Sections IV and V, followed by applications of these results to erasure and Gaussian BCs, see Section VI. The paper is concluded with a summary and conclusions, Section VII and various technical appendices contain the proofs of the results in Sections V and VI.

II. PROBLEM DEFINITION

Consider a transmitter and receivers $1, \ldots, K$. The transmitter has access to a library with D independent messages, $W_1, \ldots, W_D$, each distributed uniformly over the set $\{1, \ldots, \lfloor 2^{nR} \rfloor\}$. Here, $R \geq 0$ denotes the rate of transmission and n is the transmission blocklength. We assume that there are more messages than receivers:

$$D \geq K. \tag{1}$$

Each receiver $k \in \mathcal{K} := \{1, \ldots, K\}$ is equipped with a cache of size $\mathsf{M}_k \geq 0$. Communication takes place in two phases. For the first, i.e., the placement phase, the transmitter chooses caching functions

$$g_k \colon \{1, \ldots, \lfloor 2^{nR} \rfloor\}^D \to \{1, \ldots, \lfloor 2^{n\mathsf{M}_k} \rfloor\}, \qquad k \in \mathcal{K}, \tag{2}$$

and places

$$\mathbb{V}_k := g_k(W_1, \ldots, W_D) \tag{3}$$

in receiver k's cache. This phase takes place in a noiseless fashion.[3]

The subsequent delivery phase takes place over a *degraded* BC [59] with finite input alphabet $\mathcal{X}$, finite output alphabets $\mathcal{Y}_1, \ldots, \mathcal{Y}_K$,[4] and channel transition law

$$\Gamma(y_1, \ldots, y_K | x), \text{ for } x \in \mathcal{X}, y_1 \in \mathcal{Y}_1, \ldots, y_K \in \mathcal{Y}_K \tag{4}$$

which decomposes as

$$\Gamma(y_1, \ldots, y_K | x) = \Gamma_K(y_K | x) \cdot \Gamma_{K-1}(y_{K_1} | y_K) \cdots \Gamma_1(y_1 | y_2). \tag{5}$$

Without loss in generality, we order the receivers from the weakest to the strongest.

At the beginning of the delivery phase, each receiver k demands message W_{d_k}, $d_k \in \mathcal{D} := \{1, \ldots, D\}$. Transmitter and all the receivers are informed of the demand vector

$$\mathbf{d} := (d_1, \ldots, d_K).$$

Using this information, the transmitter forms the channel input sequence $X^n = (X_1, \ldots, X_n)$ as

$$X^n = f_{\mathbf{d}}(W_1, \ldots, W_D) \tag{6}$$

for some encoding function $f_{\mathbf{d}} : \{1, \ldots, \lfloor 2^{nR} \rfloor\}^D \to \mathcal{X}^n$.

Receiver $k \in \mathcal{K}$ observes the channel output sequence $Y_k^n := (Y_{k,1}, \ldots, Y_{k,n})$. Given the demand vector $\mathbf{d}$, cache content $\mathbb{V}_k$, and channel outputs Y_k^n, it produces its estimate of the desired message W_{d_k},

$$\hat{W}_k := \varphi_{k,\mathbf{d}}(Y_k^n, \mathbb{V}_k), \tag{7}$$

---

[3]Following previous works on caching systems, we will also assume that the placement phase takes place in low-traffic hours with abundance of bandwidth resources, and can be considered noiseless.

[4]The results of this paper readily extends to continuous alphabets. We will consider Gaussian BCs in Section VI-C.

by means of a decoding function

$$\varphi_{k,\mathbf{d}} \colon \mathcal{Y}_k^n \times \{1, \ldots, \lfloor 2^{n\mathsf{M}_k} \rfloor\} \to \{1, \ldots, \lfloor 2^{nR} \rfloor\}. \tag{8}$$

The worst-case probability of error at any receiver and any demand $\mathbf{d}$ is given by

$$\mathsf{P}_{\mathsf{e}} := \mathbb{P}\left[\bigcup_{\mathbf{d} \in \mathcal{D}^K} \bigcup_{k=1}^{K} \left\{ \hat{W}_k \neq W_{d_k} \right\} \right]. \tag{9}$$

A rate-memory tuple $(R, \mathsf{M}_1, \ldots, \mathsf{M}_K)$ is *achievable* if for any $\epsilon > 0$ there exists a sufficiently large blocklength n and caching, encoding, and decoding functions as in (3), (6), and (7) so that $\mathsf{P}_{\mathsf{e}} \leq \epsilon$.

Definition 1: The *capacity-memory tradeoff* $\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K)$ is the largest rate R for which the rate-memory tuple $(R, \mathsf{M}_1, \ldots, \mathsf{M}_K)$ is achievable:

$$\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) := \sup\{R \colon (R, \mathsf{M}_1, \ldots, \mathsf{M}_K) \text{ achievable}\}.$$

Our main goal in this paper is to optimize the cache assignment $(\mathsf{M}_1, \ldots, \mathsf{M}_K)$ to attain the largest capacity-memory tradeoff $\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K)$ under the total cache constraint:

$$\sum_{k=1}^{K} \mathsf{M}_k \leq \mathsf{M}. \tag{10}$$

Definition 2: The *global capacity-memory tradeoff* $\mathsf{C}^\star(\mathsf{M})$ is defined as:

$$\mathsf{C}^\star(\mathsf{M}) := \max_{\substack{\mathsf{M}_1, \ldots, \mathsf{M}_K > 0 \colon \\ \sum_{k=1}^{K} \mathsf{M}_k \leq \mathsf{M}}} \mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K). \tag{11}$$

Remark 1: The global capacity memory tradeoff depends on the BC law $\Gamma(y_1, \ldots, y_K|x)$ only through its marginal conditional laws. All our results thus also apply to *stochastically degraded BCs*.

A. Minimum Delivery Rate

Previous works on caching that modelled the BC as a noise-free bit-pipe, e.g., [1], adopted a "source-coding perspective" as opposed to a "channel coding perspective" as we have presented above. In the source coding perspective, each message is an $F > 0$ bits packet, the delivery communication consists of $\rho \cdot F$ channel uses, and receiver k has $m_k F$ bits of cache memory, $k = 1, \ldots, K$. The *delivery rate* ρ is said to be achievable given normalized memory sizes $m_1, \ldots, m_K$ if there exist caching, encoding, and decoding functions such that the probability of error in (9) tends to 0 as $F \to \infty$.

The following correspondence holds between the two perspectives:

$$R \text{ achievable with } (\mathsf{M}_1, \ldots, \mathsf{M}_K)$$

$$\text{under the "channel-coding perspective"}$$

$$\Longleftrightarrow$$

$$\rho = \frac{1}{R} \text{ achievable with } \left(m_1 = \frac{\mathsf{M}_1}{R}, \ldots, m_K = \frac{\mathsf{M}_K}{R} \right)$$

under the "source-coding perspective".

For simplicity, we will adopt the "source-coding perspective" in Section VI-B where we specialize the new upper bound on the capacity-memory tradeoff $C(M_1, \ldots, M_K)$ to the noise-free bit-pipe model with uniform cache assignment in [1]. For other BCs, we use the "channel-coding perspective" in line with similar setups in network information theory.

III. PRELIMINARIES: CAPACITIES WITHOUT CACHE MEMORIES

In the absence of cache memories,

$$M_1 = \ldots = M_2 = 0,$$

the capacity-memory tradeoff $C(M_1 = 0, \ldots, M_K = 0)$ is well known: It is the largest symmetric rate R with which K independent messages can be reliably sent to the K receivers. I.e.,

$$C(M_1 = 0, \ldots, M_K = 0) = C_{\mathcal{K}} \tag{12}$$

where [59]:

$$C_{\mathcal{K}} := \max \, \min \big\{ I(U_1; Y_1), \; I(U_2; Y_2|U_1), \; \ldots, \; I(U_{K-1}; Y_{K-1}|U_{K-2}), \; I(X; Y_K|U_{K-1}) \big\}, \tag{13}$$

and the maximization in (13) is over all random tuples $U_1, \ldots, U_{K-1}, X, Y_1, \ldots, Y_K$ forming the Markov chain

$$U_1 - U_2 - \cdots - U_{K-1} - X - (Y_1, \ldots, Y_K) \tag{14a}$$

satisfying the channel transition law

$$P_{Y_1 \ldots Y_K | X}(y_1, \ldots, y_K | x) = \Gamma(y_1, \ldots, y_K | x). \tag{14b}$$

To present the results in this paper, we will need the *capacity region* without cache memories of the BC to a subset of the receivers

$$\mathcal{S} := \{ j_1, \ldots, j_{|\mathcal{S}|} \} \subseteq \mathcal{K}, \quad j_1 < \cdots < j_{|\mathcal{S}|}. \tag{15}$$

This capacity region $\mathbf{C}_{\mathcal{S}}$ [59] is given by the set of all nonnegative rate-tuples $(R_1, \ldots, R_{|\mathcal{S}|})$ for which there exist random variables $U_1, \ldots, U_{|\mathcal{S}|-1}, X, Y_{j_1}, \ldots, Y_{j_{|\mathcal{S}|}}$ satisfying (14b) and forming the Markov chain

$$U_1 - U_2 - \cdots - U_{|\mathcal{S}|-1} - X - \big(Y_{j_1}, \ldots, Y_{j_{|\mathcal{S}|}} \big), \tag{16}$$

such that the following conditions hold:

$$R_1 \leq I(U_1; Y_{j_1}), \tag{17a}$$

$$R_k \leq I(U_k; Y_{j_k}|U_k), \quad k \in \{2, \ldots, |\mathcal{S}| - 1\}, \tag{17b}$$

$$R_{|\mathcal{S}|} \leq \; I\big(X; Y_{j_{|\mathcal{S}|}} \big| U_{|\mathcal{S}|-1} \big). \tag{17c}$$

We denote by $C_{\mathcal{S}}$ the largest symmetric rate $R \geq 0$ in $\mathbf{C}_{\mathcal{S}}$:

$$C_{\mathcal{S}} := \max_{R \geq 0} \{ R \colon (R, \ldots, R) \in \mathbf{C}_{\mathcal{S}} \}. \tag{18}$$

It equals

$$\mathsf{C}_{\mathcal{S}} = \max \ \min \big\{ I(U_1; Y_{j_1}), \ I(U_2; Y_{j_2}|U_1), \ \ldots, \ I(U_{|\mathcal{S}|-1}; Y_{j_{|\mathcal{S}|-1}}|U_{|\mathcal{S}|-2}), \ I(X; Y_{j_{|\mathcal{S}|}}|U_{|\mathcal{S}|-1}) \big\}, \tag{19}$$

where the maximization is over all random tuples $U_1, \ldots, U_{|\mathcal{S}|-1}, X, Y_{j_1}, \ldots, Y_{j_{|\mathcal{S}|}}$ that satisfy (14b) and (16).

Notice that $\mathsf{C}_{\{k\}}$ is simply the point-to-point capacity to receiver k and we will abbreviate it as C_k.

IV. Coding Schemes and Lower Bounds on the (Global) Capacity-Memory Tradeoff

A. The Local Caching Gain

The simplest way to use receiver cache memories is to store the same information at each and every receiver. This allows the receivers to retrieve this information locally, without transmission over the BC. Further global caching gains are not possible under this caching strategy.

Applying the described caching strategy to only a part of the cache memory that is of size $\Delta \geq 0$, while allowing a smarter use of the remaining memory, leads to the following proposition, see also [47, Proposition 1].

Proposition 1 (Local caching gain): For all $\Delta > 0$ and $\mathsf{M}_1, \ldots, \mathsf{M}_K \geq 0$:

$$\mathsf{C}(\mathsf{M}_1 + \Delta, \ldots, \mathsf{M}_K + \Delta) \geq \mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) + \frac{\Delta}{D}. \tag{20}$$

As a consequence, for all $\Delta_{\text{total}} > 0$ and $\mathsf{M} \geq 0$:

$$\mathsf{C}^{\star}(\mathsf{M} + \Delta_{\text{total}}) \geq \mathsf{C}^{\star}(\mathsf{M}) + \frac{\Delta_{\text{total}}}{K \cdot D}. \tag{21}$$

We will see that in some regimes this lower bound is tight.

B. Superposition Piggyback-Coding

We generalize the piggyback coding for erasure BCs in [26, 28] to general degraded BCs by introducing superposition coding. The idea is to piggyback information of multiple stronger receivers on that of a single weak receiver. This scheme is efficient when a receiver is strictly weaker than the others. Specifically, we assume

$$I(U_1^{\star}; Y_1) < I(U_1^{\star}; Y_k), \quad k \in \{2, \ldots, K\}, \tag{22}$$

where $(U_1^{\star}, \ldots, U_{K-1}^{\star}, X^{\star})$ is a random K-tuple that achieves the symmetric-capacity $\mathsf{C}_{\mathcal{K}}$, i.e., it is a solution to the optimization problem in (13).

Preliminaries: Let $\epsilon > 0$ be arbitrary small, and define the rates

$$R^{(\text{A})} := \mathsf{C}_{\mathcal{K}} - \epsilon, \tag{23a}$$

$$R^{(\text{B})} := \frac{1}{K-1} \big(I(U_1^{\star}; Y_2) - I(U_1^{\star}; Y_1) \big). \tag{23b}$$

The RHS of (23b) is positive by (22).

Split each message W_d, $d \in \{1, \ldots, D\}$, into two parts:

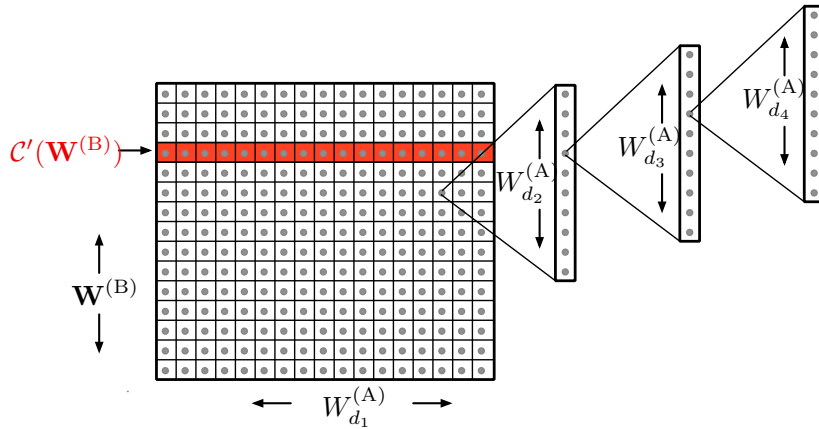$$W_d = \big(W_d^{(\text{A})}, W_d^{(\text{B})} \big),$$

Fig. 3. Codebook $\mathcal{C}$ for superposition piggyback coding for $K = 4$. Each dot represents a codeword.

where $W_d^{(A)}$ and $W_d^{(B)}$ are of rates $R^{(A)}$ and $R^{(B)}$, and thus the total message rate is

$$R = R^{(A)} + R^{(B)}. \tag{24}$$

Define

$$\mathsf{M}_1^{\mathsf{single}} := D \cdot R^{(B)} = \frac{D}{K-1}\big(I(U_1^\star; Y_2) - I(U_1^\star; Y_1)\big), \tag{25}$$

and allocate the cache size

$$\mathsf{M}_1 = \mathsf{M}_1^{\mathsf{single}} \tag{26a}$$

to receiver 1 and zero cache size to the other receivers

$$\mathsf{M}_2 = \ldots = \mathsf{M}_K = 0. \tag{26b}$$

Placement Phase: Store $W_1^{(B)}, \ldots, W_D^{(B)}$ in the cache memory of receiver 1. This is possible by (26a).

Delivery Phase: For the transmission in the delivery phase, construct a K-level superposition code $\mathcal{C}$ with a cloud center of rate $R^{(A)} + (K-1)R^{(B)}$ and satellites of rates $R^{(A)}$ in Levels $2, \ldots, K$. For the code construction, use a probability distribution

$$P_{U_1^\star} \cdot P_{U_2^\star|U_1^\star} \ldots P_{U_{K-1}^\star|U_{K-2}^\star} \cdot P_{X^\star|U_{K-1}^\star}$$

that achieves $\mathsf{C}_{\mathcal{K}}$.

It will be convenient to arrange the codewords in the cloud center in an array with $\lfloor 2^{nR^{(A)}} \rfloor$ columns and $(\lfloor 2^{nR^{(B)}} \rfloor)^{K-1}$ rows. The columns are used to encode message $W_{d_1}^{(A)}$ and the rows to encode the message tuple

$$\mathbf{W}^{(B)} := \big(W_{d_2}^{(B)}, \ldots, W_{d_{K-1}}^{(B)}, W_{d_K}^{(B)}\big). \tag{27}$$

The k-th level satellite is used to encode message $W_{d_k}^{(A)}$, for $k \in \{2, \ldots, K\}$. See Figure 3 for an illustration of the code construction.

Let $u_1^n(w_{1,\text{column}}, w_{1,\text{row}})$ denote the cloud-center codeword of $\mathcal{C}$ in column $w_{1,\text{column}}$ and row $w_{1,\text{row}}$. Similarly, let $x^n(w_{1,\text{column}}, w_{1,\text{row}}; w_2; w_3; \ldots; w_K)$ denote the Level-K satellite codeword of $\mathcal{C}$ that corresponds to the cloud center codeword $u_!^n(w_{1,\text{column}}, w_{1,\text{row}})$ and to the w_2-th, w_3-th ,w_4-th, etc. satellite codewords in Levels $2, 3, 4, \ldots$.

The transmitter chooses and sends the codeword

$$x^n\left(W_{d_1}^{(\text{A})}, \mathbf{W}^{(\text{B})};\ W_{d_2}^{(\text{A})};\ W_{d_3}^{(\text{A})};\ \ldots\ ; W_{d_K}^{(\text{A})}\right)$$

over the channel.

Decoding: Receiver $k \in \{2, \ldots, K\}$, decodes all messages in Levels $1, \ldots, k$. Recall that its desired message parts $W_{d_k}^{(\text{A})}$ and $W_{d_k}^{(\text{B})}$ are encoded in levels k and 1 (i.e., the cloud center), respectively.

Receiver 1 only has to decode $W_{d_1}^{(\text{A})}$, because it can retrieve $W_{d_1}^{(\text{B})}$ directly from its cache memory. To decode $W_{d_1}^{(\text{A})}$ it performs the following steps:

1) It retrieves the message-tuple $\mathbf{W}^{(\text{B})}$ from its cache memory.

2) It forms the subcodebook $\mathcal{C}'(\mathbf{W}^{(\text{B})}) \subseteq \mathcal{C}$ that contains all level-1 codewords that are "compatible" with the retrieved tuple $\mathbf{W}^{(\text{B})}$:

$$\mathcal{C}'(\mathbf{W}^{(\text{B})}) := \left\{u_1^n(w, \mathbf{W}^{(\text{B})})\right\}_{w=1}^{\left\lfloor 2^{nR^{(\text{A})}} \right\rfloor}. \tag{28}$$

Figure 3 illustrates such a subcodebook in red.

3) It decodes its desired message $W_{d_1}^{(\text{A})}$ using an optimal decoding rule for subcodebook $\mathcal{C}'(\mathbf{W}^{(\text{B})})$.

Error Analysis: Each receiver $k \in \{2, \ldots, K\}$ reliably decodes messages $(W_{d_1}^{(\text{A})}, W_{d_2}^{(\text{B})}, \ldots, W_{d_K}^{(\text{B})})$ and $W_{d_2}^{(\text{A})}, \ldots, W_{d_k}^{(\text{A})}$ if the following inequalities hold:

$$kR^{(\text{A})} + (K-1) \cdot R^{(\text{B})} < I(U_k^\star; Y_k), \tag{29a}$$

$$(k - \ell) \cdot R^{(\text{A})} < I(U_k^\star; Y_k | U_\ell^\star), \quad \ell \in \{1, \ldots, k-1\}. \tag{29b}$$

One can verify that for degraded BCs the choice of $R^{(\text{A})}$ and $R^{(\text{B})}$ in (23) satisfies the constraints in (29).

Finally, receiver 1 can decode with arbitrarily small probability of error because subcodebook $\mathcal{C}'(\mathbf{W}^{(\text{B})})$ contains $\lfloor 2^{nR^{(\text{A})}} \rfloor$ codewords that are generated i.i.d. according to $P_{U_1^\star}$ and because

$$R^{(\text{A})} < I(U_1^\star; Y_1).$$

Letting $\epsilon \to 0$, we obtain the following result.

Theorem 2: Under cache assignment (26), we have

$$\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) \geq \mathsf{C}_\mathcal{K} + \frac{\mathsf{M}_1}{D}. \tag{30}$$

Remark 2: Since receivers can always choose to ignore their cache memories, and because the superposition piggyback coding scheme can be time- and memory-shared with a no-caching scheme, Theorem 2 remains valid for all

$$0 \leq \mathsf{M}_1 \leq \mathsf{M}_1^{\text{single}}, \tag{31}$$

$$M_2, \ldots, M_K \geq 0. \tag{32}$$

We will see in Corollary 7 ahead, that (30) holds with equality for all $0 \leq M_1 \leq M_1^{\text{single}}$ provided that $M_2 = \ldots = M_K = 0$.

The RHS of (30) coincides with the capacity-memory tradeoff of a scenario where each and every receiver has access to receiver 1's cache memory. Superposition piggyback coding can thus be viewed as a coding technique that virtually provides all stronger receivers access to the weakest receiver's cache memory. This is achieved by transmitting the extra-message tuple $\mathbf{W}^{(B)}$ in the cloud center and by adapting the decoding at receiver 1 in a way that this additional communication does not influence its decoding performance.

C. Generalized Coded-Caching

We generalize the coded-caching scheme of [1] to noisy BCs with unequal channel conditions and to receivers with unequal cache sizes.

We first explain the scheme for a simple special case.

1) Special Case $K = 2$ and $t = 1$: Fix an input distribution P_X and a small $\epsilon > 0$, and define the rates

$$R^{(A)} = I(X; Y_1) - \epsilon \tag{33}$$

$$R^{(B)} = I(X; Y_2) - \epsilon. \tag{34}$$

Notice that by the degradedness of the BC:

$$R^{(B)} \geq R^{(A)}. \tag{35}$$

Fix a blocklength n and generate a random codebook

$$\mathcal{C} := \left\{ x^n(j) \right\}_{j=1}^{\lfloor 2^{nR^{(B)}} \rfloor} \tag{36}$$

by choosing all entries i.i.d. according to P_X. The codebook $\mathcal{C}$ is revealed to all terminals of the network.

Allocate cache memories

$$M_1 = D \cdot R^{(B)} = D \cdot (I(X; Y_2) - \epsilon), \tag{37a}$$

$$M_2 = D \cdot R^{(A)} = D \cdot (I(X; Y_1) - \epsilon), \tag{37b}$$

to receivers 1 and 2, respectively.

Split each message W_d, for $d \in \{1, \ldots, D\}$, into two parts:

$$W_d = \left(W_d^{(A)}, W_d^{(B)} \right),$$

which are of rates $R^{(A)}$ and $R^{(B)}$, respectively.

In the caching phase, the transmitter stores messages

$$W_1^{(B)}, \ldots, W_D^{(B)}$$

in receiver 1's cache memory and messages

$$W_1^{(A)}, \ldots, W_D^{(A)}$$

in receiver 2's cache memory. This is possible given the cache assignment in (37).

In the delivery phase the transmitter uses codebook $\mathcal{C}$ to send the XOR message[5]

$$W_{d_1}^{(A)} \bar{\oplus} W_{d_2}^{(B)} \tag{38}$$

to both receivers using the codeword

$$x^n \left(W_{d_1}^{(A)} \bar{\oplus} W_{d_2}^{(B)} \right).$$

Receiver 2 decodes the XOR-message, and XORs the decoded message with $W_{d_1}^{(A)}$, which it has stored in its cache memory. It then combines this guess of $W_{d_2}^{(B)}$ with the message $W_{d_2}^{(A)}$ from its cache memory.

Receiver 1 performs joint cache-channel decoding where it can exploit that it has more cache memory than receiver 2. Specifically, it retrieves $W_{d_2}^{(B)}$ from its cache memory, and extracts a subcodebook $\mathcal{C}'(W_{d_2}^{(B)}) \subseteq \mathcal{C}$ containing all codewords that are compatible with $W_{d_2}^{(B)}$:

$$\mathcal{C}'(W_{d_2}^{(B)}) := \left\{ w \bar{\oplus} W_{d_2}^{(B)} \right\}_{w=1}^{\left\lfloor 2^{nR^{(A)}} \right\rfloor} \tag{39}$$

Note that subcodebook $\mathcal{C}'(W_{d_2}^{(B)})$ is of rate $R^{(A)}$ which is smaller than the rate $R^{(B)}$ of the original codebook $\mathcal{C}$.

Receiver 1 then decodes the XOR message in (38) using an optimal decoding rule for this subcodebook $\mathcal{C}'(W_{d_2}^{(B)})$, and it XORs the decoded message with $W_{d_2}^{(B)}$, which it has stored in its cache memory. It then combines the resulting guess of $W_{d_1}^{(A)}$ with the message $W_{d_1}^{(B)}$ from its cache memory.

Since both receivers correctly guess their desired messages W_{d_1} and W_{d_2} whenever they successfully decode the XOR-message in (38), and since the rate $R^{(B)}$ of the original codebook $\mathcal{C}$ satisfies

$$R^{(B)} < I(X; Y_2), \tag{40}$$

and the rate of $R^{(A)}$ of the subcodebook $\mathcal{C}'(W_{d_2}^{(B)})$ satisfies

$$R^{(A)} < I(X; Y_1), \tag{41}$$

the probability of decoding error at both receivers tends to 0 as the blocklength n tends to infinity.

Letting $\epsilon \to 0$, we conclude that for $K = 2$ the rate-memory triple

$$R = I(X; Y_1) + I(X; Y_2),$$

$$M_1 = I(X; Y_2),$$

$$M_2 = I(X; Y_1),$$

is achievable.

---

[5]Recall that in Section I-B we defined the XOR operation $\bar{\oplus}$ over the binary representations of the two messages of same length.

Notice that the weaker receiver 1 is assigned a larger cache memory than the stronger receiver 2:

$$\mathsf{M}_1 \geq \mathsf{M}_2. \tag{42}$$

The described scheme can also be applied with a uniform cache assignment $\mathsf{M}_1 = \mathsf{M}_2 = D \cdot R^{(\mathrm{A})}$, however at the cost of a decreased achievable rate $R = 2 \cdot I(X; Y_1)$. In fact, assigning a larger cache memory M_1 to receiver 1 allows to transmit more information to receiver 2 during the communication to receiver 1.

2) General Scheme: We will need the following definitions. Let for each $t \in \mathcal{K}$

$$\mathcal{G}_1^{(t)}, \ldots, \mathcal{G}_{\binom{K}{t}}^{(t)} \tag{43a}$$

denote all unordered size-t subsets of $\mathcal{K}$. Define their complements as:

$$\mathcal{G}_\ell^{(t),c} := \mathcal{K} \backslash \mathcal{G}_\ell^{(t)}, \qquad \ell \in \left\{ 1, \ldots, \binom{K}{t} \right\}. \tag{43b}$$

Pick a small number $\epsilon > 0$ and an input distribution P_X. Pick further a parameter $t \in \{1, \ldots, K-1\}$, and assign the following cache size to receiver $k \in \mathcal{K}$:

$$\mathsf{M}_k^{(t)} := D \cdot \frac{\sum_{\{\ell : \ k \in \mathcal{G}_\ell^{(t)}\}} \prod_{k' \in \mathcal{G}_\ell^{(t),c}} I(X; Y_{k'})}{\sum_{j=1}^{\binom{K}{t+1}} \prod_{k' \in \mathcal{G}_j^{(t+1),c}} I(X; Y_{k'})}$$
$$- D \binom{K-1}{t-1} \epsilon. \tag{44}$$

Notice that

$$\mathsf{M}_1^{(t)} \leq \mathsf{M}_2^{(t)} \leq \cdots \leq \mathsf{M}_K^{(t)}, \qquad t \in \{1, \ldots, K-1\},$$

so a larger cache memory is assigned the weaker a receiver is.

Split each message W_d into $\binom{K}{t}$ independent submessages:

$$W_d = \left\{ W_{d, \mathcal{G}_\ell^{(t)}} : \quad \ell = 1, \ldots, \binom{K}{t} \right\},$$

where each submessage $W_{d, \mathcal{G}_\ell^{(t)}}$ is of rate

$$R_{\mathcal{G}_\ell^{(t)}} := \frac{\prod_{k \in \mathcal{G}_\ell^{(t),c}} I(X; Y_k)}{\sum_{j=1}^{\binom{K}{t+1}} \prod_{k \in \mathcal{G}_j^{(t+1),c}} I(X; Y_k)} - \epsilon. \tag{45}$$

The total message rate is thus

$$R := \sum_{\ell=1}^{\binom{K}{t}} R_{\mathcal{G}_\ell^{(t)}} = \frac{\sum_{\ell=1}^{\binom{K}{t}} \prod_{k \in \mathcal{G}_\ell^{(t),c}} I(X; Y_k)}{\sum_{j=1}^{\binom{K}{t+1}} \prod_{k \in \mathcal{G}_j^{(t+1),c}} I(X; Y_k)} - \binom{K}{t} \epsilon. \tag{46}$$

Notice that when $t = K - 1$ the denominator of (44), (45), and (46) all equal 1.

Placement Phase: For each $d \in \{1, \ldots, D\}$, store the tuple

$$\left\{ W_{d, \mathcal{G}_\ell^{(t)}} : \ k \in \mathcal{G}_\ell^{(t)} \right\}. \tag{47}$$

in the cache memory of receiver $k \in \mathcal{K}$. This is possible by (45) and the cache assignment in (44).

Delivery Phase: Transmission in the delivery phase takes place in $\binom{K}{t+1}$ subphases.

A given subphase $j \in \left\{1, \ldots, \binom{K}{t+1}\right\}$ is of length

$$n_j := \left\lfloor n \cdot \frac{\prod_{k \in \mathcal{G}_j^{(t+1),c}} I(X;Y_k)}{\sum_{j'=1}^{\binom{K}{t+1}} \prod_{k \in \mathcal{G}_{j'}^{(t+1),c}} I(X;Y_k)} \right\rfloor, \tag{48}$$

and is used to transmit messages

$$\left\{ W_{d_k, \mathcal{G}_j^{(t+1)} \backslash \{k\}} \right\}_{k \in \mathcal{G}_j^{(t+1)}} \tag{49}$$

to the intended receivers in $\mathcal{G}_j^{(t+1)}$. For this purpose, the transmitter creates the XOR message

$$W_{\mathrm{XOR},\mathcal{G}_j^{(t+1)}} = \overline{\bigoplus}_{k \in \mathcal{G}_j^{(t+1)}} W_{d_k, \mathcal{G}_{j+1}^{(t)} \backslash \{k\}}, \tag{50}$$

which is of rate

$$R_{\mathrm{XOR},\mathcal{G}_j^{(t+1)}} := \max_{\mathcal{G}_\ell^{(t)} \subseteq \mathcal{G}_j^{(t+1)}} R_{\mathcal{G}_\ell^{(t)}}$$

$$= \left(\max_{k' \in \mathcal{G}_j^{(t+1)}} I(X;Y_{k'}) \right) \cdot \frac{\prod_{k \in \mathcal{G}_j^{(t+1),c}} I(X;Y_k)}{\sum_{j=1}^{\binom{K}{t+1}} \prod_{k \in \mathcal{G}_j^{(t+1),c}} I(X;Y_k)} - \epsilon, \tag{51}$$

and generates a codebook

$$\mathcal{C}_j = \left\{ x_j^{n_j}(w) \colon w = 1, \ldots, \left\lfloor 2^{nR_{\mathrm{XOR},\mathcal{G}_j^{(t+1)}}} \right\rfloor \right\}, \tag{52}$$

by drawing all entries i.i.d. according to P_X.

The transmitter then sends the codeword

$$x_j^{n_j} \left(W_{\mathrm{XOR},\mathcal{G}_j^{(t+1)}} \right) \tag{53}$$

over the channel.

We now describe the decoding. Each receiver $k \in \mathcal{K}$ can retrieve messages

$$\left\{ W_{d_k, \mathcal{G}_\ell^{(t)}} \colon \ k \in \mathcal{G}_\ell^{(t)} \right\} \tag{54}$$

directly from its cache, see (47), and thus only needs to decode messages

$$\left\{ W_{d_k, \mathcal{G}_\ell^{(t)}} \colon \ k \notin \mathcal{G}_\ell^{(t)} \right\}. \tag{55}$$

For each $j \in \left\{1, \ldots, \binom{K}{t+1}\right\}$ and $k \in \mathcal{G}_j^{(t+1)}$, receiver k decodes message $W_{d_k, \mathcal{G}_j^{(t+1)} \backslash \{k\}}$ from its subphase-j outputs

$$Y_{k,j}^{n_j} := \left(Y_{k, \sum_{j'=1}^{j-1} n_{j'}+1}, \ldots, Y_{k, \sum_{j'=1}^{j} n_{j'}} \right).$$

Specifically, with the messages stored in its cache memory, it forms the XOR message

$$W_{\mathrm{XOR},j,k} := \overline{\bigoplus}_{k' \in \mathcal{G}_j^{(t+1)} \backslash \{k\}} W_{d_{k'}, \mathcal{G}_j^{(t+1)} \backslash \{k'\}}, \tag{56}$$

and it extracts a subcodebook $\mathcal{C}'_{j,k}(W_{\text{XOR},j,k})$ from $\mathcal{C}_j$ that contains all codewords that are compatible with $W_{\text{XOR},j,k}$:

$$\mathcal{C}'_{j,k}(W_{\text{XOR},j,k}) := \left\{ x_j^{n_j}\left(w \barwedge W_{\text{XOR},j,k} \right): \quad w = 1, \ldots, \left\lfloor 2^{nR_{\mathcal{G}_j^{(t+1)}\setminus\{k\}}} \right\rfloor \right\}.$$

It then decodes the XOR message $W_{\text{XOR},\mathcal{G}_j^{(t+1)}}$ by applying an optimal decoding rule for subcodebook $\mathcal{C}'_{j,k}(W_{\text{XOR},j,k})$ to the subphase-j outputs $Y_{k,j}^{n_j}$, and XORs the resulting guess $\hat{W}_{\text{XOR},\mathcal{G}_j^{(t+1)}}$ with $W_{\text{XOR},j,k}$ to obtain

$$\hat{W}_{d_k,\mathcal{G}_j^{(t+1)}\setminus\{k\}} = \hat{W}_{\text{XOR},\mathcal{G}_j^{(t+1)}} \oplus W_{\text{XOR},j,k}. \tag{57}$$

After the last sub-phase $\binom{K}{t+1}$, each receiver $k \in \mathcal{K}$ has decoded all its missing messages in (55), and can thus produce a final guess of message W_{d_k}.

Error Analysis: If each XOR-message $W_{\text{XOR},\mathcal{G}_j^{(t+1)}}$ is decoded correctly by all its intended receivers in $\mathcal{G}_j^{(t+1)}$, $j = 1, \ldots, \binom{K}{t+1}$, then all receivers $1, \ldots, K$ produce the correct estimate of their desired messages $W_{d_1}, \ldots, W_{d_K}$.

The probability that receiver $k \in \mathcal{G}_j^{(t+1)}$ wrongly decodes the XOR message $W_{\text{XOR},\mathcal{G}_j^{(t+1)}}$ tends to 0 as n (and thus n_j) $\to \infty$ because the rate of the subcodebook $\mathcal{C}'_{j,k}$ satisfies

$$\varlimsup_{n\to\infty} \frac{n}{n_j} \cdot R_{\mathcal{G}_j^{(t+1)}\setminus\{k\}} < I(X; Y_k),$$

see (45) and (48).

By letting $\epsilon \to 0$, we conclude the following result.

Theorem 3: Fix a $t \in \{1, \ldots, K-1\}$ and an input distribution P_X, and consider the corresponding cache assignment in (44). Then,

$$\mathsf{C}\left(\mathsf{M}_1^{(t)}, \ldots, \mathsf{M}_K^{(t)}\right) \geq R^{(t)}, \tag{58}$$

where $R^{(t)}$ is calculated from P_X as described in (46).

As we will see in Corollary 8, the Inequality in (58) holds with equality for $t = K-1$.

D. Lower Bound on $\mathsf{C}^\star(\mathsf{M})$

Proposition 1 and Theorems 2 and 3 readily yield a lower bound on $\mathsf{C}^\star(\mathsf{M})$. As we will see in Corollary 11 ahead, this lower bound is exact in the regimes of small and large total cache size M.

Let

$$R^{(0)} := \mathsf{C}_\mathcal{K}, \qquad \mathsf{M}^{(0)} := 0, \tag{59a}$$

and

$$R^{\text{single}} := \mathsf{C}_\mathcal{K} + \frac{\mathsf{M}^{\text{single}}}{D}, \qquad \mathsf{M}^{\text{single}} := \mathsf{M}_1^{\text{single}}, \tag{59b}$$

where $\mathsf{C}_\mathcal{K}$ is defined in (12) and $\mathsf{M}_1^{\text{single}}$ is defined in (25). Also, for given P_X, recall $\mathsf{M}^{(t)}$ and $R^{(t)}$ from (44) and (46), and define for $t \in \{1, \ldots, K-1\}$:

$$\mathsf{M}^{(t)} := \sum_{k=1}^K \mathsf{M}_k^{(t)}. \tag{59c}$$

Proposition 4: For any P_X, all rate-memory pairs in (59) are achievable. By time- and memory-sharing arguments, the upper-convex envelope of all these rate-memory pairs lower bounds $\mathsf{C}^\star(\mathsf{M})$:

$$\mathsf{C}^\star(\mathsf{M}) \geq \text{upp hull} \Big\{ \big(R^{(0)}, \mathsf{M}^{(0)}\big), \ \big(R^{\text{single}}, \mathsf{M}^{\text{single}}\big),$$

$$\bigcup_{P_X} \Big\{ \big(R^{(1)}, \mathsf{M}^{(1)}\big), \dots, \ \big(R^{(K-1)}, \mathsf{M}^{(K-1)}\big) \Big\} \Big\}. \tag{60}$$

Notice that for any P_X:

$$\mathsf{M}^{(0)} \leq \mathsf{M}^{\text{single}} \leq \mathsf{M}^{(1)} \leq \cdots \leq \mathsf{M}^{(K-1)} \tag{61}$$

and

$$R^{(0)} \leq R^{\text{single}} \leq R^{(1)} \leq \cdots \leq R^{(K-1)}. \tag{62}$$

V. UPPER BOUNDS AND EXACT RESULTS ON GLOBAL CAPACITY-MEMORY TRADEOFF

A. Results on $\mathsf{C}(\mathsf{M}_1, \dots, \mathsf{M}_K)$

The upper bound is formulated in terms of the following parameters. For each receiver set $\mathcal{S}$ as in (15), define

$$\alpha^\star_{\mathcal{S},1} := \frac{\mathsf{M}_{j_1}}{D} \tag{63a}$$

and for $k \in \{2, \dots, |\mathcal{S}|\}$:

$$\alpha^\star_{\mathcal{S},k} := \min \left\{ \frac{\sum_{i=1}^k \mathsf{M}_{j_i}}{D - k + 1}, \ \frac{1}{|\mathcal{S}| - k + 1} \left(\frac{|\mathcal{S}|}{D} \sum_{i=1}^{|\mathcal{S}|} \mathsf{M}_{j_i} - \sum_{i=1}^{k-1} \alpha_{\mathcal{S},i} \right) \right\}. \tag{63b}$$

Theorem 5: There exist random variables $X, Y_1, \dots, Y_K$ and for *every* receiver set $\mathcal{S}$ as in (15) random variables $\{U_{\mathcal{S},1}, \dots, U_{\mathcal{S},|\mathcal{S}|-1}\}$ so that the channel law (14b) and the Markov chain

$$U_{\mathcal{S},1} - U_{\mathcal{S},2} - U_{\mathcal{S},|\mathcal{S}|} - \cdots - U_{\mathcal{S},|\mathcal{S}|-1} - X - \big(Y_1, \dots, Y_K\big) \tag{64}$$

hold and so that for each $\mathcal{S}$:

$$\mathsf{C}(\mathsf{M}_1, \dots, \mathsf{M}_K) \leq I\big(U_{\mathcal{S},1}; Y_{j_1}\big) + \alpha^\star_{\mathcal{S},1}, \tag{65a}$$

$$\mathsf{C}(\mathsf{M}_1, \dots, \mathsf{M}_K) \leq I\big(U_{\mathcal{S},k}; Y_{j_k} | U_{\mathcal{S},k-1}\big) + \alpha^\star_{\mathcal{S},k}, \quad \forall k \in \{2, \dots, |\mathcal{S}| - 1\}, \tag{65b}$$

$$\mathsf{C}(\mathsf{M}_1, \dots, \mathsf{M}_K) \leq I\big(X; Y_{j_{|\mathcal{S}|}} | U_{\mathcal{S},|\mathcal{S}|-1}\big) + \alpha^\star_{\mathcal{S},|\mathcal{S}|}. \tag{65c}$$

Proof: See Appendix A. ∎

Without cache memories, $\mathsf{M}_1 = \dots = \mathsf{M}_K = 0$, the parameters $\alpha^\star_{\mathcal{S},1}, \dots, \alpha^\star_{\mathcal{S},|\mathcal{S}|}$ equal 0 for all $\mathcal{S} \subseteq \{1, \dots, K\}$, and the upper bound in Theorem 5 recovers the exact capacity-memory tradeoff $\mathsf{C}_{\mathcal{K}}$ in (13).

The upper bound in Theorem 5 is asymmetric in the different cache sizes $\mathsf{M}_1, \mathsf{M}_2, \dots, \mathsf{M}_K$, because the parameters $\alpha^\star_{\mathcal{S},j_i}$ are not symmetric. In fact, increasing the cache memories at weaker receivers generally increases the upper bound more than increasing the cache memories at stronger receivers.

The converse in Theorem 5 is weakened if constraints (65) are ignored for certain receiver sets $\mathcal{S}$, or if in these constraints the input/output random variables $X, Y_{j_1}, \ldots, Y_{j_{|\mathcal{S}|}}$ are allowed to depend on the receiver set $\mathcal{S}$. For this latter relaxation, Theorem 5 results in the following corollary.

Corollary 6: Given cache sizes $\mathsf{M}_1, \ldots, \mathsf{M}_K \geq 0$, rate R is achievable only if for every receiver set $\mathcal{S} \subseteq \mathcal{K}$:

$$\left(R - \alpha^\star_{\mathcal{S},1}, R - \alpha^\star_{\mathcal{S},2}, \ldots, R - \alpha^\star_{\mathcal{S},|\mathcal{S}|} \right) \in \mathbf{C}_{\mathcal{S}}, \tag{66}$$

where $\mathbf{C}_{\mathcal{S}}$ denotes the capacity region to receivers in $\mathcal{S}$ (ignoring receivers in $\mathcal{K}\backslash\mathcal{S}$) when there are no cache memories.

Remark 3: The upper bounds of Theorem 5 and Corollary 6 are relaxed when each $\alpha^\star_{\mathcal{S},k}$ is replaced by $\tilde{\alpha}_{\mathcal{S},k}$, where

$$\tilde{\alpha}_{\mathcal{S},1} := \frac{\mathsf{M}_{j_1}}{D}, \tag{67a}$$

$$\tilde{\alpha}_{\mathcal{S},k} := \frac{|\mathcal{S}| \cdot \sum_{i=1}^{|\mathcal{S}|} \mathsf{M}_{j_i} - \mathsf{M}_{j_1}}{(|\mathcal{S}| - 1)D}, \quad k \in \{2 \ldots, |\mathcal{S}|\}. \tag{67b}$$

The same holds if each $\alpha^\star_{\mathcal{S},k}$ is replaced by

$$\alpha'_{\mathcal{S},k} := \frac{\sum_{i=1}^{|\mathcal{S}|} \mathsf{M}_{j_i}}{D}. \tag{68}$$

Replacing in Corollary 6 each parameter $\alpha^\star_{\mathcal{S},k}$ by $\alpha'_{\mathcal{S},k}$ recovers the previous upper bound in [26, Theorem 9] and [28, Theorem 1].

Proof: The proof requires a close inspection of the proof of Theorem 5 in Appendix A. See Appendix D. ∎

By comparing the new upper bounds with the three achievability results in the previous Section IV, the exact expression for $\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K)$ can be obtained in some special cases.

The following corollary states that superposition piggyback coding is optimal when only receiver 1 has a cache memory and this cache memory is small.

Corollary 7: Under a cache assignment satisfying

$$0 \leq \mathsf{M}_1 \leq \mathsf{M}_1^{\mathsf{single}} \quad \text{and} \quad \mathsf{M}_2 = \ldots = \mathsf{M}_K = 0, \tag{69}$$

the capacity-memory tradeoff is

$$\mathsf{C}(\mathsf{M}_1, 0, \ldots, 0) = \mathsf{C}_{\mathcal{K}} + \frac{\mathsf{M}_1}{D}.$$

Proof: Achievability follows by Theorem 2. The converse from Corollary 6, where it suffices to consider only the set $\mathcal{S} = \mathcal{K}$. In fact, under (69), $\alpha_{\mathcal{K},1} = \ldots = \alpha_{\mathcal{K},K} = \frac{\mathsf{M}_1}{D}$. ∎

The next corollary states that generalized coded caching with parameter $t = K - 1$ is optimal under the corresponding cache assignment. Moreover, any extra cache memory that is uniformly distributed over the K receivers only brings local caching gain.

Proposition 8: For each $k \in \mathcal{K}$, let $\mathsf{M}_k^{\star(K-1)}$ be given by (44) when P_X is chosen as a maximizer of

$$\mathsf{C}_{\mathsf{Avg}} := \frac{1}{K} \cdot \max_{P_X} \left(\sum_{k=1}^{K} I(X; Y_k) \right). \tag{70}$$

For any $\Delta \geq 0$:

$$\mathsf{C}\left(\mathsf{M}_1^{\star(K-1)} + \Delta, \ldots, \mathsf{M}_K^{\star(K-1)} + \Delta\right)$$

$$= \mathsf{C}_{\mathsf{Avg}} + \frac{\sum_{k=1}^K \mathsf{M}_k^{\star(K-1)}}{K \cdot D} + \frac{\Delta}{D}. \tag{71}$$

Proof: See Appendix E. ∎

B. Results on $\mathsf{C}^\star(\mathsf{M})$

Theorem 5 directly yields the following result.

Proposition 9: There exist random variables $X, Y_1, \ldots, Y_K$ and for *every* receiver set $\mathcal{S}$ as in (15) random variables $\{U_{\mathcal{S},1}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1}\}$, such that (14b) and (64) hold, and such that *for some* $\mathsf{M}_1, \ldots, \mathsf{M}_K \geq 0$ summing to M and all $\mathcal{S}$:

$$\mathsf{C}^\star(\mathsf{M}) \leq I\left(U_{\mathcal{S},1}; Y_{j_1}\right) + \alpha_{\mathcal{S},1}^\star, \tag{72a}$$

$$\mathsf{C}^\star(\mathsf{M}) \leq I\left(U_{\mathcal{S},k}; Y_{j_k} | U_{\mathcal{S},k-1}\right) + \alpha_{\mathcal{S},k}^\star, \quad k \in \{2, \ldots, |\mathcal{S}| - 1\}, \tag{72b}$$

$$\mathsf{C}^\star(\mathsf{M}) \leq I\left(X; Y_{j_{|\mathcal{S}|}} | U_{\mathcal{S},|\mathcal{S}|-1}\right) + \alpha_{\mathcal{S},|\mathcal{S}|}^\star, \tag{72c}$$

where $\{\alpha_{\mathcal{S},k}^\star\}$ are defined in (63).

Solving this optimization problem numerically is computationally complex. Simpler, albeit looser, upper bounds can be obtained by either ignoring some of the constraints (72); by replacing each parameter $\alpha_{\mathcal{S},k}^\star$ in (72) by $\tilde{\alpha}_{\mathcal{S},k}$ or by $\alpha_{\mathcal{S},k}'$; or by allowing $X, Y_{j_1}, \ldots, Y_{j_{\mathcal{S}}}$ in (72) to depend on the set $\mathcal{S}$.

The following corollary presents a simpler bound that is obtained this way. Recall the definitions in (43).

Corollary 10: For each $t \in \mathcal{K}$:

$$\mathsf{C}^\star(\mathsf{M}) \leq \frac{1}{\binom{K}{t}} \sum_{\ell=1}^{\binom{K}{t}} \mathsf{C}_{\mathcal{G}_\ell^{(t)}} + \frac{t}{K} \cdot \frac{\mathsf{M}}{D}. \tag{73}$$

Proof: Fix $t \in \mathcal{K}$. For each $\ell = 1, \ldots \binom{K}{t}$, specialize Corollary 6 to $\mathcal{S} = \mathcal{G}_\ell^{(t)}$ and relax it by replacing each parameter $\alpha_{\mathcal{G}_\ell^{(t)},k}^\star$ by $\alpha_{\mathcal{G}_\ell^{(t)},k}'$. Since $\alpha_{\mathcal{G}_\ell^{(t)},1}' = \ldots = \alpha_{\mathcal{G}_\ell^{(t)},t}'$, we obtain

$$\mathsf{C}^\star(\mathsf{M}) \leq \mathsf{C}_{\mathcal{G}_\ell^{(t)}} + \alpha_{\mathcal{G}_\ell^{(t)},1}' = \mathsf{C}_{\mathcal{G}_\ell^{(t)}} + \frac{\sum_{i \in \mathcal{G}_\ell^{(t)}} \mathsf{M}_i}{D}. \tag{74}$$

Now, averaging bound (74) over all indices $\ell = 1, \ldots, \binom{K}{t}$ and upperbounding the sum $\mathsf{M}_1 + \ldots + \mathsf{M}_K$ by M yields the desired result in the corollary. ∎

The last result of this section contains two more simple upper bounds on $\mathsf{C}^\star(\mathsf{M})$. For small total cache size M one of them is achieved by assigning the entire cache memory to the weakest receiver and applying superposition piggyback coding. For large total cache size M the other is achieved by generalized coded caching with parameter $t = K - 1$, and by first applying the cache assignment corresponding to this scheme followed by a uniform cache assignment of any remaining cache memory.

Corollary 11: For total cache size $\mathsf{M} \geq 0$:

$$\mathsf{C}^\star(\mathsf{M}) \leq \mathsf{C}_\mathcal{K} + \frac{\mathsf{M}}{D} \tag{75}$$

and

$$\mathsf{C}^\star(\mathsf{M}) \leq \mathsf{C}_{\mathsf{avg}} + \frac{1}{K} \cdot \frac{\mathsf{M}}{D}. \tag{76}$$

For small cache sizes,

$$0 \leq \mathsf{M} \leq \mathsf{M}^{\mathsf{single}}, \tag{77}$$

(75) holds with equality.

For large cache sizes,

$$\mathsf{M} \geq D \cdot (K - 1) \cdot K \cdot \mathsf{C}_{\mathsf{avg}}, \tag{78}$$

(76) holds with equality.

Proof: Upper bound (75) follows by specializing Corollary 10 to $t = K$. Upper bound (76) is proved as follows. Relax Theorem 9 by replacing each parameter $\alpha_{\mathcal{S},k}^\star$ by $\alpha_{\mathcal{S},k}'$ and considering only the constraints (72) that correspond to sets $\mathcal{S} = \{k\}$, for $k \in \mathcal{K}$. Finally, average the K resulting inequalities and maximize over the input distribution P_X.

The tightness of (75) for $\mathsf{M} \leq \mathsf{M}^{\mathsf{single}}$ follows from Theorem 2. The tightness of (78) for $\mathsf{M} \geq D(K-1)K\mathsf{C}_{\mathsf{Avg}}$ follows from Proposition 8 because

$$\mathsf{M}_1^{\star(K-1)} + \ldots + \mathsf{M}_K^{\star(K-1)} = D(K-1)K\mathsf{C}_{\mathsf{Avg}}.$$

$\blacksquare$

We remark that for small total cache sizes, $\mathsf{C}^\star(\mathsf{M})$ grows as $\frac{\mathsf{M}}{D}$. This corresponds to a *perfect global caching gain*, i.e., the same performance as in a system where each receiver can directly access all cache contents in the network. For large total cache sizes, $\mathsf{C}^\star(\mathsf{M})$ grows only as $\frac{1}{K} \cdot \frac{\mathsf{M}}{D}$. This corresponds to the local caching gain achieved by Proposition 1.

VI. EXAMPLES

A. Erasure BCs

We specialize our results to erasure BCs where at time t receiver k's output $Y_{k,t}$ equals the channel input X_t with probability $1 - \delta_k$ and it equals an erasure symbol "?" with probability δ_k. The erasure probabilities satisfy:

$$1 > \delta_1 \geq \delta_2 \geq \ldots \geq \delta_K \geq 0. \tag{79}$$

For erasure BCs,

$$\mathsf{C}_\mathcal{S} = \left(\sum_{s \in \mathcal{S}} \frac{1}{1 - \delta_s} \right)^{-1}, \qquad \mathcal{S} \subseteq \mathcal{K}. \tag{80}$$

Moreover, a Bernoulli-1/2 input distribution P_X maximizes $I(X; Y_k)$ and $I(X; Y_k|U)$ simultaneously for all $k \in \mathcal{K}$ and auxiliaries U that form the Markov chain $U - X - Y_k$. Therefore, Theorem 5 and Corollary 6 coincide. Also,

$$\mathsf{C}_{\mathrm{avg}} = \frac{1}{K} \sum_{k=1}^{K} \mathsf{C}_k = 1 - \frac{\sum_{k=1}^{K} \delta_k}{K}. \tag{81}$$

Figure 4, depicts the upper and lower bounds on $\mathsf{C}^\star(\mathsf{M})$ in Propositions 4 and 9. For comparison, also the upper bound in Theorem 5 under a uniform cache assignment

$$\mathsf{M}_1 = \ldots = \mathsf{M}_K = \frac{\mathsf{M}}{K}$$

is plotted. This proves numerically that a smart allocation of the total cache memory M significantly increases the global capacity-memory tradeoff of erasure BCs when different receivers have different erasure probabilities.
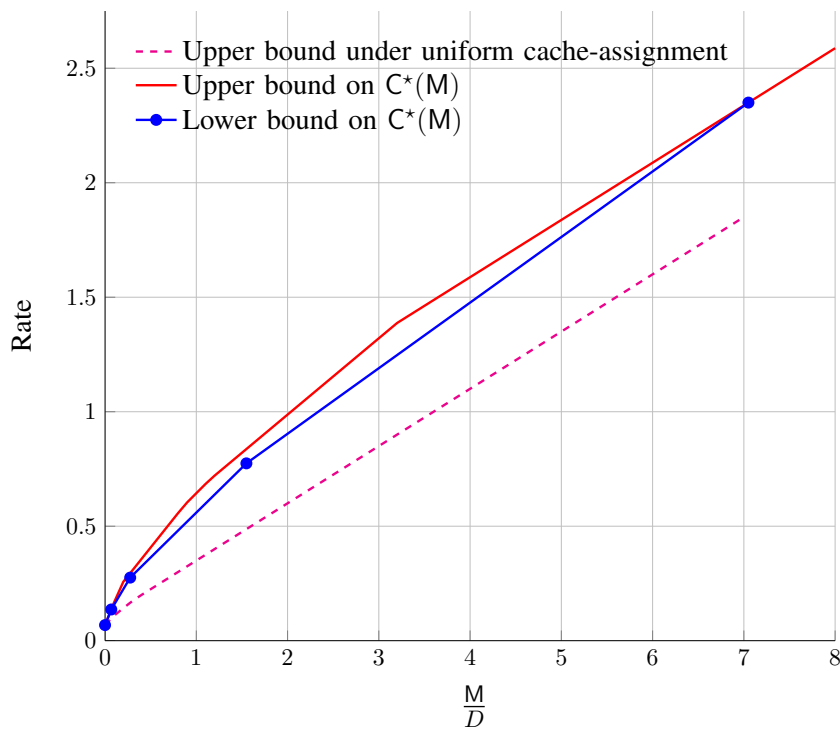

Fig. 4. Bounds on $\mathsf{C}^\star(\mathsf{M})$ for a 4-user Erasure BC with $\delta_1 = 0.9$, $\delta_2 = 0.6$, $\delta_3 = 0.1$, $\delta_4 = 0.051$.

Analytically, we can prove that for small total cache size $\mathsf{M} \leq \mathsf{M}^{\mathrm{single}}$ any cache assignment that does not allocate all cache memory to the weakest receiver is suboptimal on the erasure BC. This follows from the achievability in Corollary 11 and the following Proposition 12.

Proposition 12: For given $\mathsf{M}_1 \geq 0$ and $\mathsf{M} := \sum_{k=1}^{K} \mathsf{M}_k \geq 0$,

$$\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K)$$
$$\leq \min \left\{ \mathsf{C}_\mathcal{K} + \frac{\mathsf{M}_1}{D} + \frac{(\mathsf{M} - \mathsf{M}_1)}{D} \cdot \frac{K \cdot \mathsf{C}_\mathcal{K}}{(K-1)\mathsf{C}_{\{2,...,K\}}}, \right.$$

$$\left. \mathsf{C}_1 + \frac{\mathsf{M}_1}{D} \right\} \tag{82}$$

The RHS of (82) is strictly less than $\mathsf{C}_\mathcal{K} + \frac{\mathsf{M}}{D}$ unless $\mathsf{M} = \mathsf{M}_1$ or $\delta_1 = \ldots = \delta_K$.

Proof: See Appendix F. ∎

B. Noise-Free Bit-Pipe

Consider now the noise-free bit-pipe model with uniform cache assignment in [1]. It corresponds to an erasure BC where each receiver has zero erasure probability,

$$\delta_1 = \ldots = \delta_K = 0. \tag{83}$$

We adopt the "source-coding perspective" of [1], and assume equal cache size

$$m_1 = \cdots = m_K = m.$$

From the upper bound on $\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K)$ in Theorem 5, the following *lower bound* on the minimum achievable delivery rate $\rho^\star$ can be obtained as a function of the normalized symmetric cache size m:

Corollary 13: For the noise-free bit-pipe model in [1]:

$$\rho^\star \geq t - m \cdot \min\left\{\frac{t^2}{D}, \sum_{k=1}^{t} \frac{k}{D-k+1}\right\}, \quad m \leq D. \tag{84}$$

Proof: See Appendix G. ∎

Figure 5 compares this new converse result on ρ with the existing converse results in [1], [9], and [10], and with the achievability result in [43]. The converse result in [10] is generally cumbersome to evaluate. The plot shows the numerical value calculated in [10].

C. Gaussian BCs

Finally, we specialize our results to memoryless Gaussian BCs. At time t, the received symbol at receiver k is

$$Y_{k,t} = X_t + Z_{k,t}, \tag{85}$$

where X_t is the input to the channel and $\{Z_{k,t}\}$ is an i.i.d. Gaussian process with zero mean and variance $\sigma_k^2 > 0$. The channel inputs are subject to an average block-power constraint P. The receivers are ordered in increasing strength:

$$\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_K^2 > 0.$$

By [60], for every set $\mathcal{S}$ as defined in (15),

$$\mathsf{C}_\mathcal{S} = \frac{1}{2} \log_2\left(1 + \frac{\beta_1 P}{\sum_{k=2}^{|\mathcal{S}|} \beta_k P + \sigma_1^2}\right), \tag{86}$$

where $\beta_1, \ldots, \beta_{|\mathcal{S}|}$ form the unique choice of $|\mathcal{S}|$ real numbers in $[0,1]$ that sum to 1 and satisfy

$$\frac{\beta_1 P}{\sum_{k=2}^{|\mathcal{S}|} \beta_k P + \sigma_1^2} = \frac{\beta_i P}{\sum_{k=i+1}^{|\mathcal{S}|} \beta_k P + \sigma_{j_i}^2}, \quad i \in \{1, \ldots, |S|\}. \tag{87}$$
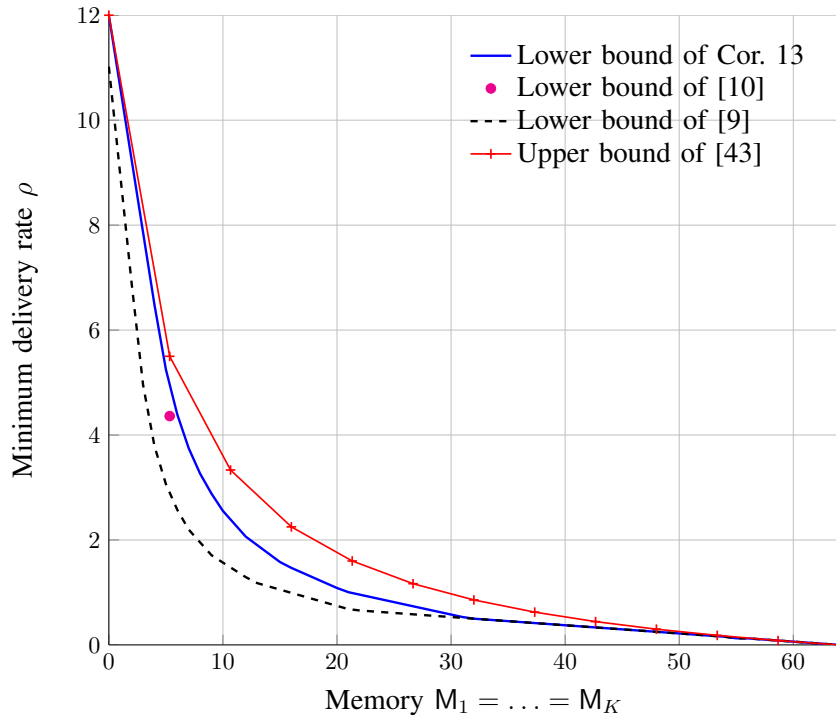
Fig. 5. Upper and lower bounds on the minimum delivery rate ρ in the source coding model of [1] for $K = 12$ and $D = 64$.

In particular,

$$\mathsf{C}_k = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_k^2} \right), \quad k \in \{1, \ldots, K\}. \tag{88}$$

Moreover, given a power constraint $P > 0$, a zero-mean variance-P Gaussian input distribution P_X maximizes $I(X; Y_k)$ and $I(X; Y_k|U)$ simultaneously for all $k \in \mathcal{K}$ and auxiliaries U that form the Markov chain $U - X - Y_k$. Therefore, Theorem 5 and Corollary 6 coincide. Also,

$$\mathsf{C}_{\text{avg}} = \frac{1}{K} \sum_{k=1}^{K} \mathsf{C}_k = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{2} \log \left(1 + \frac{P}{\sigma_k^2} \right). \tag{89}$$

Figure 6 shows the upper and lower bounds on $\mathsf{C}^\star(\mathsf{M})$ in Propositions 4 and 9. The five blue points indicate the rate-memory points $(R^{(0)}, \mathsf{M}^{(0)})$, $(R^{\text{single}}, \mathsf{M}^{\text{single}})$, $(R^{(1)}, \mathsf{M}^{(1)})$, $(R^{(2)}, \mathsf{M}^{(2)})$, and $(R^{(3)}, \mathsf{M}^{(3)})$ for a zero-mean variance-P Gaussian distribution P_X. For comparison, the figure also shows the upper bound in Theorem 5 for a setup with uniform cache assignment $\frac{\mathsf{M}}{K}$ across all receivers. We observe that a smart cache assignment provides substantial gains in the capacity-memory tradeoff.

VII. SUMMARY AND CONCLUSION

We have provided close upper and lower bounds on the global capacity-memory tradeoff $\mathsf{C}^\star(\mathsf{M})$ of degraded BCs. The bounds coincide in the regimes of small and large total cache memory with thresholds depending on the BC statistics. For small cache memory sizes, the weakest receiver needs to be assigned all. In this regime,
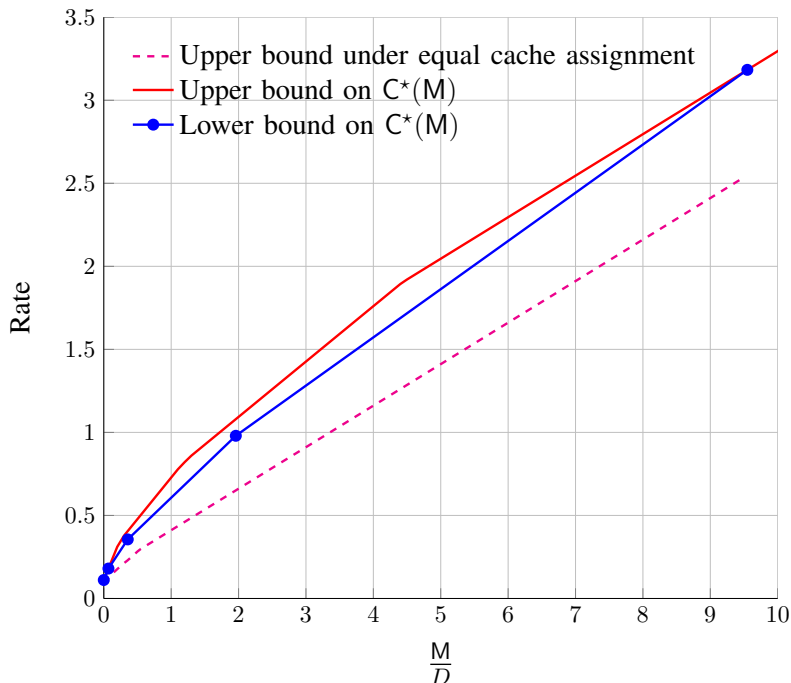
Fig. 6. Upper and lower bounds on $\mathsf{C}^\star(\mathsf{M})$ on a 4-receiver Gaussian BC with input power $P = 1$ and noise variances $\sigma_1 = 4$, $\sigma_2^2 = 1$, $\sigma_3^2 = 0.5$, and $\sigma_4^2 = 0.1$. We have $D = 10$.

$\mathsf{C}^\star(\mathsf{M})$ grows as $\frac{\mathsf{M}}{D}$, which corresponds to a perfect global caching gain where all receivers can benefit from all the cache contents of the network. This performance is achieved by the proposed superposition piggyback coding scheme, which provides each receiver virtual access to the weakest receiver's cache contents. For the regime of moderate M, we propose a generalized coded caching scheme, which assigns cache memories to all the receivers, with a larger cache memory the weaker a receiver is. Notice that the larger the total cache budget M, the larger the coded caching parameter $t \in \{1, \ldots, K-1\}$ needs to be chosen. This leads to a decreasing global caching gain because with increasing t the various cache memories have more and more overlapping contents which cannot provide global caching gains. As a consequence, the slope of the rate-memory tradeoff achieved by generalized coded caching decreases with increasing total cache budget M. The same behaviour is also suggested by the upper bound. For parameter $t = K-1$ generalized coded caching and the corresponding cache assignment exactly achieve the global capacity-memory tradeoff. Once the total cache memory budget exceeds the corresponding cache budget, it is optimal to uniformly allocate all the remaining cache memory across all the receivers and to store the same content in the extra portions of the receivers' cache memories. Here, $\mathsf{C}^\star(\mathsf{M})$ grows as $\frac{1}{K} \cdot \frac{\mathsf{M}}{D}$, which corresponds to a local caching gain. We conclude that assigning the total cache memory uniformly across all the receivers is highly suboptimal over noisy BCs, in contrast to the noiseless setup considered in [1].

APPENDIX A

PROOF OF UPPER BOUND IN THEOREM 5

Fix the rate of communication

$$R < \mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K).$$

Since R is achievable, for each sufficiently large blocklength n and for each demand vector $\mathbf{d}$, there exist K caching functions $\{g_k^{(n)}\}$, an encoding function $\{f_{\mathbf{d}}^{(n)}\}$, and K decoding functions $\{\varphi_{k,\mathbf{d}}^{(n)}\}$ so that the probability of worst-case error $\mathsf{P}_e^{(n)}(\mathbf{d})$ tends to 0 as $n \to \infty$.

Fix $\epsilon > 0$ and a sufficiently large blocklength n (depending on this ϵ). Let

$$\mathbb{V}_k = g_k^{(n)}(W_1, \ldots, W_D), \qquad k \in \{1, \ldots, K\} \tag{90}$$

denote the cache contents corresponding to the chosen caching function, and let for each demand vector $\mathbf{d} = (d_1, \ldots, d_K)$ with all different entries

$$X_{\mathbf{d}}^n = f_{\mathbf{d}}^{(n)}(W_1, \ldots, W_D) \tag{91}$$

denote the input of the degraded BC corresponding to the chosen encoding functions. Let $Y_{k,\mathbf{d}}^n$ denote the corresponding channel outputs at receiver k.

Lemma 14: There exist random variables $X_{\mathbf{d}}, Y_{1,\mathbf{d}}, \ldots, Y_{K,\mathbf{d}}$ and for each set $\mathcal{S}$ as in (15) random variables $\{U_{\mathcal{S},1,\mathbf{d}}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d}}\}$, so that given $X_{\mathbf{d}} = x \in \mathcal{X}$:

$$(Y_{1,\mathbf{d}}, Y_{2,\mathbf{d}}, \ldots, Y_{K,\mathbf{d}}) \sim \Gamma(\cdots | x); \tag{92a}$$

and for each $\mathcal{S}$:

$$U_{\mathcal{S},1,\mathbf{d}} - \cdots - U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d}} - X_{\mathbf{d}} - Y_{K,\mathbf{d}} - Y_{K-1,\mathbf{d}} \cdots - Y_{1,\mathbf{d}} \tag{92b}$$

forms a Markov chain and the following $|\mathcal{S}|$ inequalities hold:

$$R - \epsilon \leq \frac{1}{n} I\big(W_{d_{j_1}}; \mathbb{V}_{j_1}\big) + I\big(U_{\mathcal{S},1,\mathbf{d}}; Y_{j_1,\mathbf{d}}\big), \tag{93a}$$

$$R - \epsilon \leq \frac{1}{n} I\big(W_{d_{j_k}}; \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_k} | W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}\big)$$

$$+ I\big(U_{\mathcal{S},k,\mathbf{d}}; Y_{j_k,\mathbf{d}} | U_{\mathcal{S},k-1,\mathbf{d}}\big),$$

$$\forall k \in \{2, \ldots, |\mathcal{S}| - 1\}, \tag{93b}$$

$$R - \epsilon \leq \frac{1}{n} I\big(W_{d_{j_{|\mathcal{S}|}}}; \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_{|\mathcal{S}|}} | W_{d_{j_1}}, \ldots, W_{d_{j_{|\mathcal{S}|-1}}}\big)$$

$$+ I\big(X_{\mathbf{d}}; Y_{j_{|\mathcal{S}|},\mathbf{d}} | U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d}}\big). \tag{93c}$$

Proof: The proof is similar to the converse proof of the capacity of degraded BCs without caching [59].

Since the worst case error probability is bounded by ϵ, using Fano's inequality we have

$$R - \epsilon \leq \frac{1}{n} I\big(W_{d_{j_1}}; Y_{j_1,\mathbf{d}}^n, \mathbb{V}_{j_1}\big)$$

$$= \frac{1}{n} I\big(W_{d_{j_1}}; \mathbb{V}_{j_1}\big) + \frac{1}{n} I\big(W_{d_{j_1}}; Y_{j_1,\mathbf{d}}^n | \mathbb{V}_{j_1}\big). \tag{94a}$$

Similarly, for $k \in \{2, \ldots, K\}$:

$$R - \epsilon \overset{(a)}{\leq} \frac{1}{n} I\big(W_{d_{j_k}}; Y_{j_k, \mathbf{d}}^n, \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_k} \big| W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}\big)$$

$$= \frac{1}{n} I\big(W_{d_{j_k}}; \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_k} \big| W_{d_1}, \ldots, W_{d_{j_{k-1}}}\big)$$

$$+ \frac{1}{n} I\big(W_{d_{j_k}}; Y_{j_k, \mathbf{d}}^n \big| \mathbb{V}_1, \ldots, \mathbb{V}_{j_k}, W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}\big), \tag{94b}$$

where (a) uses Fano's inequality as well as the fact that all messages are independent. Recall that the demand vector $\mathbf{d}$ has all different entries.

We next develop the second summands in (94a) and (94b). For the second summand in (94a) we write

$$\frac{1}{n} I\big(W_{d_{j_1}}; Y_{j_1, \mathbf{d}}^n \big| \mathbb{V}_{j_1}\big) = \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_{j_1}}; Y_{j_1, \mathbf{d}, t} \big| \mathbb{V}_{j_1}, Y_{j_1, \mathbf{d}}^{t-1}\big)$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_{j_1}}, Y_{j_1, \mathbf{d}}^{t-1}; Y_{j_1, \mathbf{d}, t} \big| \mathbb{V}_{j_1}\big)$$

$$= I\big(U_{\mathcal{S}, 1, \mathbf{d}, T}; Y_{j_1, \mathbf{d}, T} \big| \mathbb{V}_{j_1}, T\big)$$

$$\leq I\big(U_{\mathcal{S}, 1, \mathbf{d}}; Y_{j_1, \mathbf{d}} \big| \mathbb{V}_{j_1}\big) \tag{95}$$

where T denotes a random variable that is uniformly distributed over $\{1, \ldots, n\}$ and independent of all previously defined random variables, and where

$$U_{\mathcal{S}, 1, \mathbf{d}, T} := \big(\mathbb{V}_{j_1}, W_{d_{j_1}}, Y_{j_1, \mathbf{d}}^{t-1}\big),$$

$$U_{\mathcal{S}, 1, \mathbf{d}} := (U_{\mathcal{S}, 1, \mathbf{d}, T}, T),$$

$$Y_{j_1, \mathbf{d}} := Y_{j_1, \mathbf{d}, T}.$$

Define further for $k \in \{2, \ldots, |\mathcal{S}| - 1\}$:

$$U_{\mathcal{S}, k, \mathbf{d}, T} := \big(U_{\mathcal{S}, k-1, \mathbf{d}, T}, \mathbb{V}_{j_k}, W_{d_{j_k}}, Y_{j_k, \mathbf{d}}^{t-1}\big),$$

$$U_{\mathcal{S}, k, \mathbf{d}} := U_{\mathcal{S}, k, \mathbf{d}, T},$$

$$Y_{j_k, \mathbf{d}} := Y_{j_k, \mathbf{d}, T},$$

and

$$Y_{j_{|\mathcal{S}|}, \mathbf{d}} := Y_{j_{|\mathcal{S}|}, \mathbf{d}, T}$$

$$X_{\mathbf{d}} := X_{\mathbf{d}, T}.$$

For $k \in \{2, \ldots, K-1\}$, we expand the second summand in (94b) as:

$$\frac{1}{n} I\big(W_{d_{j_k}}; Y_{j_k, \mathbf{d}}^n \big| \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_k}, W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}\big)$$

$$= \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_{j_k}}; Y_{j_k, \mathbf{d}, t} \big| \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_k}, W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}, Y_{j_k, \mathbf{d}}^{t-1}\big)$$

$$\overset{(a)}{=} \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_{j_k}}; Y_{j_k,\mathbf{d},t} \big| \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_k}, W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}},$$

$$Y_{j_1,\mathbf{d}}^{t-1}, \ldots, Y_{j_{k-1},\mathbf{d}}^{t-1}, Y_{j_k,\mathbf{d}}^{t-1}\big)$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_{j_k}}, Y_{j_k,\mathbf{d}}^{t-1}, \mathbb{V}_{j_k}; Y_{j_k,\mathbf{d},t} \big| \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_{k-1}},$$

$$W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}, Y_{j_1,\mathbf{d}}^{t-1}, \ldots, Y_{j_{k-1},\mathbf{d}}^{t-1}\big)$$

$$= I\big(U_{\mathcal{S},k,\mathbf{d},T}; Y_{j_k,\mathbf{d},T} \big| U_{\mathcal{S},k-1,\mathbf{d},T}, T\big)$$

$$= I\big(U_{\mathcal{S},k,\mathbf{d}}; Y_{j_k,\mathbf{d}} \big| U_{\mathcal{S},k-1,\mathbf{d}}\big), \tag{96}$$

where (a) follows from the degradedness of the outputs.

Similarly, we also have

$$\frac{1}{n} I\big(W_{d_{j_{|\mathcal{S}|}}}; Y_{j_{|\mathcal{S}|},\mathbf{d}}^{n} \big| \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_{|\mathcal{S}|}}, W_{d_{j_1}}, \ldots, W_{d_{j_{|\mathcal{S}|-1}}}\big)$$

$$= \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_{j_{|\mathcal{S}|}}}; Y_{j_{|\mathcal{S}|},\mathbf{d},t} \big| \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_{|\mathcal{S}|}}, W_{d_{j_1}}, \ldots, W_{d_{j_{|\mathcal{S}|-1}}},$$

$$Y_{j_1,\mathbf{d}}^{t-1}, \ldots, Y_{j_{|\mathcal{S}|-1},\mathbf{d}}^{t-1}, Y_{j_{|\mathcal{S}|},\mathbf{d}}^{t-1}\big)$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_{j_{|\mathcal{S}|}}}, Y_{j_{|\mathcal{S}|},\mathbf{d}}^{t-1}, \mathbb{V}_{j_{|\mathcal{S}|}}; Y_{j_{|\mathcal{S}|},\mathbf{d},t} \big| \mathbb{V}_{j_1}, \ldots, \mathbb{V}_{j_{|\mathcal{S}|-1}},$$

$$W_{d_{j_1}}, \ldots, W_{d_{j_{|\mathcal{S}|-1}}}, Y_{j_1,\mathbf{d}}^{t-1}, \ldots, Y_{j_{|\mathcal{S}|-1},\mathbf{d}}^{t-1}\big)$$

$$\leq I\big(X_{\mathbf{d},T}; Y_{j_{|\mathcal{S}|},\mathbf{d},T} \big| U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d},T}, T\big)$$

$$= I\big(X_{\mathbf{d}}; Y_{j_{|\mathcal{S}|},\mathbf{d}} \big| U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d}}\big). \tag{97}$$

It can be verified that the defined random variables satisfy Conditions (92). Combining this observation with (94)–(97) concludes the proof. ∎

We average the bounds in (93) over demand vectors. Let $\mathcal{Q}_K^{\text{dist}}$ be the set of all the $\binom{D}{K} K!$ K-dimensional demand vectors with all distinct entries. Also, let Q be a uniform random variable over the elements of $\mathcal{Q}_K^{\text{dist}}$ and independent of all other random variables. Define for each set $\mathcal{S}$ as in (15): $U_{\mathcal{S},1} := (U_{\mathcal{S},1,Q}, Q)$; $U_{\mathcal{S},k} := U_{\mathcal{S},k,Q}$, for $k \in \{2, \ldots, |\mathcal{S}|-1\}$; $X := X_Q$; and $Y_k := Y_{k,Q}$ for $k \in \mathcal{K}$.

Notice that the defined random variables satisfy conditions (14b) and (64) in the theorem. It remains to prove that they also satisfy (65). To this end, we average inequalities (93) over all the demand vectors in $\mathcal{Q}_K^{\text{dist}}$. Using standard arguments to take care of the time-sharing random variable Q, and defining

$$\alpha_{\mathcal{S},1} := \frac{1}{\binom{D}{K} K!} \sum_{\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}} \frac{1}{n} I(W_{d_{j_1}}; \mathbb{V}_1), \tag{98a}$$

$$\alpha_{\mathcal{S},k} := \frac{1}{\binom{D}{K} K!} \sum_{\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}} \frac{1}{n} I(W_{d_{j_k}}; \mathbb{V}_1, \ldots, \mathbb{V}_{j_k} | W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}),$$

$$k \in \{2, \ldots, |\mathcal{S}|\}, \tag{98b}$$

we obtain for each $\mathcal{S}$ as in (15):

$$R - \epsilon \leq I\big(U_{\mathcal{S},1}; Y_{j_1}\big) + \alpha_{\mathcal{S},1}, \tag{99a}$$

$$R - \epsilon \leq I\big(U_{\mathcal{S},k}; Y_{j_k} | U_{\mathcal{S},k-1}\big) + \alpha_{\mathcal{S},k}, \quad \forall k \in \{2, \ldots, |\mathcal{S}| - 1\}, \tag{99b}$$

$$R - \epsilon \leq I\big(X; Y_{j_{|\mathcal{S}|}} | U_{\mathcal{S},|\mathcal{S}|-1}\big) + \alpha_{\mathcal{S},|\mathcal{S}|}, \tag{99c}$$

Lemma 15: For each set $\mathcal{S}$, parameters $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$ satisfy the following constraints:

$$0 \leq \alpha_{\mathcal{S},k} \leq \frac{\sum_{i=1}^{k} \mathsf{M}_{j_i}}{D - k + 1}, \qquad k \in \{1, \ldots, |\mathcal{S}|\}, \tag{100a}$$

$$\alpha_{\mathcal{S},k'} \leq \alpha_{\mathcal{S},k}, \qquad k, k' \in \{1, \ldots, |\mathcal{S}|\}, \ k' \leq k, \tag{100b}$$

$$\sum_{k=1}^{|\mathcal{S}|} \alpha_{\mathcal{S},k} \leq \frac{|\mathcal{S}|}{D} \sum_{k=1}^{|\mathcal{S}|} \mathsf{M}_{j_k}. \tag{100c}$$

Proof: See Appendix B. ∎

By (99)–(100) and letting $\epsilon \to 0$, the following intermediate result—which is used in other proofs in this paper—is obtained.

Lemma 16: There exist random variables $X, Y_1, \ldots, Y_K$ and for *every* receiver set $\mathcal{S}$ as in (15) random variables $\{U_{\mathcal{S},1}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1}\}$, so that (14b) and (64) hold, and for all $\mathcal{S}$:

$$\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) \leq I\big(U_{\mathcal{S},1}; Y_{j_1}\big) + \alpha_{\mathcal{S},1}, \tag{101a}$$

$$\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) \leq I\big(U_{\mathcal{S},k}; Y_{j_k} | U_{\mathcal{S},k-1}\big) + \alpha_{\mathcal{S},k},$$

$$\forall k \in \{2, \ldots, |\mathcal{S}|\}, \tag{101b}$$

for parameters $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$ satisfying (100).

By the following Lemma 17, because constraints (101) are increasing in $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$, and by constraint (100c), we conclude that the choice $\alpha_{\mathcal{S},k} = \alpha_{\mathcal{S},k}^{\star}$ in (63) makes the upper bound (101) loosest. The following Lemma 17 thus concludes the proof.

Lemma 17: Lemma 16 remains valid, if parameters $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$ are further constrained to satisfy for each $k \in \{1, \ldots, |\mathcal{S}| - 1\}$ one of the two following conditions:

- $\alpha_{\mathcal{S},k} = \frac{\sum_{i=1}^{k} \mathsf{M}_{j_i}}{D - k + 1}$; or

- $\alpha_{\mathcal{S},k} = \alpha_{\mathcal{S},k+1}$.

Proof: See Appendix C. ∎

APPENDIX B

PROOF OF LEMMA 15

We only prove the lemma for $\mathcal{S} = \mathcal{K}$. The other proofs are similar.

We first prove (100a). Every $\alpha_{\mathcal{K},k}$ is non-negative, because mutual information is non-negative. To prove the upper bound in (100a), we proceed as follows. Let $\mathcal{Q}_K^{\mathrm{dist}}$ be the set of K-dimensional demand vectors that have K distinct

entries in $\{1,\ldots,D\}$; and for each $k \in \{1,\ldots,K\}$ and each $k-1$ dimensional demand vector $\tilde{\mathbf{d}} = (d_1,\ldots,d_{k-1})$, define $W_{\tilde{\mathbf{d}}} := (W_{d_1},\ldots,W_{d_{k-1}})$. We have:

$$\alpha_{\mathcal{K},k}$$

$$= \frac{1}{K!\binom{D}{K}} \sum_{\mathbf{d} \in \mathcal{Q}_K^{\mathrm{dist}}} I(W_{d_k}; \mathbb{V}_1,\ldots,\mathbb{V}_k | W_{d_1},\ldots,W_{d_{k-1}})$$

$$= \frac{1}{K!\binom{D}{K}} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\mathrm{dist}}} \sum_{\substack{\mathbf{d} \in \mathcal{Q}_K^{\mathrm{dist}}: \\ (d_1,\ldots,d_{k-1})=\tilde{\mathbf{d}}}} I(W_{d_k}; \mathbb{V}_1,\ldots,\mathbb{V}_k | W_{\tilde{\mathbf{d}}})$$

$$\overset{(a)}{=} \frac{1}{K!\binom{D}{K}} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\mathrm{dist}}} \sum_{j \in \mathcal{D} \setminus \tilde{\mathbf{d}}} I(W_j; \mathbb{V}_1,\ldots,\mathbb{V}_k | W_{\tilde{\mathbf{d}}})$$

$$\cdot \binom{D-k}{K-k}(K-k)!$$

$$= \frac{1}{k!\binom{D}{k}} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\mathrm{dist}}} \sum_{j \in \mathcal{D} \setminus \tilde{\mathbf{d}}} I(W_j; \mathbb{V}_1,\ldots,\mathbb{V}_k | W_{\tilde{\mathbf{d}}})$$

$$\overset{(b)}{=} \frac{1}{k!\binom{D}{k}} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\mathrm{dist}}} \Big[H(W_1,\ldots,W_N | W_{\tilde{\mathbf{d}}})$$

$$- \sum_{j \in \mathcal{D} \setminus \tilde{\mathbf{d}}} H(W_j | \mathbb{V}_1,\ldots,\mathbb{V}_k, W_{\tilde{\mathbf{d}}}) \Big]$$

$$\overset{(c)}{\leq} \frac{1}{k!\binom{D}{k}} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\mathrm{dist}}} I(W_1,\ldots,W_N; \mathbb{V}_1,\ldots,\mathbb{V}_k | W_{\tilde{\mathbf{d}}})$$

$$\overset{(d)}{\leq} \frac{(k-1)!\binom{D}{k-1}}{k!\binom{D}{k}} \sum_{i=1}^{k} \mathsf{M}_i$$

$$= \frac{\sum_{i=1}^{k} \mathsf{M}_i}{D-k+1} \tag{102}$$

where (a) holds because for each value of K and j there are $\binom{D-k}{K-k}(K-k)!$ ordered demand vectors $\mathbf{d} \in \mathcal{Q}_K^{\mathrm{dist}}$ with $(d_1,\ldots,d_{k-1}) = \tilde{\mathbf{d}}$ and with $d_k = j$; (b) holds by the independence of the messages; (c) holds because for any random tuple $(A_1,\ldots,A_L)$ it holds that $\sum_{l=1}^{L} H(A_l) \geq H(A_1,\ldots,A_L)$; and (d) holds because $I(W_1,\ldots,W_N; \mathbb{V}_1,\ldots,\mathbb{V}_k | W_{\tilde{\mathbf{d}}})$ cannot exceed $\sum_{i=1}^{k} \mathsf{M}_i$. This concludes the proof of (100a).

To prove constraint (100b), we fix a K-dimensional demand vector $\mathbf{d} \in \mathcal{Q}_K^{\mathrm{dist}}$, and consider the cyclic shifts of this vector. For $\ell \in \{0,\ldots,K-1\}$, let $\mathbf{d}^{(\ell)}$ be the vector obtained from $\mathbf{d}$ when the elements are cyclically shifted ℓ positions to the right. (For example, if $\mathbf{d} = (1,2,3)$ then $\mathbf{d}^{(2)} = (2,3,1)$.) For each $\ell \in \{0,\ldots,K-1\}$ and $k \in \{1,\ldots,K\}$, let $d_k^{(\ell)}$ denote the k-th index of demand vector $\mathbf{d}^{(\ell)}$. So,

$$d_k^{(\ell)} = d_{(k-\ell) \bmod K} \tag{103}$$

where for each positive integer ξ the term $(\xi \bmod K)$ takes value in $\{1,\ldots,K\}$ so that

$$\xi \bmod K = \xi - bK \quad \text{for some positive integer } b. \tag{104}$$

For each $\ell \in \{1, \ldots, K-1\}$ and $k, k' \in \{2, \ldots, K\}$ with $k' \leq k$, we write

$$I(W_{d_1}; \mathbb{V}_1) \overset{(a)}{=} I(W_{d_{k'}^{(k'-1)}}; \mathbb{V}_1)$$

$$\overset{(b)}{\leq} I(W_{d_{k'}^{(k'-1)}}; \mathbb{V}_1, \ldots, \mathbb{V}_{k'} | W_{d_1^{(k'-1)}}, \ldots, W_{d_{k'-1}^{(k'-1)}})$$

$$\overset{(a)}{=} I(W_{d_k^{(k-1)}}; \mathbb{V}_1, \ldots, \mathbb{V}_{k'} | W_{d_{1+k-k'}^{(k-1)}}, \ldots, W_{d_{k-1}^{(k-1)}})$$

$$\overset{(b)}{\leq} I(W_{d_k^{(k-1)}}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1^{(k-1)}}, \ldots, W_{d_{k-1}^{(k-1)}}) \tag{105}$$

where (a) follows by (103) and (b) is by the independence of messages.

Fix a demand vector $\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}$ and sum up the above inequality (105) over all K cyclic shifts $\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots,$ $\mathbf{d}^{(K-1)}$ of $\mathbf{d}$ to obtain:

$$\sum_{\ell=0}^{K-1} I(W_{d_1^{(\ell)}}; \mathbb{V}_1)$$

$$\leq \sum_{\ell=0}^{K-1} I(W_{d_{k'}^{(\ell)}}; \mathbb{V}_1, \ldots, \mathbb{V}_{k'} | W_{d_1^{(\ell)}}, \ldots, W_{d_{k'-1}^{(\ell)}})$$

$$\leq \sum_{\ell=0}^{K-1} I(W_{d_k^{(\ell)}}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1^{(\ell)}}, \ldots, W_{d_{k-1}^{(\ell)}}). \tag{106}$$

Since the set $\mathcal{Q}_K^{\text{dist}}$ can be partitioned into subsets of demand vectors that are cyclic shifts of each others and all cyclic shifts of a demand vector in $\mathcal{Q}_K^{\text{dist}}$ are also in $\mathcal{Q}_K^{\text{dist}}$, we conclude from (106):

$$\sum_{\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}} I(W_{d_1}; \mathbb{V}_1)$$

$$\leq \sum_{\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}} I(W_{d_{k'}}; \mathbb{V}_1, \ldots, \mathbb{V}_{k'} | W_{d_1}, \ldots, W_{d_{k'-1}})$$

$$\leq \sum_{\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, \ldots, W_{d_{k-1}}). \tag{107}$$

This proves (100b).

We proceed to prove constraint (100c). For each $\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}$:

$$I(W_{d_1}; \mathbb{V}_1) + \sum_{k=2}^{K} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, W_{d_2}, \ldots, W_{d_{k-1}})$$

$$\leq I(W_{d_1}, W_{d_2}, \ldots, W_{d_K}; \mathbb{V}_1, \ldots, \mathbb{V}_K). \tag{108}$$

So,

$$\sum_{\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}} \Bigg[I(W_{d_1}; \mathbb{V}_1)$$

$$+ \sum_{k=2}^{K} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, W_{d_2}, \ldots, W_{d_{k-1}}) \Bigg]$$

$$\leq \sum_{\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}} I(W_{d_1}, W_{d_2}, \ldots, W_{d_K}; \mathbb{V}_1 \ldots, \mathbb{V}_K)$$

$$\stackrel{(a)}{=} \sum_{\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}} \Big[H(W_{d_1}) + H(W_{d_2}) + \ldots + H(W_{d_K})$$

$$- H(W_{d_1}, \ldots, W_{d_K} | \mathbb{V}_1, \ldots, \mathbb{V}_K) \Big]$$

$$\stackrel{(b)}{=} \frac{K}{D} |\mathcal{Q}_K^{\text{dist}}| H(W_1, \ldots, W_D)$$

$$- \sum_{\mathbf{d} \in \mathcal{Q}_K^{\text{dist}}} H(W_{d_1}, \ldots, W_{d_K} | \mathbb{V}_1, \ldots, \mathbb{V}_K)$$

$$\stackrel{(c)}{\leq} \frac{K}{D} K! \binom{D}{K} H(W_1, \ldots, W_D)$$

$$- \frac{K}{D} K! \binom{D}{K} H(W_1, \ldots, W_D | \mathbb{V}_1, \ldots, \mathbb{V}_K)$$

$$\stackrel{(b)}{=} \frac{K}{D} K! \binom{D}{K} I(W_1, \ldots, W_D; \mathbb{V}_1, \ldots, \mathbb{V}_K)$$

$$\leq \frac{K}{D} K! \binom{D}{K} n \sum_{k=1}^{K} M_k,$$

where (a) holds by the chain rule of mutual information, (b) by the independence and uniform rate of messages $W_1, \ldots, W_D$ and the definition of the set $\mathcal{Q}_K^{\text{dist}}$, which is of size $\binom{D}{K} K!$, and (c) by the generalized Han-Inequality (the following Proposition 18).

Proposition 18: Let L be a positive integer and $A_1, \ldots, A_L$ be a finite random L-tuple. Denote by $A_{\mathcal{J}}$ the subset $\{A_l, \ l \in \mathcal{J}\}$. For every $i \in \{1, \ldots, L\}$:

$$\frac{1}{\binom{L}{i}} \sum_{\substack{\mathcal{J} \subseteq \{1, \ldots, L\}: \\ |\mathcal{J}| = i}} \frac{H(A_{\mathcal{J}})}{i} \geq \frac{1}{L} H(A_1, \ldots, A_L). \tag{109}$$

Proof: See [62, Theorem 17.6.1]. ∎

APPENDIX C

PROOF OF LEMMA 17

We prove the lemma by contradiction. Fix a random tuple $(X, Y_1, \ldots, Y_K)$ satisfying (14b) and for each set $\mathcal{S}$ as in (15) a random tuple $U_{\mathcal{S},1}, U_{\mathcal{S},2}, \ldots, U_{\mathcal{S},|\mathcal{S}-1}$ satisfying (64) and real numbers $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$ satisfying (100).

Assume that for some set $\mathcal{S}$ as in (15) and some $\tilde{k} \in \{1, \ldots, |\mathcal{S}| - 1\}$:

$$\alpha_{\mathcal{S},\tilde{k}} \neq \alpha_{\mathcal{S},\tilde{k}+1} \tag{110}$$

and

$$\alpha_{\mathcal{S},\tilde{k}} < \frac{\sum_{i=1}^{\tilde{k}} \mathsf{M}_{j_i}}{D - \tilde{k} + 1}. \tag{111}$$

Let

$$\gamma := \max \left\{ \frac{1}{2}, \frac{\alpha_{\mathcal{S},\tilde{k}+1} - \frac{\sum_{i=1}^{\tilde{k}} \mathsf{M}_{j_i}}{D - \tilde{k} + 1}}{\alpha_{\mathcal{S},\tilde{k}+1} - \alpha_{\mathcal{S},\tilde{k}}} \right\}. \tag{112}$$

Notice that by (111):

$$\gamma \in \left[\frac{1}{2}, 1\right). \tag{113}$$

Define the new parameters

$$\bar{\alpha}_{\mathcal{S},k} := \alpha_{\mathcal{S},k}, \quad k \in \{1, \ldots, |\mathcal{S}|\} \backslash \{\tilde{k}, \tilde{k}+1\} \tag{114a}$$

$$\bar{\alpha}_{\mathcal{S},\tilde{k}} := \gamma \alpha_{\mathcal{S},\tilde{k}} + (1-\gamma)\alpha_{\mathcal{S},\tilde{k}+1} \tag{114b}$$

$$\bar{\alpha}_{\mathcal{S},\tilde{k}+1} := (1-\gamma)\alpha_{\mathcal{S},\tilde{k}} + \gamma \alpha_{\mathcal{S},\tilde{k}+1}. \tag{114c}$$

Notice that this new set of parameters satisfies constraints (100) when $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$ are replaced by $\bar{\alpha}_{\mathcal{S},1}, \ldots, \bar{\alpha}_{\mathcal{S},|\mathcal{S}|}$. In particular,

$$\bar{\alpha}_{\mathcal{S},k} \leq \bar{\alpha}_{\mathcal{S},k+1}, \qquad k \in \{1, \ldots, |\mathcal{S}|-1\}. \tag{115}$$

We will show that there exist new auxiliary random variables $\bar{U}_{\mathcal{S},1}, \bar{U}_{\mathcal{S},2}, \ldots, \bar{U}_{\mathcal{S},|\mathcal{S}|-1}$ satisfying the Markov chain (64), and so that upper bound (93) is looser for these new auxiliares and the new parameters $\bar{\alpha}_{\mathcal{S},1}, \ldots, \bar{\alpha}_{\mathcal{S},|\mathcal{S}|}$ than for the original auxiliaries $U_{\mathcal{S},1}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1}$ and parameters $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|-1}$.

To simplify notation in the following, we define

$$U_{\mathcal{S},|\mathcal{S}|} := X. \tag{116}$$

Notice that since $\alpha_{\mathcal{S},\tilde{k}} \neq \alpha_{\mathcal{S},\tilde{k}+1}$ and by (100b), the strict inequality

$$\alpha_{\mathcal{S},\tilde{k}} < \alpha_{\mathcal{S},\tilde{k}+1} \tag{117}$$

must hold. Choose

$$\bar{U}_{\mathcal{S},k} = U_{\mathcal{S},k}, \qquad k \in \{1, \ldots, |\mathcal{S}|-1\} \backslash \{\tilde{k}\}, \tag{118}$$

and

$$\bar{U}_{\mathcal{S},|\mathcal{S}|} = U_{\mathcal{S},|\mathcal{S}|} = X. \tag{119}$$

The choice of $\bar{U}_{\mathcal{S},\tilde{k}}$ depends on whether

$$I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|U_{\mathcal{S},\tilde{k}-1}) \leq I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\mathcal{S},\tilde{k}}), \tag{120a}$$

or

$$I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|U_{\mathcal{S},\tilde{k}-1}) > I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\mathcal{S},\tilde{k}}). \tag{120b}$$

If (120a) holds, choose

$$\bar{U}_{\mathcal{S},\tilde{k}} = U_{\mathcal{S},\tilde{k}}. \tag{121}$$

If (120b) holds, let $E \in \{0,1\}$ be a Bernoulli-β random variable independent of everything else, where

$$\beta := (1-\gamma) - (1-\gamma) \cdot \frac{I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\mathcal{S},\tilde{k}})}{I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|U_{\mathcal{S},\tilde{k}-1})}. \tag{122}$$

Choose

$$\bar{U}_{\mathcal{S},\tilde{k}} = \begin{cases} (U_{\mathcal{S},\tilde{k}}, E), & \text{if } E = 0 \\ (U_{\mathcal{S},\tilde{k}-1}, E), & \text{if } E = 1. \end{cases} \tag{123}$$

Notice that in both cases the proposed choice satisfies the Markov chain $\bar{U}_{\mathcal{S},1} - \bar{U}_{2,\mathcal{S}} - \cdots - \bar{U}_{\mathcal{S},|\mathcal{S}|-1} - X$.

Trivially, for $k \notin \{\tilde{k}, \tilde{k}+1\}$, constraint (93) is unchanged if we replace $(U_{\mathcal{S},1}, U_{\mathcal{S},2}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1}, X)$ by $(\bar{U}_{\mathcal{S},1}, \bar{U}_{\mathcal{S},2}, \ldots, \bar{U}_{\mathcal{S},K-1}, X)$ and $(\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|})$ by $(\bar{\alpha}_{\mathcal{S},1}, \ldots, \bar{\alpha}_{\mathcal{S},|\mathcal{S}|})$.

If (120a) holds, then the proposed replacement relaxes constraint (93) for $k = \tilde{k}$ $\big($because $\bar{\alpha}_{\mathcal{S},\tilde{k}} > \alpha_{\mathcal{S},\tilde{k}}\big)$ and it tightens it for $k = \tilde{k}+1$ $\big($because $\bar{\alpha}_{\mathcal{S},\tilde{k}+1} < \alpha_{\mathcal{S},\tilde{k}+1}\big)$. However, the new constraint for $k = \tilde{k}+1$ is less stringent than the original constraint for $k = \tilde{k}$:

$$\begin{aligned} \bar{\alpha}_{\mathcal{S},\tilde{k}+1} &+ I(\bar{U}_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1}|\bar{U}_{\mathcal{S},\tilde{k}}) \\ &\overset{(a)}{=} (1-\gamma) \cdot \alpha_{\mathcal{S},\tilde{k}} + \gamma \cdot \alpha_{\mathcal{S},\tilde{k}+1} + I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\mathcal{S},\tilde{k}}) \\ &\overset{(b)}{>} \alpha_{\mathcal{S},\tilde{k}} + I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\mathcal{S},\tilde{k}}) \\ &\overset{(c)}{\geq} \alpha_{\mathcal{S},\tilde{k}} + I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|U_{\mathcal{S},\tilde{k}-1}), \end{aligned} \tag{124}$$

where (a) holds by (114c); (b) holds by (117); and (c) holds by holds by assumption (120a). We conclude that when (120a) holds, the upper bound on $\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K)$ in (93) is relaxed if everywhere one replaces $(U_{\mathcal{S},1}, U_{\mathcal{S},2} \ldots, U_{|\mathcal{S}|,|\mathcal{S}|-1})$ and $(\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|})$ by $(\bar{U}_{\mathcal{S},1}, \bar{U}_{\mathcal{S},2}, \ldots, \bar{U}_{\mathcal{S},|\mathcal{S}|-1})$ and $(\bar{\alpha}_{\mathcal{S},1}, \ldots, \bar{\alpha}_{\mathcal{S},|\mathcal{S}|})$.

We now assume that (120b) holds. We show that the new constraints obtained for $k = \tilde{k}$ and for $k = \tilde{k}+1$ cannot be more stringent then the tighter of the two original constraints for $k = \tilde{k}$ and $k = \tilde{k}+1$.

Consider $k = \tilde{k}$. By (122) and (123) we have

$$\begin{aligned} I(\bar{U}_{\mathcal{S},\tilde{k}}; & Y_{\tilde{k}}|\bar{U}_{\mathcal{S},\tilde{k}-1}) \\ &= I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|\bar{U}_{\mathcal{S},\tilde{k}-1}E) \\ &= (1-\beta) \cdot I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|U_{\mathcal{S},\tilde{k}-1}) \\ &= \gamma \cdot I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|U_{\mathcal{S},\tilde{k}-1}) \\ &\quad + (1-\gamma) \cdot I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\mathcal{S},\tilde{k}}). \end{aligned} \tag{125}$$

By (114b) and (125):

$$\begin{aligned} \bar{\alpha}_{\mathcal{S},\tilde{k}} &+ I(\bar{U}_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|\bar{U}_{\mathcal{S},\tilde{k}-1}) \\ &= \big(\gamma\alpha_{\mathcal{S},\tilde{k}} + (1-\gamma)\alpha_{\mathcal{S},\tilde{k}+1}\big) \\ &\quad + \gamma I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|U_{\mathcal{S},\tilde{k}-1}) + (1-\gamma)I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\mathcal{S},\tilde{k}}) \\ &\geq \min\big\{\alpha_{\mathcal{S},\tilde{k}} + I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}}|U_{\mathcal{S},\tilde{k}-1}), \\ &\qquad\qquad \alpha_{\mathcal{S},\tilde{k}+1} + I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\mathcal{S},\tilde{k}})\big\}. \end{aligned} \tag{126}$$

Let now $k = \tilde{k} + 1$. We have:

$$
\begin{aligned}
& I(\bar{U}_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | \bar{U}_{\mathcal{S},\tilde{k}}) \\
& \stackrel{(a)}{=} (1-\beta) I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) \\
& \quad + \beta I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}-1}) \\
& \stackrel{(b)}{=} (1-\beta) I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) \\
& \quad + \beta I(U_{\mathcal{S},\tilde{k}+1}, U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}-1}) \\
& \stackrel{(c)}{=} I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) + \beta I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}-1}) \\
& \stackrel{(d)}{\geq} I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) + \beta I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}) \\
& \stackrel{(e)}{=} \gamma I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) + (1-\gamma) I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}),
\end{aligned}
$$
$$(127)$$

where (a) follows by the definition of $\bar{U}_{\mathcal{S},\tilde{k}}$ and $\bar{U}_{\mathcal{S},\tilde{k}+1}$; (b) by the Markov chain (64); (c) by the chain rule of mutual information and Markov chain (64); (d) by the degradedness of the channel (14b); (e) by the definition of β in (122).

Therefore, by (114c):

$$
\begin{aligned}
& \bar{\alpha}_{\mathcal{S},k+1} + I(\bar{U}_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | \bar{U}_{\mathcal{S},\tilde{k}}) \\
& \geq (1-\gamma) \cdot \alpha_{\mathcal{S},\tilde{k}} + \gamma \cdot \alpha_{\mathcal{S},\tilde{k}+1} \\
& \quad + (1-\gamma) \cdot I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}) + \gamma \cdot I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) \\
& \geq \min \big\{ \alpha_{\mathcal{S},\tilde{k}} + I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}), \\
& \qquad\qquad \alpha_{\mathcal{S},\tilde{k}+1} + I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) \big\}.
\end{aligned}
$$
$$(128)$$

We thus conclude that also when (120b) holds, the upper bound on $\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K)$ in (93) is relaxed if one replaces $(U_{\mathcal{S},1}, U_{\mathcal{S},2}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1})$ and $(\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},K})$ by $(\bar{U}_{\mathcal{S},1}, \bar{U}_{\mathcal{S},2}, \ldots, \bar{U}_{\mathcal{S},|\mathcal{S}|-1})$ and $(\bar{\alpha}_{\mathcal{S},1}, \ldots, \bar{\alpha}_{\mathcal{S},|\mathcal{S}|})$.

APPENDIX D
PROOF OF REMARK 3

We first prove that the bound in Theorem 5 is loosened when each $\alpha_{\mathcal{S},k}^{\star}$ is replaced by $\tilde{\alpha}_{\mathcal{S},k}$. Consider the intermediate Lemma 16 in the proof of Theorem 5, Appendix A. Relax the upper bound in this lemma by replacing for $k = 2, \ldots, K$ constraint (100a) by

$$
\alpha_{\mathcal{S},k} \geq 0. \tag{129}
$$

Following similar steps as in the proof of Lemma 17, see also [26, Lemma 12], it can be shown that this relaxed upper bound is not changed when one imposes that

$$
\alpha_{\mathcal{S},2} = \alpha_{\mathcal{S},3} = \ldots = \alpha_{\mathcal{S},|\mathcal{S}|},
$$

and

$$\alpha_{\mathcal{S},1} = \frac{\mathsf{M}_1}{D} \qquad \text{or} \qquad \alpha_{\mathcal{S},1} = \alpha_{\mathcal{S},2}.$$

Since constraints (101) are increasing in $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$, by constraint (100c), we conclude that the relaxed upper bound is loosest for

$$\alpha_{\mathcal{S},1} = \frac{\mathsf{M}_1}{D}$$

$$\alpha_{\mathcal{S},k} = \frac{|\mathcal{S}| \sum_{i=1}^{|\mathcal{S}|} \mathsf{M}_{j_i} - \mathsf{M}_{j_1}}{(|\mathcal{S}| - 1)D}, \quad k \in \{2, \ldots, |\mathcal{S}| - 1\},$$

i.e., for $\alpha_{\mathcal{S},k} = \tilde{\alpha}_{\mathcal{S},k}$.

We now prove that the bound in Theorem 5 is loosened when each $\alpha_{\mathcal{S},k}^\star$ is replaced by $\alpha_{\mathcal{S},k}'$. Consider again the intermediate Lemma 16 in Appendix A. Relax constraint (100a) by replacing it with $\alpha_{\mathcal{S},k} \geq 0$, for *all $k = 1, \ldots, K$*. Following the steps in [26, Lemma 12], it can be shown that the new constraints are loosest if each

$$\alpha_{\mathcal{S},k} = \alpha_{\mathcal{S},k}'. \tag{131}$$

This concludes the proof.

APPENDIX E

PROOF OF PROPOSITION 8

For $\Delta = 0$, achievability follows by specializing Theorem 3 to $t = K - 1$ and to the input distribution P_X that maximizes (70). In fact, for this input distribution:

$$R^{(K-1)} = K\mathsf{C}_{\text{avg}} = \mathsf{C}_{\text{avg}} + \frac{\sum_{k=1}^{K} \mathsf{M}_k^{\star(K-1)}}{K \cdot D}.$$

For $\Delta > 0$, achievability follows from Proposition 1.

The converse is proved as follows. Apply Theorem 5, but consider only the constraints (65) corresponding to the sets $\mathcal{S} = \{k\}$, for $k \in \mathcal{K}$. Taking the average over the resulting K constraints, establishes that there exists a random variable $(X, Y_1, \ldots, Y_K)$ satisfying (14b) and so that

$$\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) \leq \frac{1}{K} \sum_{k \in \mathcal{K}} I(X; Y_k) + \frac{1}{K} \sum_{k \in \mathcal{K}} \frac{\mathsf{M}_k}{D}. \tag{132}$$

Maximizing the right-hand side over input distributions P_X yields the desired converse.

APPENDIX F

PROOF OF PROPOSITION 12

Relax the upper bound in Theorem 5 by considering constraints (65) only for the set of all receivers $\mathcal{S} = \mathcal{K}$, and by replacing each $\alpha_{\mathcal{S},k}^\star$ by $\tilde{\alpha}_{\mathcal{S},k}$. Specializing the resulting relaxed bound to the erasure BC, one obtains the following upper bound:

$$\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) \leq \max \min \left\{ (1 - \delta_1)\beta_1 + \frac{\mathsf{M}_1}{D}, (1 - \delta_2)\beta_2 + \frac{K\mathsf{M} - \mathsf{M}_1}{D \cdot (K-1)}, \ldots, (1 - \delta_K)\beta_K + \frac{K\mathsf{M} - \mathsf{M}_1}{D \cdot (K-1)} \right\},$$

$$(133)$$

where the maximization is over the choice of parameters $\beta_1, \beta_2, \ldots, \beta_K \geq 0$ satisfying

$$\sum_{k=1}^{K} \beta_k \leq 1. \tag{134}$$

The upper bound in the proposition is established by solving this maximization problem. In fact, by noticing that the bound is increasing in $\beta_1, \beta_2, \ldots, \beta_K \geq 0$, and by first fixing β_1 and optimizing over the choices $\beta_2, \ldots, \beta_K \geq 0$ summing to $1 - \beta_1$, we obtain

$$
\begin{aligned}
\mathsf{C}(\mathsf{M}_1, &\ldots, \mathsf{M}_K) \\
\leq &\max_{\beta_1 \in [0,1]} \min \left\{ \beta_1 \mathsf{C}_1 + \frac{\mathsf{M}_1}{D}, \right. \\
&\qquad \left. (1 - \beta_1) \mathsf{C}_{\{2,\ldots,K\}} + \frac{K\mathsf{M} - \mathsf{M}_1}{(K-1) \cdot D} \right\}, \\
= &\max_{\beta_1 \in [0,1]} \min \left\{ \beta_1 \mathsf{C}_1, (1 - \beta_1) \mathsf{C}_{\{2,\ldots,K\}} + \frac{K(\mathsf{M} - \mathsf{M}_1)}{(K-1) \cdot D} \right\} \\
&+ \frac{\mathsf{M}_1}{D}.
\end{aligned}
$$

$$(135)$$

If

$$\frac{K(\mathsf{M} - \mathsf{M}_1)}{(K-1) \cdot D} \geq \mathsf{C}_1,$$

then the maximum is achieved at $\beta_1 = 1$ and the upper bound results in

$$\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) \leq \mathsf{C}_1 + \frac{\mathsf{M}_1}{D}. \tag{136}$$

Otherwise the maximum is at $\beta = \beta^\star$, where

$$\beta_1^\star := \frac{\mathsf{C}_{\{2,\ldots,K\}} + \frac{K(\mathsf{M}-\mathsf{M}_1)}{(K-1) \cdot D}}{\mathsf{C}_1 + \mathsf{C}_{\{2,\ldots,K\}}}, \tag{137}$$

and the upper bound results in

$$
\begin{aligned}
\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) &\leq \mathsf{C}_{\mathcal{K}} + \frac{K(\mathsf{M} - \mathsf{M}_1)}{(K-1) \cdot D} \cdot \frac{\mathsf{C}_1}{\mathsf{C}_1 + \mathsf{C}_{\{2,\ldots,K\}}} + \frac{\mathsf{M}_1}{D}, \\
&= \mathsf{C}_{\mathcal{K}} + \frac{K(\mathsf{M} - \mathsf{M}_1)}{(K-1) \cdot D} \cdot \frac{\mathsf{C}_{\mathcal{K}}}{\mathsf{C}_{\{2,\ldots,K\}}} + \frac{\mathsf{M}_1}{D},
\end{aligned}
\tag{138}
$$

where we used that for erasure BCs

$$\mathsf{C}_{\mathcal{K}} = \frac{\mathsf{C}_1 \cdot \mathsf{C}_{\{2,\ldots,K\}}}{\mathsf{C}_1 + \mathsf{C}_{\{2,\ldots,K\}}}. \tag{139}$$

APPENDIX G

PROOF OF COROLLARY 13

Fix $t \in \mathcal{K}$ and $\mathcal{S} = \{1, \ldots, t\}$. For the considered channel

$$(r_1 \ldots, r_t) \in \mathbf{C}_{\mathcal{S}} \quad \Longleftrightarrow \quad \sum_{k=1}^{t} r_k \leq 1. \tag{140}$$

The upper bound in Corollary 6 thus states that for this noise-free BC a rate-memory tuple $(R, \mathsf{M}_1, \ldots, \mathsf{M}_K)$ is achievable only if

$$tR - \sum_{k=1}^{t} \alpha^{\star}_{\mathcal{S},k} \leq 1. \tag{141}$$

This is equivalent to the following bound on the capacity-memory tradeoff

$$\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_K) \leq \frac{1}{t} \left(1 + \sum_{k=1}^{t} \alpha^{\star}_{\mathcal{S},k} \right). \tag{142}$$

Notice that the sum $\sum_{k=1}^{t} \alpha^{\star}_{\mathcal{S},k}$ takes on only two different values, depending on the outcomes of the minimizations defining $\alpha^{\star}_{\mathcal{S},k}$. It is either

$$\sum_{k=1}^{t} \alpha^{\star}_{\mathcal{S},k} = \frac{t \sum_{k=1}^{t} \mathsf{M}_k}{D} \tag{143a}$$

or

$$\sum_{k=1}^{t} \alpha^{\star}_{\mathcal{S},k} = \sum_{k=1}^{t} \frac{\sum_{i=1}^{k} \mathsf{M}_i}{D - k + 1}. \tag{143b}$$

Combining (142) with (143), applying the correspondence $\rho = R^{-1}$ and $m_k = \frac{\mathsf{M}_k}{R}$, and setting $m_1 = m_2 = \ldots = m_k = m$ yields,

$$1 \leq \frac{1}{t} \left(\rho + m \cdot \min \left\{ \frac{t^2}{D}, \sum_{k=1}^{t} \frac{k}{D - k + 1} \right\} \right), \tag{144}$$

which is equivalent to the bound in the corollary.

REFERENCES

[1] M. A. Maddah-Ali, U. Niesen, "Fundamental limits of caching," in *IEEE Trans. on Inform. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[2] Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: Improved bounds for small buffer users," *IET Commun.*, 2016, Vol. 10, Iss. 17, pp. 2315–2318.

[3] C. Tian, "A note on the fundamental limits of coded caching," *arXiv*, 1503.00010v1, Feb. 2015.

[4] K. Wan, D. Tuninetti, and P. Piantanida, "On the optimality of uncoded cache placement," in Proc. *IEEE ITW*, Cambridge,UK, 2016, pp. 161–165.

[5] K. Wan, D. Tuninetti, and P. Piantanida, "On caching with more users than files," in Proc. *IEEE ISIT*, Barcelona, Spain, July 2016, pp. 135–139.

[6] A. Sengupta, R. Tandon, and T. C. Clancy, "Improved approximation of storage-rate tradeoff for caching via new outer bounds," in Proc. *IEEE ISIT*, Hong Kong, China June 2015, pp. 1691–1695.

[7] S. Sahraei and M. Gastpar, "K users caching two files: An improved achievable rate," in Proc. *CISS*, pp. 620–624, Mar. 2016.

[8] C. Tian and J. Chen, "Caching and delivery via interference elimination," in Proc. *IEEE ISIT*, Barcelona, Spain, July 2016, pp. 830–834.

[9] C.-Y. Wang, S. H. Lim, and M. Gastpar, "A new converse bound for coded caching," in Proc. *IEEE ITA*, La Jolla (CA), Feb. 2016.

[10] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," in Proc. *IEEE ISIT*, Hong Kong, China, June 2015, pp. 1696–1700.

[11] H. Ghasemi and A. Ramamoorthy, "Further results on lower bounds for coded caching," in Proc. *IEEE ISIT*, Barcelona, Spain, July 2016, pp. 2319–2323.

[12] C. Tian, "Symmetry, demand types and outer bounds in caching system," in Proc. *IEEE ISIT*, Barcelona, Spain, July 2016, pp. 825–829.

[13] A. Amiri, Q. Yang, and D. Gündüz, "Coded caching for a large number of users," in Proc. *IEEE ITW*, Cambridge, UK, July 2016, pp. 171–175.

[14] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," in *arXiv*, 1609.07817, Sep. 2016.

[15] M. M. Amiri, Q. Yang, D. Gündüz, "Decentralized coded caching with distinct cache capacities," *arXiv*, 1610.03792, Oct., 2016.

[16] S. Wang, W. Li, X. Tian, and H. Liu, "Fundamental limits of heterogenous cache," *arXiv*, 1504.01123v1, Apr., 2015.

[17] R. Pedarsani, M. A. Maddah-Ali and U. Niesen, "Online coded caching," *IEEE/ACM Trans. Networking*, vol. 24, iss. 2, pp. 836–845, April 2016.

[18] U. Niesen and M. A. Maddah-Ali, "Coded caching with nonuniform demands," *IEEE Trans. on Inf. Theory*, vol. 63, no. 2, pp. 1146–1158, Feb. 2017.

[19] M. Ji, A. M. Tulino, J. Llorca, and G. Caire, "Order-optimal rate of caching and coded multicasting with random demands," *arXiv*, 1502.03124, Feb. 2015.

[20] J. Zhang, X. Lin, and X. Wang, "Coded caching under arbitrary popularity distributions," in Proc. *ITA*, La Jolla, USA, Feb. 2015, pp. 98–107.

[21] R. Timo, S. Saeedi Bidokhti, M. Wigger and B. Gieger, "A rate-distortion approach to caching," in Proc. *IZS*, Zurich, Switzerland, March, 2016, pp. 125–129.

[22] R. Timo, S. Saeedi Bidokhti, M. Wigger and B. Geiger, "A rate-distortion approach to caching," *arXiv*, 1610.07304, Oct. 2016.

[23] C.-Y. Wang, S. H. Lim, and M. Gastpar, "Information-theoretic caching: Sequential coding for computing," *IEEE Trans. on Inf. Theory*, vol. 62, no. 11, pp. 6393–6406, Nov. 2016.

[24] S. H. Lim, C.-Y. Wang, and M. Gastpar, "Information theoretic caching: The multi-user case," *arXiv*, 1604.02333, 2016.

[25] R. Timo and M. Wigger, "Joint cache-channel coding over erasure broadcast channels," in *Proc. ISWCS 2015*, Bruxelles, Belgium, Aug. 2015, pp. 201–205.

[26] S. Saeedi Bidokhti, M. Wigger, and R. Timo, "Noisy broadcast networks with receiver caching," submitted to *IEEE Trans. on Inf. Theory*, *arXiv*, 1605.02317, May, 2016.

[27] S. Saeedi Bidokhti, M. Wigger, and R. Timo, "Erasure broadcast networks with receiver caching," in Proc. *IEEE ISIT*, July 2016. pp. 1819–1823.

[28] S. Saeedi Bidokhti, M. Wigger, and R. Timo, "An upper bound on the capacity-memory tradeoff of degraded broadcast channels," in Proc. of *IEEE ISTC*, Brest, France, Sep., 2016, pp. 350–354.

[29] W. Huang, S. Wang, N. Ding, F. Yang, and W. Zhang, "The performance analysis of coded cache in wireless fading channel," *arXiv*,1504.01452, Apr. 2015.

[30] S. Wang, X. Tian and H. Liu, "Exploiting the unexploited of coded caching for wireless content distribution," in *Proc. IEEE ICNC*, 2015, pp. 700–706.

[31] P. Hassanzadeh, E. Erkip, J. Llorca,and A. Tulino, "Distortion-memory tradeoffs in cache-aided wireless video delivery," in *Proc. of Allerton Conference* Monticello (IL), USA, Oct. 2015, pp. 1150–1157.

[32] A. S. Cacciapuoti, M. Caleffi, M. Ji, J. Llorca, A. M. Tulino,"Speeding up future video distribution via channel-aware caching-aided coded multicast," *IEEE JSAC in Comm.*, vol. 34, no. 8, pp. 2207–2218, Aug., 2016.

[33] A. Ghorbel, M. Kobayashi, and S. Yang, "Cache-enabled broadcast packet erasure channels with state feedback," in *Proc. of Allerton Conference,* Monticello (IL), USA, Oct. 2015, pp. 1446–1453.

[34] A. Ghorbel, M. Kobayashi, and S. Yang "Content delivery in erasure broadcast channels with cache and feedback," *IEEE Trans. on Inf. Theory*, vol. 62, no. 11, pp. 6407–6422, Nov. 2016.

[35] A. Ghorbel, M. Kobayashi, and S. Yang, "Content delivery in erasure broadcast channels with cache and feedback," in Proc. *IEEE ISIT*, July 2016, Barcelona, Spain, pp. 835–839.

[36] S. Yang, K.-H. Ngo, and M. Kobayashi, "Content delivery with coded caching and massive MIMO in 5G," in Proc. of *9th ISTC*, Sep. 2016, Brest, France, pp. 370–374.

[37] J. Zhang and P. Elia, "Fundamental limits of cache-aided wireless BC: interplay of coded-caching and CSIT feedback," *to appear in IEEE Trans. on Inf. Theory, 2017.*

[38] J. Hachem, N. Karamchandani, and S. Diggavi, "Content caching and delivery over heterogeneous wireless networks," in Proc. *IEEE Infocom*, Kowloon, China, March 2015, pp. 756–764.

[39] J. Zhang, F. Engelmann, and P. Elia, "Coded caching for reducing CSIT-feedback in wireless communications," in Proc. *Allerton Conf.*, Oct. 2015, pp. 1099–1105.

[40] J. Zhang and P. Elia, "Wireless coded caching: A topological perspective," *arXiv*, 1606.08253.

[41] J. Zhang and P. Elia, "Feedback-aided coded caching for the MISO BC with small caches," *arXiv*, 1606.05396.

[42] X. Yi and G. Caire, "Topological coded caching," in Proc. *IEEE ISIT*, July 2016, Barcelona, Spain, pp. 2039–2043.

[43] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "Edge-facilitated wireless distributed computing," in Proc. *IEEE GLOBECOM*, Dec. 2016.

[44] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in Proc. *IEEE ISIT*, Hong Kong, China, June 2015, pp. 809–813

[45] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr, "Fundamental limits of cache-aided interference management," *to appear in IEEE Trans. on Inf. Theory*, 2017.

[46] S. Pooya Shariatpanah, S. Abolfazl Motahari, B. Hossein Khalaj, "Multi-server coded caching," *IEEE Trans. on Inf. Theory*, vol. 62, no. 12, pp. 7253–7271, Dec. 2016.

[47] M. Wigger, R. Timo, and S. Shamai, "Complete interference mitigation through receiver-caching in Wyner's networks," in Proc. *IEEE ITW*, Cambridge, UK, Sep. 2016, pp. 335–339.

[48] V. Ntranos, M. A. Maddah-Ali, and G. Caire, "Cellular interference alignment," *IEEE Trans. Inform. Theory,* vol. 61, no. 3, March 2015, pp. 1194–1217.

[49] Y. Ugur, Z. H. Awan and A. Sezgin, "Cloud radio access networks with coded caching." *arXiv*, 1512.02385, Dec. 2015.

[50] S.-H. Park, O. Simeone, and S. Shamai (Shitz), "Joint optimization of cloud and edge processing for fog radio access networks," *IEEE Trans. on Wireless Comm.,* vol. 15, no. 11, pp. 7621–7632, Nov. 2016.

[51] R. Tandon and O. Simeone, "Cloud-aided wireless networks with edge caching: Fundamental latency trade-offs in fog radio access networks," in Proc. *IEEE ISIT*, Barcelona, Spain, July 2016, pp. 2029–2033.

[52] B. Azari, O. Simeone, U. Spagnolini and A. Tulino, "Hypergraph-based analysis of clustered cooperative beamforming with application to edge caching," *IEEE Wireless Comm. Letters*, vol. 5, no. 1, pp. 84–87, Feb. 2016.

[53] M. Peng, S. Yan, K. Zhang and C. Wang, "Fog computing based radio access networks: issues and challenges," *IEEE Network*, vol. 30, iss. 4, pp. 46–53, July-Aug. 2016.

[54] N. Karamchandani, U. Niesen, M. A. Maddah-Ali, and S. Diggavi, "Hierarchical coded caching," in Proc. *IEEE ISIT*, Honolulu, USA, June 2014, pp. 2142–2146.

[55] J. Hachem, N. Karamchandani, and S. Diggavi, "Effect of number of users in multi-level coded caching," in Proc. *IEEE ISIT*, Hong Kong, China, June 2015, pp. 1701–1705.

[56] J. Hachem, N. Karamchandani and S. Diggavi, "Multi-level coded caching," iin Proc. *IEEE ISIT*, Honolulu, USA, June 2014, pp. 56–60.

[57] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," in Proc. *ICC*, London, UK, July 2015, pp. 3358–3363.

[58] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 1469 –1482, April, 2006.

[59] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problem. Inf. Transm.*, vol 10, no. 3, pp. 3–14, 1974.

[60] P. P. Bergmans, "A simple converse for broadcast channels with additive white Gaussian noise," *IEEE Trans. on Inf. Theory*, vol. 20, no. 2, pp. 279–280, 1974.

[61] A. El Gamal and Y-H Kim, *Network information theory.* Cambridge, U.K.: Cambridge Univ. Press, 2011.

[62] T. M. Cover and J. Y. Thomas. *Elements of Information Theory*, Wiley-Interscience, New York, 2006.