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 interesting special cases. The lower bounds are achieved by two new coding schemes that benefit from *non-uniform cache assignments*. 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, where the thresholds depend on 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 by 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 the 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. It is observed numerically that a uniform 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 times. The main challenge is that the contents have to be cached before knowing which files the users will request in peak-traffic periods. A conventional approach is to store popular contents in the cache memories of all the users. This allows the receivers to locally retrieve the contents without burdening the network and attain the so called *local caching gain*. Maddah-Ali and Niesen [1] have shown that further caching gains, i.e., the so called *global caching gains*, are achievable if different contents are stored at different users in a careful manner.

In particular, reference [1] considered a broadcast scenario with a transmitter that has access to a library of N independent files and with K receivers that are all equipped with individual cache memories of the same size. Communication during off-peak periods, when contents are cached, is assumed error-free and constrained only by the amount of information that can be placed in the cache memories at the receivers. This communication is henceforth called *cache placement* or *placement phase*. The subsequent peak-traffic communication is called *delivery phase*. In this phase, each receiver requests a single file from the library and the transmitter delivers the requested files by communicating over a common noise-free link to all K receivers. Reference [1] has proposed to diversify the placed contents across cache memories so as to allow for coding opportunities during the delivery phase. These coding opportunities allow to simultaneously serve multiple receivers in each transmission, providing gains that scale with the total size of all cache memories in the network, i.e., global caching gains. By contrast, the previously reported gains depend only on individual cache sizes and are thus referred to as local caching gains.

The performance metric in [1] is the minimum required delivery rate for given cache sizes, leading to the fundamental quantity of interest, the *delivery rate-memory tradeoff*. Upper and lower bounds on this tradeoff are provided in [1]. Improved upper bounds (achievability results) have subsequently been presented in [2]–[8] and improved lower bounds (converse results) in [9]–[13]. The common noise-free link model of [1] has also been studied for networks in which receivers have cache memories of different sizes [14]–[16]. Rate-limited links in delivery have been considered in [17].

In this paper, we relax the assumption that delivery takes place over a noise-free link. Instead, we model the delivery phase by a degraded broadcast channel (BC). The class of degraded broadcast channels is a fairly general class that includes practical noisy channel models such as broadcast erasure channels and Gaussian channels. Our model is depicted in Figure 1. A transmitter communicates with receivers $1, \ldots, \mathsf{K}$ which are equipped with cache memories of sizes $n\mathsf{M}_1, \ldots, n\mathsf{M}_\mathsf{K}$, when communication is of blocklength n.

Noisy broadcast channels with caching receivers have been studied in different settings [18]–[30]. For example, references [18]–[20] explore the benefits of coded caching for Gaussian or slow fading BCs when all users have the same cache sizes. References [26]–[30] study the interplay between coded caching with spatial multiplexing (MIMO), channel state information (CSI), or feedback. Most related to the current work are the works in [21]–[24] which focus on erasure BCs with a set of weak receivers that are equipped with cache memories of equal size and a set of strong receivers without cache
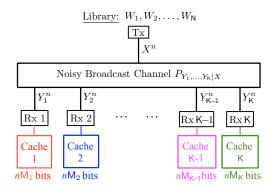
Fig. 1. Noisy broadcast channel with cache memories at the receivers.

Fig. 2. Separate cache-channel coding architecture.

memories or with smaller cache memories. In accordance with multi-user information theory metrics, performance in these works is measured in terms of the *capacity-memory tradeoff*, i.e., the largest *message rate* for which receivers can decode their requested messages reliably as a function of the cache sizes. Lower bounds (achievability results) and upper bounds (converse results) are presented on the capacity-memory tradeoff. The lower bounds are based on *joint cache-channel coding schemes* where encoders and decoders exploit both the knowledge of the channel statistics and the cache contents. This is in contrast to previous works, e.g., [18], which adopt a *separate cache-channel coding* architecture. In separate cache-channel coding, the encoders (resp. decoders) consist of (i) a cache encoder (resp. decoder) that only exploits the cache contents and (ii) a channel encoder (resp. decoder) that only exploits the channel statistics; see Figure 2. As the results in [21, 22] show, when receivers have different channel statistics and weaker receivers have larger cache sizes, then adopting a joint cache-channel coding architecture can significantly improve performance. Moreover, [21] illustrates that it is beneficial to assign larger cache memories to weaker receivers than to stronger receivers.

Joint cache-channel coding schemes[1] have also been employed for transmission over noisy BCs with caching receivers when the files to be sent are correlated [31, 32] or when receivers have different fidelity constraints [24, 25, 33]. In these applications, improvements are possible even when users have perfectly symmetric channels and cache sizes.

In this work, we consider the problem of efficient cache assignment in degraded broadcast channels and how to code under these assignments. We quantify new caching gains obtained through such assignments and appropriate coding.

A. A Motivating Example

Consider an erasure BC with $K = 10$ users where receiver 1 has erasure probability $\delta_1 = 0.4$ and all other receivers $2, \ldots, K$ have erasure probability $\delta_2 = \ldots = \delta_K = 0.1$. Assuming no cache memories, denote the point to point channel capacity of user k by $C_{\{k\}}$, and the symmetric capacity
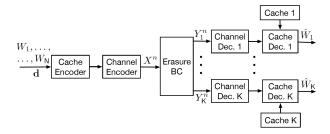
---

[1]Joint cache-channel coding schemes are also referred to as *joint source channel coding* schemes because once the cache contents are fixed, they can be recast as realizations of sources that contribute side-information to the receivers.

of the K-user broadcast channel by $C_{\{1,\ldots,K\}}$ (see Section III for definitions). Now suppose that we are given a total cache size of nM to be distributed across the receivers (with any desired assignment), hence $M = n(M_1 + \ldots + M_K)$. We seek cache assignments, as well as coding schemes, that achieve high message rate. In particular, we are interested in the *global capacity-memory tradeoff* which we define as the largest rate achievable given the total cache budget M.

The traditional approach assigns the same cache size $n\frac{M}{K}$ to each receiver. We assign cache memories according to the channel strengths. For example, when the total cache rate M is small, we assign all the cache rate to the weak receiver, so $M_1 = M$. For large total cache rate, in the example at hand, we propose to assign a larger portion to the weak receiver and to distribute the rest uniformly over all strong receivers. The meaning of small and large cache sizes are made precise in Corollary 11 in Section VI. Tables I and II compare the rates that are achievable for this example by the cache assignments and coding schemes of this paper as well as traditional uniform cache assignments and standard (separate cache-channel) codes. Table I treats the regime of small cache memories and Table II treats the regime of large cache memories. In both tables, the first column presents an upper bound on the capacity-memory tradeoff $C\left(\frac{M}{K}, \ldots, \frac{M}{K}\right)$ under uniform cache assignment (i.e., when the total cache memory M is assigned uniformly over the K users). The second column shows the rate that is achievable using the cache assignments and coding schemes proposed in this paper. We will show that these rates equal the global capacity-memory tradeoff $C^\star(M)$ in the regimes of small and large cache sizes. The third column considers the same cache assignment as in column 2, but presents the achievable rate R using standard (separate cache-channel) codes. More specifically, in the regime of small cache sizes, a standard BC code is used to communicate to each receiver the part of its requested file that is not in its cache memory. Since in this example only the weak receiver has a cache memory, coded caching (multicasting) is not possible. Under the cache assignment in Table II (large cache memories), the rate achieved with separate source-channel coding is sufficiently small so that the weak receiver 1 can store all files in its cache memory, thus precluding delivery communication to this receiver. The optimal separation-based strategy is then to perform coded caching with parameter $K - 2$ (designed for the strong receivers) followed by a standard BC multicast code to those receivers. A comparison of columns 2 and 3 in Tables (I) and (II) shows that a smart

$M_1 = M_2 = \ldots = M_K$	$M_1 = M$ and $M_2 = \ldots = M_K = 0$	
Upper bound with uniform assignment	New coding	Standard coding
$C\left(\frac{M}{K}, \ldots, \frac{M}{K}\right) \leq C_{\{1,\ldots,K\}} + \frac{83}{530} \cdot \frac{M}{N}$	$C^\star(M) = C_{\{1,\ldots,K\}} + \frac{M}{N}$	$R = C_{\{1,\ldots,K\}} + \frac{1}{7} \cdot \frac{M}{N}$

TABLE I

COMPARISON OF TRADITIONAL SCHEMES WITH THE ONES PROPOSED IN THIS PAPER IN THE REGIME OF "SMALL" TOTAL CACHE SIZES M.

$M_1 = M_2 = \ldots = M_K$	$M_1 = (8.1 + \Delta)N$ and $M_2 = \ldots = M_K = (7.8 + \Delta)N$, for any $\Delta > 0$	
Upper bound with uniform assignment	New coding	Standard coding
$C\left(\frac{M}{K}, \ldots, \frac{M}{K}\right) \leq C_{\{1\}} + \frac{M}{K \cdot N}$ $= 0.6 + \frac{M}{K \cdot N}$	$C^\star(M) = \frac{1}{K} \sum_{k=1}^{K} C_{\{k\}} + \frac{M}{K \cdot N}$ $= 0.87 + \frac{M}{K \cdot N}$	$R = \frac{1}{K} \sum_{k=1}^{K} \left(C_{\{k\}} - C_{\{1\}} \right) + \frac{M}{K \cdot N}$ $= 0.27 + \frac{M}{K \cdot N}$

TABLE II

COMPARISON OF TRADITIONAL SCHEMES WITH THE ONES PROPOSED IN THIS PAPER IN THE REGIME OF "LARGE" TOTAL CACHE SIZES M.

cache assignment creates new coding opportunities that can be exploited by joint cache-channel coding.

From Table I, column 2, we further observe the following behavior of the global capacity-memory tradeoff $C^\star(M)$: First of all, without cache memory, $C^\star(M = 0)$ equals the largest symmetric rate $C_{\{1,\ldots,K\}}$ that is achievable to all the receivers in the BC. Now consider the slope with which $C^\star(M)$ increases with M. For small total cache budget M, smart cache assignment and coding (see column 2) allow to attain a steeper slope than traditional uniform cache assignment (see column 1) as well as separate cache-channel coding (see column 3). In particular, our proposed cache assignment and coding allow to achieve what we call a *perfect caching gain*, where the capacity-memory tradeoff grows as $\frac{M}{N}$, i.e., like the *total size of all cache memories* in the network divided by the number of files N. This is the same performance as if each receiver had access to all cache memories in the network.

From Table II, we observe that for large cache memories, a smart cache assignment increases the capacity from $C_{\{1\}} + \frac{M}{KN}$ to $\frac{1}{K} \sum_{k=1}^{K} C_{\{k\}} + \frac{M}{KN}$. So the weak channel of user 1 is no longer limiting in the delivery phase. The gain of additional cache memories is, however, only local, i.e., the capacity-memory tradeoff only grows as $\frac{M}{KN}$.

We remark that the results in this paper are not restricted to erasure BCs and hold for general memoryless degraded BCs.

B. Main Contributions and Implications

The main contributions of this paper are as follows.

- *New coding schemes*: We propose two new joint cache-channel coding schemes for degraded broadcast networks with heterogeneous cache sizes: *superposition piggyback-coding* and *generalized coded caching*. In superposition piggyback-coding, we assume a single cache memory at the weakest receiver and our delivery scheme loads (piggybacks) the information that is intended for stronger receivers and cached at weakest receiver onto the infor-

mation that is communicated to this weakest receiver[2]. When the rate of the piggybacked information is modest, the decoding at the strong receivers can be done without harming the performance at the weak receivers. The communication to the stronger receivers can thus be viewed as being almost for free. In some sense, piggyback coding provides the stronger receivers virtual access to the weak receivers' cache-memories as if these cache contents were locally present at the stronger receivers. *All* receivers gain virtual access to the weakest receiver's cache memory (the only cache memory in this case) and hence *perfect caching gain* is achieved.

In generalized coded caching, all receivers have cache memories, but weaker receivers have larger cache sizes than stronger receivers. We build the placement and delivery similar to the coded-caching scheme in [1]. However, by assigning larger cache memories to the weaker users, we create a new coding opportunity: piggyback coding. We piggyback information for the stronger receivers on the communication to the weaker receivers (without harming the weaker receivers' decodability). Using our coding scheme, 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. We show that generalized coded caching is optimal for a specific cache assignment.

- *A New Converse Result:* We prove a general converse result for degraded BCs with arbitrary cache sizes at the receivers. Our result strictly improves over the existing converse results for degraded BCs in [22, 35], and at the time of submission[3], also over all previous converse results for the noise-free bit-pipe model [1, 10]–[12].
- *Global Capacity-Memory Tradeoff:* We study the problem of cache assignment on cache-aided noisy broadcast

---

[2]Piggyback coding was proposed in a version without superposition coding in [21, 22]. This original version can be seen as a simplified version of the coding scheme in [34] (without binning) in the context of Slepian-Wolf coding over broadcast channels.

[3]The parallel work [36] slightly improves on this bound for the noise-free bit-pipe model (but does not generalize to noisy channels); see also [37].

networks, and derive new upper and lower bounds on the global capacity-memory tradeoff, i.e., on the largest rate that is achievable under an optimized cache assignment. The bounds match when the total available cache budget is small or large. For a small total cache budget M, a perfect global caching gain is achievable. For larger cache budgets M, the caching gain diminishes as M increases. Finally for M larger than a certain threshold, only a *local caching gain* is possible; i.e., the performance corresponds to a system where all receivers store the same content and no coded delivery is possible. In this case, the global capacity-memory tradeoff grows as $\frac{M}{K \cdot N}$, where K denotes the number of receivers. Finally, we demonstrate numerically that the popular approach of assigning equal cache sizes to all receivers is suboptimal over Gaussian and erasure BCs. We further prove this analytically for erasure BCs in the small cache size regime.

C. 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 N. 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 the sequence $A_1, \ldots, A_n$ where n is an integer. 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 $\mathcal{A}^n$. Further, $|\mathcal{A}|$ denotes the cardinality of $\mathcal{A}$. The notation $(a)^+$, for $a \in \mathcal{R}$, refers to $\max(0, a)$.

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

For indices w_1 and w_2 taking value in $\{1, \ldots, \lfloor 2^{\ell_1} \rfloor\}$ and $\{1, \ldots, \lfloor 2^{\ell_2} \rfloor\}$, respectively, we define the generalized XOR as the index in $\{1, \ldots, \lfloor 2^{\ell_{\max}} \rfloor\}$ that corresponds to the componentwise XOR of the length-$\ell_{\max}$ binary representations of w_1 and w_2, where $\ell_{\max} := \max\{\ell_1, \ell_2\}$. We denote the generalized XOR of w_1 and w_2 by

$$w_1 \overline{\bigoplus} w_2.$$

D. Outline

The remainder of the paper is organized as follows. Section II describes the problem setup. Section III summarizes known results for the scenario where there is no cache memory in the network. The main results of this paper are described in Sections IV–VI. Section VII specializes these results to the examples of erasure and Gaussian BCs, and to the noise-free bit-pipe model. The paper is concluded in Section VIII.

II. PROBLEM DEFINITION

Consider a network with a transmitter and receivers $1, \ldots, \mathsf{K}$. The transmitter has access to a library with N independent messages, $W_1, \ldots, W_\mathsf{N}$, 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 $\mathsf{N} \geq \mathsf{K}$.

Each receiver $k \in \mathcal{K} := \{1, \ldots, \mathsf{K}\}$ is equipped with a cache of size $n\mathsf{M}_k$ bits, where we have $\mathsf{M}_k \geq 0$. The sizes of the cache memories thus scale linearly in the blocklength n. Communication takes place in two phases. First, in the placement phase, the transmitter chooses caching functions

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

and places

$$V_k := g_k(W_1, \ldots, W_\mathsf{N}) \tag{2}$$

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

The subsequent delivery phase takes place over a *degraded* BC [38] with finite input alphabet $\mathcal{X}$, finite output alphabets $\mathcal{Y}_1, \ldots, \mathcal{Y}_\mathsf{K}$,[4] and the channel transition law

$$\Gamma(y_1, \ldots, y_\mathsf{K}|x), \text{ for } x \in \mathcal{X}, y_1 \in \mathcal{Y}_1, \ldots, y_\mathsf{K} \in \mathcal{Y}_\mathsf{K} \tag{3}$$

which decomposes as

$$\Gamma(y_1, \ldots, y_\mathsf{K}|x) = \Gamma_\mathsf{K}(y_\mathsf{K}|x) \cdot \Gamma_{\mathsf{K}-1}(y_{\mathsf{K}-1}|y_\mathsf{K}) \cdots \Gamma_1(y_1|y_2). \tag{4}$$

Without loss of generality, this decomposition assumes that the K users of the degraded BC are ordered from weakest to strongest.

At the beginning of the delivery phase, each receiver k produces a random demand D_k from the set $\mathcal{N} := \{1, \ldots, \mathsf{N}\}$ to indicate that it wishes to learn message W_{D_k}. The transmitter and all the receivers are informed about the entire demand vector

$$\mathbf{D} := (D_1, \ldots, D_\mathsf{K}).$$

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

$$X^n = f(W_1, \ldots, W_\mathsf{N}, \mathbf{D}) \tag{5}$$

for some encoding function $f : \{1, \ldots, \lfloor 2^{nR} \rfloor\}^\mathsf{N} \times \mathcal{N}^\mathsf{K} \to \mathcal{X}^n$.

Each receiver $k \in \mathcal{K}$ observes the outputs $Y_k^n := (Y_{k,1}, \ldots, Y_{k,n})$ of the DMC $\Gamma(y_1, \ldots, y_K|x)$ for inputs X^n. With the previously learned demand vector $\mathbf{D}$, its local cache content V_k, and the channel outputs Y_k^n, it then produces its estimate of the desired message W_{D_k}:

$$\hat{W}_k := \varphi_k(Y_k^n, V_k, \mathbf{D}) \tag{6}$$

by means of a decoding function

$$\varphi_k \colon \mathcal{Y}_k^n \times \{1, \ldots, \lfloor 2^{n\mathsf{M}_k} \rfloor\} \times \mathcal{N}^\mathsf{K} \to \{1, \ldots, \lfloor 2^{nR} \rfloor\}. \tag{7}$$

We define the probability of error as:

$$\mathsf{P}_\mathsf{e}^{(n)} := \mathbb{E}_\mathbf{D}[\mathsf{P}_\mathsf{e}^{(n)}(\mathbf{D})], \tag{8}$$

where

$$\mathsf{P}_\mathsf{e}^{(n)}(\mathbf{d}) := \mathbb{P}\left[\bigcup_{k=1}^\mathsf{K} \{\hat{W}_k \neq W_{d_k}\} \middle| \mathbf{D} = \mathbf{d} \right], \tag{9}$$

---

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

and we assume that the random demand vector $\mathbf{D}$ has a uniform distribution on $\mathcal{N}^{\mathsf{K}}$; i.e., $\mathbf{D} = \mathbf{d}$ with probability $\frac{1}{N^{\mathsf{K}}}$ for every $\mathbf{d} \in \mathcal{N}^{\mathsf{K}}$. Notice that the probability of error cannot be reduced by using *stochastic* instead of *deterministic* caching, encoding, and/or decoding functions. It is thus without loss in optimality that in this paper we assume deterministic functions.

A rate-memory tuple $(R, \mathsf{M}_1, \ldots, \mathsf{M}_{\mathsf{K}})$ is *achievable* if for any $\epsilon > 0$ there exists a sufficiently large blocklength n and caching, encoding, and decoding functions as in (2), (5), and (6) so that $\mathsf{P}_{\mathrm{e}}^{(n)} \leq \epsilon$.

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

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

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

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

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

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

Remark 1: The global capacity memory tradeoff depends on the BC law $\Gamma(y_1, \ldots, y_{\mathsf{K}} | x)$ only through its marginal conditional laws $\Gamma_1(y_1 | x), \ldots, \Gamma_{\mathsf{K}}(y_{\mathsf{K}} | x)$. All our results thus also apply to *stochastically degraded BCs*.

A. Minimum Delivery Rate

Most previous works on caching that modeled the BC as a noise-free bit-pipe, e.g. [1], have fixed the size of the messages to F bits and assumed that delivery takes place over $\rho \cdot F$ channel uses and that each receiver $k \in \mathcal{K}$, has $m_k F$ bits of cache memory. *Delivery rate* ρ is then said to be achievable given normalized cache memory sizes $m_1, \ldots, m_{\mathsf{K}}$ if there exist caching, encoding, and decoding functions such that the probability of error in (8) tends to 0 as $F \to \infty$. In Section VII-B, we specialize our results to the noise-free bit-pipe channel model and use the notion of delivery rate to compare our results with the state-of-the-art.

It is not difficult to see the following correspondence between the two definitions:

Message rate R is achievable with $(\mathsf{M}_1, \ldots, \mathsf{M}_{\mathsf{K}})$

$$\Longleftrightarrow$$

Delivery rate $\rho = \dfrac{1}{R}$ is achievable with $\left(m_1, \ldots, m_{\mathsf{K}}\right)$

where $m_1 = \dfrac{\mathsf{M}_1}{R}, \ldots, m_{\mathsf{K}} = \dfrac{\mathsf{M}_{\mathsf{K}}}{R}$.

III. PRELIMINARIES: CAPACITIES WITHOUT CACHE MEMORIES

In this section, we recall known results for our network in the special case when there are no cache memories:

$$\mathsf{M}_1 = \ldots = \mathsf{M}_{\mathsf{K}} = 0. \qquad (12)$$

These results will be utilized in the subsequent sections to state our results for cache-aided broadcast networks.

When (12) holds, it is know from [38, 39] that it is optimal to have the stronger receivers also decode messages intended to weaker receivers. This is done by the so called superposition coding. Here, the worst case probability of error in (8) is attained for a demand vector $\mathbf{d}$ that has all different entries and the capacity-memory tradeoff $\mathsf{C}(\mathsf{M}_1 = 0, \ldots, \mathsf{M}_{\mathsf{K}} = 0)$ is the largest symmetric rate R with which K independent messages can be sent reliably to the K receivers. We thus have

$$\mathsf{C}(\mathsf{M}_1 = 0, \ldots, \mathsf{M}_{\mathsf{K}} = 0) = \mathsf{C}_{\mathcal{K}} \qquad (13)$$

where $\mathsf{C}_{\mathcal{K}}$ is found from the capacity of degraded broadcast channels [38, 39]:

$$\mathsf{C}_{\mathcal{K}} := \max \min \big\{ I(U_1; Y_1), I(U_2; Y_2 | U_1), \ldots,$$
$$I(U_{\mathsf{K}-1}; Y_{\mathsf{K}-1} | U_{\mathsf{K}-2}), I(X; Y_{\mathsf{K}} | U_{\mathsf{K}-1}) \big\}. \qquad (14)$$

The maximization in (14) is over all auxiliary random tuples $U_1, \ldots, U_{\mathsf{K}-1}, X, Y_1, \ldots, Y_{\mathsf{K}}$ that satisfy the following Markov chain:

$$U_1 \to U_2 \to \cdots \to U_{\mathsf{K}-1} \to X \to (Y_1, \ldots, Y_{\mathsf{K}}) \qquad (15a)$$

and the channel transition law:

$$P_{Y_1 \ldots Y_{\mathsf{K}} | X}(y_1, \ldots, y_{\mathsf{K}} | x) = \Gamma(y_1, \ldots, y_{\mathsf{K}} | x). \qquad (15b)$$

Denote the alphabet sets of the auxiliary random variables $U_1, \ldots, U_{\mathsf{K}}$ by $\mathcal{U}_1, \ldots, \mathcal{U}_{\mathsf{K}}$. Using the Fenchel-Eggleston-Carathéodory theorem [40, Appendix A], without loss of generality, one can restrict the cardinality of the sets as follows:

$$|\mathcal{U}_k| \leq |\mathcal{U}_{k+1}| + 2 \qquad (16)$$
$$|\mathcal{U}_{\mathsf{K}-1}| \leq |\mathcal{X}| + 2. \qquad (17)$$

To present the results in this paper, we will also need the *no-cache capacity region* of the BC to a subset of the receivers

$$\mathcal{S} := \{j_1, \ldots, j_{|\mathcal{S}|}\} \subseteq \mathcal{K}, \qquad j_1 < \cdots < j_{|\mathcal{S}|}. \qquad (18)$$

The no-cache capacity region $\mathcal{C}_{\mathcal{S}}$ is naturally given by the set of all nonnegative rate-tuples $(R_1, \ldots, R_{|\mathcal{S}|})$ for which there exist random variables $U_1, \ldots, U_{|\mathcal{S}|-1}, X, Y_{j_1}, \ldots, Y_{j_{|\mathcal{S}|}}$ that satisfy (15b) and form the Markov chain

$$U_1 \to U_2 \to \cdots \to U_{|\mathcal{S}|-1} \to X \to \left(Y_{j_1}, \ldots, Y_{j_{|\mathcal{S}|}}\right), \qquad (19)$$

such that the following conditions hold:

$$R_1 \leq I(U_1; Y_{j_1}), \qquad (20a)$$
$$R_k \leq I(U_k; Y_{j_k} | U_{k-1}), \quad k \in \{2, \ldots, |\mathcal{S}|-1\}, \qquad (20b)$$
$$R_{|\mathcal{S}|} \leq I\big(X; Y_{j_{|\mathcal{S}|}} | U_{|\mathcal{S}|-1}\big). \qquad (20c)$$

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

$$C_{\mathcal{S}} := \max_{R \geq 0} \{R \colon (R, \ldots, R) \in \mathcal{C}_{\mathcal{S}}\}. \tag{21}$$

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

By (20) and (21):

$$C_{\mathcal{S}} = \max \min \big\{ I(U_1; Y_{j_1}), I(U_2; Y_{j_2}|U_1), \ldots,$$
$$I(U_{|\mathcal{S}|-1}; Y_{j_{|\mathcal{S}|-1}}|U_{|\mathcal{S}|-2}), I(X; Y_{j_{|\mathcal{S}|}}|U_{|\mathcal{S}|-1}) \big\}, \tag{22}$$

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 (15b) and (19).

IV. CODING SCHEMES AND LOWER BOUNDS ON THE CAPACITY-MEMORY TRADEOFF

We present three lower bounds on the capacity-memory tradeoff along with coding schemes that achieve them. The first coding scheme and lower bound apply for general cache assignments $M_1, \ldots, M_K$. The second and third ones apply only to specific cache assignments. Nevertheless, the proposed schemes are useful for a broad set of cache-assignments by time- (and memory-) sharing different coding schemes, or equivalently, taking convex combinations of different lower bounds.

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. Global caching gains are not possible under this caching strategy.

Applying the described simple caching strategy to only a part of the cache memory that is of size $\Delta \geq 0$, while allowing a smart use of the remaining memory, leads to the following proposition, see also [41, Proposition 1].

Proposition 1 (Local caching gain): For all $\Delta > 0$ and $M_1, \ldots, M_K \geq 0$:

$$C(M_1 + \Delta, \ldots, M_K + \Delta) \geq C(M_1, \ldots, M_K) + \frac{\Delta}{N}. \tag{23}$$

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

$$C^\star(M + \Delta_{\text{total}}) \geq C^\star(M) + \frac{\Delta_{\text{total}}}{K \cdot N}. \tag{24}$$

We will see that this lower bound is tight in certain regimes of operation, namely, when the cache budget is larger than a threshold.

B. Superposition Piggyback-Coding

We generalize the piggyback coding scheme of [22, 35], that was specific to erasure BCs, to general degraded BCs by introducing superposition coding. The scheme assigns all the available cache memory to the weakest receiver, and uses a layered superposition code for delivery, see Fig. 3. In this superposition code:

- the lower-most layer encodes the part of message W_{d_1} (intended for receiver 1) that is *not* stored in receiver 1's
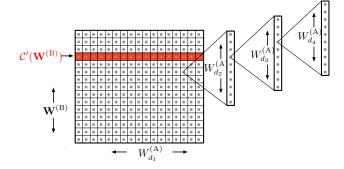

Fig. 3. Superposition piggyback code construction $\mathcal{C}$ for $K = 4$ receivers. Recall that receiver 1 is the weakest. Each dot represents a codeword.

cache memory and the parts of messages $W_{d_2}, \ldots, W_{d_K}$ that *are* stored at receiver 1.
- the k-th lowest layer, for $k \in \{2, \ldots, K\}$, encodes the part of the message W_{d_k} (intended for receiver k) that is *not* stored in the cache memory of receiver 1.

When decoding, each receiver $k \in \{2, \ldots, K\}$, decodes the k lowest layers using, e.g., a standard joint typicality decoder. In particular, receiver 1 (the weakest user) only decodes the lowest layer. This layer encodes a part of message W_{d_1} that is desired at receiver 1 together with parts of messages $W_{d_2}, \ldots, W_{d_K}$ that are not desired at receiver 1, but are locally available at its cache memory. As we will see, the cache content allows receiver 1 to achieve the same decoding performance as if the additional messages to the other receivers were not encoded in the lowest layer. In other words, we can encode information desired at the stronger receivers $2, \ldots, K$ through the lowest superposition layer without affecting the decoding performance at the weakest receiver.

1) Lower Bound on Capacity-Memory Tradeoff: Let $(U_1^\star, \ldots, U_{K-1}^\star, X^\star)$ be a random K-tuple that achieves the symmetric-capacity $C_{\mathcal{K}}$, i.e., is a solution to the optimization problem in (14). Define

$$M_1^{\text{single}} := \frac{N}{K-1} \big(I(U_1^\star; Y_2) - I(U_1^\star; Y_1) \big). \tag{25}$$

Theorem 2: Under the following cache assignment

$$M_1 = M_1^{\text{single}} \tag{26a}$$
$$M_2 = \ldots = M_K = 0, \tag{26b}$$

we have

$$C(M_1, \ldots, M_K) \geq C_{\mathcal{K}} + \frac{M_1}{N}. \tag{27}$$

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

$$0 \leq M_1 \leq M_1^{\text{single}}, \tag{28}$$
$$M_2, \ldots, M_K \geq 0. \tag{29}$$

We will see in Corollary 6 ahead that (27) holds with equality for all $0 \leq \mathsf{M}_1 \leq \mathsf{M}_1^{\text{single}}$ provided that $\mathsf{M}_2 = \ldots = \mathsf{M}_\mathsf{K} = 0$.

The RHS of (27) 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.

2) Coding Scheme: Let $(U_1^\star, \ldots, U_{\mathsf{K}-1}^\star, X^\star)$ be a solution to the optimization problem in (14) so that the following inequality is strict:

$$I(U_1^\star; Y_1) < I(U_1^\star; Y_k), \quad k \in \{2, \ldots, \mathsf{K}\}. \tag{30}$$

(If no such choice exists, Theorem 2 reduces to $\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_\mathsf{K}) \geq \mathsf{C}_\mathcal{K}$ and is trivial.)

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

$$R^{(A)} := \mathsf{C}_\mathcal{K} - \epsilon, \tag{31a}$$
$$R^{(B)} := \frac{1}{\mathsf{K}-1}\big(I(U_1^\star; Y_2) - I(U_1^\star; Y_1)\big). \tag{31b}$$

The RHS of (31b) is positive by (30). Split each message W_d, $d \in \{1, \ldots, \mathsf{N}\}$, into two parts:

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

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

$$R = R^{(A)} + R^{(B)}. \tag{32}$$

Placement Phase: Store $W_1^{(B)}, \ldots, W_\mathsf{N}^{(B)}$ in the cache memory of receiver 1. This is possible by (26a) and the definition of $\mathsf{M}_1^{\text{single}}$ in (28).

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

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

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 $(\lfloor 2^{nR^{(B)}} \rfloor)^{\mathsf{K}-1}$ rows. The columns are used to encode message $W_{d_1}^{(A)}$ and the rows to encode the message tuple

$$\mathbf{W}^{(B)} := \big(W_{d_2}^{(B)}, \ldots, W_{d_{\mathsf{K}-1}}^{(B)}, W_{d_\mathsf{K}}^{(B)}\big). \tag{33}$$

The k-th level satellite is used to encode message $W_{d_k}^{(A)}$, for $k \in \{2, \ldots, \mathsf{K}\}$. 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 $x^n(w_{1,\text{column}}, w_{1,\text{row}}; w_2; w_3; \ldots; w_\mathsf{K})$ denote the level-K satellite codeword of $\mathcal{C}$ that corresponds to the cloud center codeword $u_1^n(w_{1,\text{column}}, w_{1,\text{row}})$ 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

$$x^n\big(W_{d_1}^{(A)}, \mathbf{W}^{(B)}; \ W_{d_2}^{(A)}; \ W_{d_3}^{(A)}; \ \ldots \ ; W_{d_\mathsf{K}}^{(A)}\big)$$

over the channel.

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

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

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

$$\mathcal{C}'(\mathbf{W}^{(B)}) := \big\{u_1^n\big(w, \mathbf{W}^{(B)}\big)\big\}_{w=1}^{\lfloor 2^{nR^{(A)}} \rfloor}. \tag{34}$$

Figure 3 illustrates such a subcodebook in red.

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

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

$$kR^{(A)} + (\mathsf{K}-1) \cdot R^{(B)} < I(U_k^\star; Y_k), \tag{35a}$$
$$(k-\ell) \cdot R^{(A)} < I(U_k^\star; Y_k|U_\ell^\star), \ \ell \in \{1, \ldots, k-1\}, \tag{35b}$$

where we define $U_K^\star = X^\star$. One can verify that for degraded BCs the choice of $R^{(A)}$ and $R^{(B)}$ in (31) satisfies the constraints in (35). In fact, by (31) and the degradedness of the BC, we have for all $j \in \{2, \ldots, k\}$

$$R^{(A)} \leq I(U_j^\star; Y_j|U_{j-1}^\star) - \epsilon < I(U_j^\star; Y_k|U_{j-1}^\star) - \epsilon \tag{36}$$
$$R^{(A)} + (\mathsf{K}-1) \cdot R^{(B)} \leq I(U_1^\star; Y_2) - \epsilon < I(U_1^\star; Y_k). \tag{37}$$

Summing (37) over all $j \in \{2, \ldots, k\}$, adding (36), and using the chain rule of mutual information and the Markov chain $U_1^\star \to U_2^\star \to \cdots \to U_K^\star$, establishes (35a). Similarly, summing up (36) for $j \in \{\ell, \ldots, k\}$ establishes (35b).

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

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

Letting $\epsilon \to 0$ proves achievability of Theorem 2.

C. Generalized Coded-Caching

We generalize the coded-caching scheme of [1] to degraded BCs and to unequal cache sizes. In [1], the authors have proposed a scheme for error free channels, parametrized by an integer $t \in [1 : \mathsf{K}-1]$, that can simultaneously communicate to groups of $t+1$ users and hence offer global caching gains. In (noisy) broadcast channels, users have different channel statistics. The main idea in this section is to assign a

larger cache memory at weaker receivers to balance the worse channel condition and to send generalized XOR-messages to groups of $t+1$ receivers at a time. Our cache assignment and delivery scheme are designed such that, in the transmission to any group of $t+1$ receivers, each of the involved receivers is served at a rate close to its capacity. This is possible because each receiver has stored all other transmitted messages in its cache memory, and can exploit this knowledge in the decoding. Notice that if separate cache-channel coding was applied, the rate to each receiver was limited by the worst channel capacity.

1) Lower Bound on Capacity-Memory Tradeoff: We will need the following definitions. Let for each $t \in \mathcal{K}$

$$\mathcal{G}_1^{(t)}, \ldots, \mathcal{G}_{\binom{\mathsf{K}}{t}}^{(t)} \tag{38a}$$

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

$$\mathcal{G}_\ell^{(t),c} := \mathcal{K} \backslash \mathcal{G}_\ell^{(t)}, \qquad \ell \in \left\{1, \ldots, \binom{\mathsf{K}}{t}\right\}. \tag{38b}$$

For any given distribution P_X and $t = 1, \ldots, \mathsf{K}-1$, define the cache sizes and rates:

$$\mathsf{M}_k^{(t)} = \mathsf{N} \cdot \frac{\sum_{\{\ell: \ k \in \mathcal{G}_\ell^{(t)}\}} \prod_{k' \in \mathcal{G}_\ell^{(t),c}} I(X; Y_{k'})}{\sum_{j=1}^{\binom{\mathsf{K}}{t+1}} \prod_{k' \in \mathcal{G}_j^{(t+1),c}} I(X; Y_{k'})}, \tag{39}$$

$$R^{(t)} = \frac{\sum_{\ell=1}^{\binom{\mathsf{K}}{t}} \prod_{k' \in \mathcal{G}_\ell^{(t),c}} I(X; Y_k)}{\sum_{j=1}^{\binom{\mathsf{K}}{t+1}} \prod_{k' \in \mathcal{G}_j^{(t+1),c}} I(X; Y_k)}. \tag{40}$$

Note that when $t = \mathsf{K}-1$ the denominators of (39) and (40) are equal to 1, and hence we have

$$\mathsf{M}_k^{(\mathsf{K}-1)} = N \sum_{\ell \in \mathcal{K} \backslash \{k\}} I(X; Y_\ell), \tag{41}$$

$$R^{(\mathsf{K}-1)} = \sum_\ell I(X; Y_\ell). \tag{42}$$

Observe that for any given P_X, we have

$$\mathsf{M}_1^{(t)} \geq \mathsf{M}_2^{(t)} \geq \cdots \geq \mathsf{M}_\mathsf{K}^{(t)}, \qquad t \in \{1, \ldots, \mathsf{K}-1\}, \tag{43}$$

so a larger cache memory is assigned the weaker a receiver is. The choices of $\mathsf{M}^{(t)}$ in (39) and $R^{(t)}$ in (40) become clear in the description of the coding scheme. In particular, these choices ensure decodability of (sub-)messages in different phases of the coding scheme.

Theorem 3: Fix a $t \in \{1, \ldots, \mathsf{K}-1\}$ and an input distribution P_X, and consider the corresponding cache assignment in (39). Then,

$$\mathsf{C}\left(\mathsf{M}_1^{(t)}, \ldots, \mathsf{M}_\mathsf{K}^{(t)}\right) \geq R^{(t)}, \tag{44}$$

where $M_1^{(t)}, \ldots, M_\mathsf{K}^{(t)}$ and $R^{(t)}$ are calculated from P_X as described in (39) and (40).

As we will see in Corollary 7, the inequality in (44) holds with equality for $t = \mathsf{K}-1$.

We first explain the scheme for the special case of two users.

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

$$R^{(\mathrm{A})} = I(X; Y_1) - \epsilon \tag{45}$$
$$R^{(\mathrm{B})} = I(X; Y_2) - \epsilon. \tag{46}$$

Notice that by the degradedness of the BC:

$$R^{(\mathrm{B})} \geq R^{(\mathrm{A})}. \tag{47}$$

Fix a blocklength n and generate a random codebook

$$\mathcal{C} := \left\{x^n(j)\right\}_{j=1}^{\lfloor 2^{nR^{(\mathrm{B})}} \rfloor} \tag{48}$$

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

$$\mathsf{M}_1 = \mathsf{N} \cdot R^{(\mathrm{B})} = \mathsf{N} \cdot (I(X; Y_2) - \epsilon), \tag{49a}$$
$$\mathsf{M}_2 = \mathsf{N} \cdot R^{(\mathrm{A})} = \mathsf{N} \cdot (I(X; Y_1) - \epsilon), \tag{49b}$$

to receivers 1 and 2, respectively.

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

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

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

In the caching phase, the transmitter stores messages

$$W_1^{(\mathrm{B})}, \ldots, W_\mathsf{N}^{(\mathrm{B})}$$

in receiver 1's cache memory and messages

$$W_1^{(\mathrm{A})}, \ldots, W_\mathsf{N}^{(\mathrm{A})}$$

in receiver 2's cache memory. This is possible given the cache assignment in (49).

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

$$W_{d_1}^{(\mathrm{A})} \overline{\oplus} W_{d_2}^{(\mathrm{B})} \tag{50}$$

to both receivers using the codeword

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

To estimate $W_{d_2}^{(\mathrm{B})}$, Receiver 2 decodes the XOR-message and XORs the decoded message with $W_{d_1}^{(\mathrm{A})}$, which it has stored in its cache memory. The remaining part of its desired message, $W_{d_2}^{(\mathrm{A})}$, is retrieved 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 $W_{d_2}^{(\mathrm{B})}$ from its cache memory, and extracts a subcodebook $\mathcal{C}'(W_{d_2}^{(\mathrm{B})}) \subseteq \mathcal{C}$ containing all codewords that are compatible with $W_{d_2}^{(\mathrm{B})}$:

$$\mathcal{C}'(W_{d_2}^{(\mathrm{B})}) := \left\{x^n(w \overline{\oplus} W_{d_2}^{(\mathrm{B})})\right\}_{w=1}^{\lfloor 2^{nR^{(\mathrm{A})}} \rfloor} \tag{51}$$

---

[5]Recall that in Section I-C we defined the generalized XOR operation $\overline{\oplus}$ as the component-wise XOR over the binary representations of two messages zero-padded to the length of the longer one.

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 $\mathcal{C}$.

To estimate $W_{d_1}^{(A)}$, Receiver 1 decodes the XOR message in (50) using an optimal decoding rule for this subcodebook $\mathcal{C}'(W_{d_2}^{(B)})$, and XORs the decoded message with $W_{d_2}^{(B)}$, which it has stored in its cache memory. The remaining part of its desired message, $W_{d_1}^{(B)}$, is retrieved from its cache memory.

With this scheme, both receivers correctly recover their desired messages W_{d_1} and W_{d_2} whenever they successfully decode the XOR-message in (50). Since the rate $R^{(B)}$ of the original codebook $\mathcal{C}$ satisfies

$$R^{(B)} < I(X; Y_2), \tag{52}$$

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

$$R^{(A)} < I(X; Y_1), \tag{53}$$

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 $\mathsf{K} = 2$ 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.

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

$$\mathsf{M}_1 \geq \mathsf{M}_2. \tag{54}$$

The described scheme can also be applied with a uniform cache assignment $\mathsf{M}_1 = \mathsf{M}_2 = \mathsf{N} \cdot R^{(A)}$, 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.

3) General Coding Scheme: Fix a positive integer $t \in \{1, \ldots, \mathsf{K} - 1\}$. This parameter is an indicator of the number of receivers that cache each part of a file. Similar to the work of [1], this caching scheme ensures that $t + 1$ receivers can be simultaneously served in each transmission during the delivery phase. Pick a small number $\epsilon > 0$ and an input distribution P_X. Consider the cache assignment in (39), where mutual informations are calculated with respect to P_X.

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

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

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

$$R_{\mathcal{G}_\ell^{(t)}} := \frac{\prod_{k \in \mathcal{G}_\ell^{(t), c}} I(X; Y_k)}{\sum_{j=1}^{\binom{\mathsf{K}}{t+1}} \prod_{k \in \mathcal{G}_j^{(t+1), c}} I(X; Y_k)} - \epsilon. \tag{55}$$

The total message rate is thus

$$R = \sum_{\ell=1}^{\binom{\mathsf{K}}{t}} R_{\mathcal{G}_\ell^{(t)}} = \frac{\sum_{\ell=1}^{\binom{\mathsf{K}}{t}} \prod_{k \in \mathcal{G}_\ell^{(t), c}} I(X; Y_k)}{\sum_{j=1}^{\binom{\mathsf{K}}{t+1}} \prod_{k \in \mathcal{G}_j^{(t+1), c}} I(X; Y_k)} - \binom{\mathsf{K}}{t} \epsilon. \tag{56}$$

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

$$\left\{ W_{d, \mathcal{G}_\ell^{(t)}}: \ k \in \mathcal{G}_\ell^{(t)} \right\} \tag{57}$$

in the cache memory of receiver $k \in \mathcal{K}$. This is possible by (55) and the cache assignment in (39).

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

We define the subphase $j \in \left\{ 1, \ldots, \binom{\mathsf{K}}{t+1} \right\}$ to be of length

$$n_j := \left\lfloor n \cdot \frac{\prod_{k \in \mathcal{G}_j^{(t+1), c}} I(X; Y_k)}{\sum_{j'=1}^{\binom{\mathsf{K}}{t+1}} \prod_{k \in \mathcal{G}_{j'}^{(t+1), c}} I(X; Y_k)} \right\rfloor, \tag{58}$$

and to transmit messages

$$\left\{ W_{d_k, \mathcal{G}_j^{(t+1)} \setminus \{k\}} \right\}_{k \in \mathcal{G}_j^{(t+1)}} \tag{59}$$

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

$$W_{\mathrm{XOR}, \mathcal{G}_j^{(t+1)}} = \overline{\bigoplus}_{k \in \mathcal{G}_j^{(t+1)}} W_{d_k, \mathcal{G}_{j+1}^{(t)} \setminus \{k\}}, \tag{60}$$

which is of rate

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

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

and generates a codebook

$$\mathcal{C}_j = \left\{ x_j^{n_j}(w): \ w = 1, \ldots, \left\lfloor 2^{n R_{\mathrm{XOR}, \mathcal{G}_j^{(t+1)}}} \right\rfloor \right\}, \tag{61}$$

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

The transmitter then sends the codeword

$$x_j^{n_j}\left(W_{\mathrm{XOR}, \mathcal{G}_j^{(t+1)}} \right) \tag{62}$$

over the channel.

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

$$\left\{ W_{d_k, \mathcal{G}_\ell^{(t)}}: \ k \in \mathcal{G}_\ell^{(t)} \right\} \tag{63}$$

directly from its cache, see (57), and thus only needs to decode messages

$$\left\{ W_{d_k, \mathcal{G}_\ell^{(t)}}: \ k \notin \mathcal{G}_\ell^{(t)} \right\}. \tag{64}$$

For each $j \in \left\{ 1, \ldots, \binom{\mathsf{K}}{t+1} \right\}$ and $k \in \mathcal{G}_j^{(t+1)}$, receiver k decodes message $W_{d_k, \mathcal{G}_j^{(t+1)} \setminus \{k\}}$ 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

$$W_{\mathrm{XOR}, j, k} := \overline{\bigoplus}_{k' \in \mathcal{G}_j^{(t+1)} \setminus \{k\}} W_{d_{k'}, \mathcal{G}_j^{(t+1)} \setminus \{k'\}}, \tag{65}$$

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

$$\mathcal{C}'_{j,k}(W_{\mathrm{XOR},j,k}) := \left\{ x_j^{n_j}\left(w \,\overline{\oplus}\, W_{\mathrm{XOR},j,k}\right) \right\}_{w=1}^{\left\lfloor 2^{nR_{\mathcal{G}_j^{(t+1)}\backslash\{k\}}} \right\rfloor}.$$

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

$$\hat{W}_{d_k,\mathcal{G}_j^{(t+1)}\backslash\{k\}} = \hat{W}_{\mathrm{XOR},\mathcal{G}_j^{(t+1)}} \,\overline{\oplus}\, W_{\mathrm{XOR},j,k}. \tag{66}$$

After the last sub-phase $\binom{\mathsf{K}}{t+1}$, each receiver $k \in \mathcal{K}$ has decoded all its missing messages in (64), and can thus recover W_{d_k}.

Error Analysis: If each XOR-message $W_{\mathrm{XOR},\mathcal{G}_j^{(t+1)}}$ is decoded correctly by all its intended receivers in $\mathcal{G}_j^{(t+1)}$, $j = 1,\ldots,\binom{\mathsf{K}}{t+1}$, then all receivers $1,\ldots,\mathsf{K}$ recover their desired messages $W_{d_1},\ldots,W_{d_\mathsf{K}}$.

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

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

see (55) and (58). Letting then $\epsilon \to 0$, establishes Theorem 3.

V. Upper Bounds and Exact Results on the Capacity-Memory Tradeoff

We present a general upper bound on the capacity-memory tradeoff. We further show that the upper bound matches the lower bounds derived in the previous section in certain regimes of cache sizes. In these regimes, we can thus characterize the exact capacity-memory tradeoff.

A. Upper Bounds

Our upper bound is formulated in terms of the following parameters. Consider any set $\mathcal{S} \subseteq \mathcal{K}$ and represent it by $\mathcal{S} = \{j_1,\ldots,j_{|\mathcal{S}|}\}$ as in (18). Define

$$\alpha^\star_{\mathcal{S},1} := \frac{\mathsf{M}_{j_1}}{\mathsf{N}} \tag{67a}$$

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

$$\alpha^\star_{\mathcal{S},k} := \min\left\{ \frac{\sum_{i=1}^k \mathsf{M}_{j_i}}{\mathsf{N}-k+1}, \right.$$
$$\left. \frac{1}{|\mathcal{S}|-k+1}\left(\frac{|\mathcal{S}|}{\mathsf{N}}\sum_{i=1}^{|\mathcal{S}|}\mathsf{M}_{j_i} - \sum_{i=1}^{k-1}\alpha^\star_{\mathcal{S},i} \right) \right\}. \tag{67b}$$

Theorem 4: There exist random variables $X,Y_1,\ldots,Y_\mathsf{K}$ and for *every* receiver set $\mathcal{S}$ as in (18) random variables $\{U_{\mathcal{S},1},\ldots,U_{\mathcal{S},|\mathcal{S}|-1}\}$ so that the channel law (15b) and the following Markov chain hold:

$$U_{\mathcal{S},1} \to U_{\mathcal{S},2} \to \cdots \to U_{\mathcal{S},|\mathcal{S}|-1} \to X \to (Y_1,\ldots,Y_\mathsf{K}), \tag{68}$$

and so that for each $\mathcal{S}$ we have

$$\mathsf{C}(\mathsf{M}_1,\ldots,\mathsf{M}_\mathsf{K}) \leq I(U_{\mathcal{S},1};Y_{j_1}) + \alpha^\star_{\mathcal{S},1}, \tag{69a}$$
$$\mathsf{C}(\mathsf{M}_1,\ldots,\mathsf{M}_\mathsf{K}) \leq I(U_{\mathcal{S},k};Y_{j_k}|U_{\mathcal{S},k-1}) + \alpha^\star_{\mathcal{S},k},$$
$$\forall k \in \{2,\ldots,|\mathcal{S}|-1\}, \tag{69b}$$
$$\mathsf{C}(\mathsf{M}_1,\ldots,\mathsf{M}_\mathsf{K}) \leq I(X;Y_{j_{|\mathcal{S}|}}|U_{\mathcal{S},|\mathcal{S}|-1}) + \alpha^\star_{\mathcal{S},|\mathcal{S}|}. \tag{69c}$$

Proof: See Appendix A. ■

Without cache memories, $\mathsf{M}_1 = \ldots = \mathsf{M}_\mathsf{K} = 0$, the parameters $\alpha^\star_{\mathcal{S},1},\ldots,\alpha^\star_{\mathcal{S},|\mathcal{S}|}$ equal 0 for all $\mathcal{S} \subseteq \{1,\ldots,\mathsf{K}\}$, and hence the upper bound in Theorem 4 recovers the exact capacity-memory tradeoff $\mathsf{C}_\mathcal{K}$ in (14).

The upper bound in Theorem 4 is asymmetric in the different cache sizes $\mathsf{M}_1,\mathsf{M}_2,\ldots,\mathsf{M}_\mathsf{K}$, 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 upper bound in Theorem 4 is weakened if the constraints in (69) 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 $\mathcal{S}$. Using the latter relaxation, Theorem 4 results in the following corollary.

Corollary 5: Given cache sizes $\mathsf{M}_1,\ldots,\mathsf{M}_\mathsf{K} \geq 0$, rate $R \geq 0$, is achievable only if for every receiver set $\mathcal{S} \subseteq \mathcal{K}$:

$$\left((R-\alpha^\star_{\mathcal{S},1})^+, (R-\alpha^\star_{\mathcal{S},2})^+, \ldots, (R-\alpha^\star_{\mathcal{S},|\mathcal{S}|})^+ \right) \in \mathcal{C}_\mathcal{S}, \tag{70}$$

where $\mathcal{C}_\mathcal{S}$ denotes the no-cache capacity region to the receivers in $\mathcal{S}$ (assuming no cache memories at the receivers and ignoring all receivers in $\mathcal{K}\backslash\mathcal{S}$).

Proof: If R is achievable, then Theorem 4 ensures that for every set $\mathcal{S} = \{j_1,\ldots,j_{|\mathcal{S}|}\}$, there exists a set of random variables $(U_1,U_2,\ldots,U_{|\mathcal{S}|-1},X,Y_{j_1},\ldots,Y_{j_{|\mathcal{S}|}})$ that satisfy (19)-(20) and hence (70) holds for every set $\mathcal{S}$. Note that this does not necessarily hold in the reverse direction, meaning that if (70) holds for every $\mathcal{S}$, it is not clear if the conditions of Theorem 4 are satisfied in general. This is because the choice of random variable X that is found from (70) (for every $\mathcal{S}$), may implicitly depend on $\mathcal{S}$ while this is not permissible in condition (68) in Theorem 4. ■

Remark 3: The upper bounds of Theorem 4 and Corollary 5 are relaxed when each $\alpha^\star_{\mathcal{S},k}$ is replaced by $\tilde{\alpha}_{\mathcal{S},k}$, defined below:

$$\tilde{\alpha}_{\mathcal{S},1} := \frac{\mathsf{M}_{j_1}}{\mathsf{N}}, \tag{71a}$$
$$\tilde{\alpha}_{\mathcal{S},k} := \frac{|\mathcal{S}|\cdot\sum_{i=1}^{|\mathcal{S}|}\mathsf{M}_{j_i} - \mathsf{M}_{j_1}}{(|\mathcal{S}|-1)\mathsf{N}}, \quad k \in \{2\ldots,|\mathcal{S}|\}. \tag{71b}$$

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

$$\alpha'_{\mathcal{S},k} := \frac{\sum_{i=1}^{|\mathcal{S}|}\mathsf{M}_{j_i}}{\mathsf{N}}. \tag{72}$$

In particular, Corollary 5 recovers the previous upper bound in [22, Theorem 9] and [35, Theorem 1] if $\alpha^\star_{\mathcal{S},k}$ is replaced by $\alpha'_{\mathcal{S},k}$:

Proof: The proof requires a close inspection of the proof of Theorem 4 in Appendix A. See Appendix D. ■

B. Exact Results

By comparing the new upper bounds with the three lower bounds presented in Section IV, the exact expression for $C(M_1, \ldots, M_K)$ can be obtained in some special cases. For example, as the following corollary states, the lower bound achieved by superposition piggyback coding matches the upper bound when only receiver 1 has a cache memory and this cache memory is small.

Corollary 6: Under a cache assignment satisfying

$$0 \leq M_1 \leq M_1^{\text{single}} \quad \text{and} \quad M_2 = \ldots = M_K = 0, \quad (73)$$

the capacity-memory tradeoff is

$$C(M_1, 0, \ldots, 0) = C_{\mathcal{K}} + \frac{M_1}{N}.$$

Proof: Achievability follows by Theorem 2 (see also Remark 2) and the converse follows by Corollary 5, where it suffices to consider only the set $\mathcal{S} = \mathcal{K}$. In fact, under (73), $\alpha_{\mathcal{K},1}^{\star} = \ldots = \alpha_{\mathcal{K},K}^{\star} = \frac{M_1}{N}$. ∎

The next corollary states that the lower bound attained by generalized coded caching with parameter $t = K - 1$ matches the upper bound under the corresponding cache assignment in (41). Moreover, any extra cache memory that is uniformly distributed over the K receivers only brings local caching gain.

Proposition 7: For each $k \in \mathcal{K}$, let $M_k^{\star(K-1)}$ be given by (41) when P_X is chosen as a maximizer of

$$C_{\text{Avg}} := \frac{1}{K} \cdot \max_{P_X} \left(\sum_{k=1}^{K} I(X; Y_k) \right). \quad (74)$$

For any $\Delta \geq 0$, we have

$$C\left(M_1^{\star(K-1)} + \Delta, \ \ldots, \ M_K^{\star(K-1)} + \Delta \right)$$
$$= C_{\text{Avg}} + \frac{\sum_{k=1}^{K} M_k^{\star(K-1)}}{K \cdot N} + \frac{\Delta}{N}. \quad (75)$$

Proof: See Appendix E. ∎

VI. BOUNDS ON THE GLOBAL CAPACITY-MEMORY TRADEOFF $C^{\star}(M)$

The two preceding sections presented lower and upper bounds on the capacity-memory tradeoff for given cache assignments. In this section, we assume that a system designer is given a total cache budget $M \geq 0$ that it can arbitrarily distribute across users. We are thus interested in the largest capacity-memory tradeoff optimized over the cache assignment $M_1, \ldots, M_K$ subject to a total cache budget $M_1 + M_2 + \ldots + M_K \leq M$. We introduced this quantity as the global capacity-memory tradeoff $C^{\star}(M)$ in (11). This section presents lower and upper bounds on $C^{\star}(M)$, for any value of M, as well as exact results for $C^{\star}(M)$ when M is below a certain threshold or above another threshold.

A. Lower Bound

Proposition 1 and Theorems 2 and 3 readily yield a lower bound on $C^{\star}(M)$, see (77). As we will see in Corollary 11 ahead, this lower bound holds with equality when the total cache size M is small or large.

Let

$$R^{(0)} := C_{\mathcal{K}}, \qquad M^{(0)} := 0, \quad (76a)$$

and

$$R^{\text{single}} := C_{\mathcal{K}} + \frac{M^{\text{single}}}{N}, \qquad M^{\text{single}} := M_1^{\text{single}}, \quad (76b)$$

where $C_{\mathcal{K}}$ is defined in (13) and M_1^{single} is defined in (25). Also, for any given P_X, recall $M^{(t)}$ and $R^{(t)}$ from (39) and (40), and define for $t \in \{1, \ldots, K-1\}$:

$$M^{(t)} := \sum_{k=1}^{K} M_k^{(t)}. \quad (76c)$$

We have proved the achievability of each pair by proposing a corresponding scheme. Consider two schemes achieving the memory-rate pairs $(M^{(t)}, R^{(t)})$ and $(M^{(t')}, R^{(t')})$. By time sharing between the two schemes, we can achieve all memory-rate pairs that lie on the line connecting $(M^{(t)}, R^{(t)})$ and $(M^{(t')}, R^{(t')})$. I.e., the upper-convex envelope of all these rate-memory pairs thus lower bounds $C^{\star}(M)$. This is formulated in Corollary 8 below.

Corollary 8: The global capacity-memory tradeoff is lower bounded by:

$$C^{\star}(M) \geq \text{upp hull}\Big\{ \left(R^{(0)}, M^{(0)}\right), \left(R^{\text{single}}, M^{\text{single}}\right),$$
$$\bigcup_{P_X} \Big\{ \left(R^{(1)}, M^{(1)}\right), \ldots, \left(R^{(K-1)}, M^{(K-1)}\right) \Big\} \Big\}. \quad (77)$$

Notice that for any P_X:

$$M^{(0)} \leq M^{\text{single}} \leq M^{(1)} \leq \cdots \leq M^{(K-1)} \quad (78)$$

and

$$R^{(0)} \leq R^{\text{single}} \leq R^{(1)} \leq \cdots \leq R^{(K-1)}. \quad (79)$$

B. Upper Bounds

Theorem 4 directly yields the following upper bound on the global capacity-memory tradeoff.

Corollary 9: There exist random variables $X, Y_1, \ldots, Y_K$ and for *every* receiver set $\mathcal{S}$ as in (18) random variables $\{U_{\mathcal{S},1}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1}\}$, such that (15b) and (68) hold, and such that *for some* $M_1, \ldots, M_K \geq 0$ summing to M and all $\mathcal{S}$:

$$C^{\star}(M) \leq I\left(U_{\mathcal{S},1}; Y_{j_1}\right) + \alpha_{\mathcal{S},1}^{\star}, \quad (80a)$$
$$C^{\star}(M) \leq I\left(U_{\mathcal{S},k}; Y_{j_k} | U_{\mathcal{S},k-1}\right) + \alpha_{\mathcal{S},k}^{\star}, \quad k \in \{2, \ldots, |\mathcal{S}|-1\}, \quad (80b)$$
$$C^{\star}(M) \leq I\left(X; Y_{j_{|\mathcal{S}|}} | U_{\mathcal{S},|\mathcal{S}|-1}\right) + \alpha_{\mathcal{S},|\mathcal{S}|}^{\star}, \quad (80c)$$

where $\{\alpha_{\mathcal{S},k}^{\star}\}$ are defined in (67). Evaluating this bound numerically is cumbersome because for each possible subset $\mathcal{S}$ the coefficients $\alpha_{\mathcal{S},1}^{\star}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}^{\star}$ have to be computed and then the optimal choice of $U_{\mathcal{S},1}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1}, X$ needs to be found in order to find the loosest upper bound in (80). To find upper bounds that have easier close-form expressions, we loosen the bound by either relaxing some of the constraints in (80); by replacing each parameter $\alpha_{\mathcal{S},k}^{\star}$ in (80) by $\tilde{\alpha}_{\mathcal{S},k}$ or by $\alpha_{\mathcal{S},k}'$ (defined in (71) and (72)); and/or by allowing $X, Y_{j_1}, \ldots, Y_{j_{\mathcal{S}}}$ in (80) to depend

on the set $\mathcal{S}$. The following corollary presents such a simpler upper bound. Recall the definitions in (38).

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

$$C^\star(M) \leq \frac{1}{\binom{K}{t}} \sum_{\ell=1}^{\binom{K}{t}} C_{\mathcal{G}_\ell^{(t)}} + \frac{t}{K} \cdot \frac{M}{N}. \tag{81}$$

Proof: Fix $t \in \mathcal{K}$. For each $\ell = 1, \ldots, \binom{K}{t}$, specialize Corollary 5 to $\mathcal{S} = \mathcal{G}_\ell^{(t)}$ and relax it by replacing each parameter $\alpha^\star_{\mathcal{G}_\ell^{(t)},k}$ by $\alpha'_{\mathcal{G}_\ell^{(t)},k}$ defined in (72). Since $\alpha'_{\mathcal{G}_\ell^{(t)},1} = \ldots = \alpha'_{\mathcal{G}_\ell^{(t)},t}$, we obtain

$$C^\star(M) \leq C_{\mathcal{G}_\ell^{(t)}} + \alpha'_{\mathcal{G}_\ell^{(t)},1} = C_{\mathcal{G}_\ell^{(t)}} + \frac{\sum_{i \in \mathcal{G}_\ell^{(t)}} M_i}{N}. \tag{82}$$

Now, averaging (82) over all indices $\ell = 1, \ldots, \binom{K}{t}$ and upper-bounding the sum $M_1 + \ldots + M_K$ by M yields the desired result. ∎

C. Exact Results

Lower and upper bounds on $C^\star(M)$ presented above match for small and large total cache sizes M. Corollary 11 below states this more formally. Recall M^{single} from (25) and define M^L as follows:

$$M^L = N \cdot (K - 1) \cdot K \cdot C_{\text{avg}} \tag{83}$$

where C_{avg} is defined in (74).

Corollary 11: For any positive total cache size $M \leq M_{\text{single}}$:

$$C^\star(M) = C_{\mathcal{K}} + \frac{M}{N}, \qquad 0 \leq M \leq M^{\text{single}}, \tag{84}$$

and for any $M \geq M^L$:

$$C^\star(M) = K C_{\text{avg}} + \frac{1}{K} \cdot \frac{M - M^L}{N}, \quad M \geq M^L. \tag{85}$$

Before presenting the proof of this Corollary, let us briefly discuss the implications: For small cache sizes, the entire cache memory should be assigned to the weakest receiver and the superposition piggyback coding scheme of Section IV-B is optimal. For large total cache sizes M, a careful assignment of the available cache memory is needed. In particular, for $M = M^L$, the generalized coded caching of Section IV-C (and its corresponding cache assignment) is optimal.

Remark 4: For small total cache sizes, $C^\star(M)$ grows as $\frac{M}{N}$. This corresponds to a perfect global caching gain as if each receiver could access all cache contents in the network locally. For large total cache sizes, the global benefit of receivers' cache memories is fully exploited. Any additional cache budget exceeding M^L should be distributed among the receivers uniformly and it only offers a local caching gain. In particular, $C^\star(M)$ grows as $\frac{1}{K} \cdot \frac{M}{N}$. This is similar to the insights from [17] (for rate-limited links). For moderate cache sizes, $C^\star(M)$ grows with $\frac{M}{N}$ at a slope equal to:

$$\frac{R^{(t+1)} - R^{(t)}}{\frac{M^{(t+1)}}{N} - \frac{M^{(t)}}{N}} = \frac{1}{t+1} - \frac{M^{(t)}}{t(t+1)\left(M^{(t+1)} - M^{(t)}\right)},$$

for $t = 1, \ldots, K - 1$.

Proof of Corollary 11: The global capacity-memory tradeoff $C^\star(M)$ is upper bounded by the right-hand side of (84) and this follows by specializing Corollary 10 to $t = K$. Equality in (84) for $M \leq M^{\text{single}}$ follows by Theorem 2.

The capacity memory tradeoff $C^\star(M)$ is also upper bounded by the right-hand side of (85). To see this, relax Corollary 9 by (i) replacing each parameter $\alpha^\star_{\mathcal{S},k}$ with $\alpha'_{\mathcal{S},k}$ and (ii) considering only the constraints (80) that correspond to sets $\mathcal{S} = \{k\}$, for $k \in \mathcal{K}$. Next, average the K resulting inequalities and maximize over the input distribution P_X. This yields:

$$C^\star(M) \leq C_{\text{avg}} + \frac{1}{K} \frac{M}{N}. \tag{86}$$

Using the definition of M^L as given in (83), we can equivalently write the RHS of (86) in the form of $K C_{\text{avg}} + \frac{1}{K} \frac{M - M^L}{N}$. Equality in (85) for $M \geq N(K - 1) K C_{\text{Avg}}$ follows from Proposition 7 with $\Delta = 0$. Recall that $M_k^{(K-1)}$ is given in (41) as the sum of mutual information terms $I(X; Y_\ell)$ for all $\ell \in \mathcal{K} \setminus \{k\}$. So we have

$$M_1^{\star(K-1)} + \ldots + M_K^{\star(K-1)} = N \sum_{i=1}^{K} (K - 1) I(X; Y_i) \tag{87}$$

$$= N(K - 1) K C_{\text{Avg}} \tag{88}$$

where in the first step we have noted that each mutual information term appears $(K - 1)$ times in the sum on the LHS of (87). ∎

VII. Examples

A. Erasure BCs

We specialize our results to erasure BCs where at time i receiver k's output $Y_{k,i}$ equals the channel input X_i with probability $1 - \delta_k$ and it equals an erasure symbol "?" with probability δ_k. Without loss of generality, we assume that the erasure probabilities satisfy:

$$1 > \delta_1 \geq \delta_2 \geq \ldots \geq \delta_K \geq 0. \tag{89}$$

For erasure BCs,

$$C_{\mathcal{S}} = \left(\sum_{s \in \mathcal{S}} \frac{1}{1 - \delta_s} \right)^{-1}, \qquad \mathcal{S} \subseteq \mathcal{K}. \tag{90}$$

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 4 and Corollary 5 coincide. Also,

$$C_{\text{avg}} = \frac{1}{K} \sum_{k=1}^{K} C_k = 1 - \frac{\sum_{k=1}^{K} \delta_k}{K}. \tag{91}$$

Figure 4, depicts the upper and lower bounds on $C^\star(M)$ in Corollaries 8 and 9. For comparison, the upper bound of Theorem 4 is also plotted under the uniform cache assignment:

$$M_1 = \ldots = M_K = \frac{M}{K}.$$

One observes that a smart allocation of the total cache memory M significantly increases the global capacity-memory tradeoff
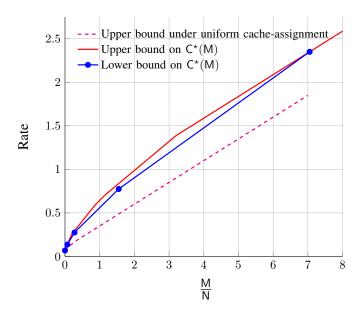
Fig. 4. Bounds on $C^\star(M)$ for a 4-user Erasure BC with $\delta_1 = 0.9$, $\delta_2 = 0.6$, $\delta_3 = 0.1$, $\delta_4 = 0.05$.

of erasure BCs when different receivers have different erasure probabilities.

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

Proposition 12: For a given $M_1 \geq 0$ and $M := \sum_{k=1}^{K} M_k$, $M \geq 0$,

$$C(M_1, \ldots, M_K)$$
$$\leq \min \left\{ C_K + \frac{M_1}{N} + \frac{(M - M_1)}{N} \cdot \frac{K \cdot C_K}{(K-1)C_{\{2,\ldots,K\}}}, \right.$$
$$\left. C_1 + \frac{M_1}{N} \right\}. \quad (92)$$

The RHS of (92) is strictly less than $C_K + \frac{M}{N}$ unless $M = M_1$ or $\delta_1 = \ldots = \delta_K$.

Proof: See Appendix F. ∎

B. Noise-Free Bit-Pipe

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

$$\delta_1 = \ldots = \delta_K = 0. \quad (93)$$

Consider also the notation introduced in Section II-A. We adopt the system model of [1] with equal cache sizes

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

and the *delivery rate* ρ.

From the upper bound on $C(M_1, \ldots, M_K)$ in Theorem 4, 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], the delivery rate is bounded from below as follows.

$$\rho^\star \geq t - m \cdot \min \left\{ \frac{t^2}{N}, \sum_{k=1}^{t} \frac{k}{N - k + 1} \right\}, \quad m \leq N. \quad (94)$$

Proof: See Appendix G. ∎

This lower bound improves on the lower bounds [1], [10]–[12] that existed at the time when this manuscript was submitted. It is within a constant gap of 2.35 from the optimal rate-memory tradeoff [37]. A slightly improved bound has been established in a parallel work [36]. This latter bound, however, is specific to the noise-free bit-pipe model.

C. Gaussian BCs

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

$$Y_{k,t} = X_t + Z_{k,t}, \quad (95)$$

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. Without loss of generality, the receivers are ordered in increasing strength:

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

By [39], for every set $\mathcal{S}$ as defined in (18), we have

$$C_\mathcal{S} = \frac{1}{2} \log_2 \left(1 + \frac{\beta_1 P}{\sum_{k=2}^{|\mathcal{S}|} \beta_k P + \sigma_1^2} \right), \quad (96)$$

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

$$\frac{\beta_1 P}{\sum_{k=2}^{|\mathcal{S}|} \beta_k P + \sigma_1^2} = \frac{\beta_i P}{\sum_{k=i+1}^{|\mathcal{S}|} \beta_k P + \sigma_{j_i}^2}, \quad i \in \{1, \ldots, |S|\}. \quad (97)$$

In particular,

$$C_k = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_k^2} \right), \quad k \in \{1, \ldots, K\}. \quad (98)$$

Moreover, given a power constraint $P > 0$, a zero-mean variance-P Gaussian input distribution maximizes $I(X; Y_k)$ and $I(X; Y_k|U)$ simultaneously for all $k \in \mathcal{K}$ and Gaussian auxiliaries U that form the Markov chain $U - X - Y_k$. Therefore, Theorem 4 and Corollary 5 coincide. Also,

$$C_{avg} = \frac{1}{K} \sum_{k=1}^{K} C_k = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{2} \log \left(1 + \frac{P}{\sigma_k^2} \right). \quad (99)$$

Figure 5 shows the upper and lower bounds on $C^\star(M)$ in Corollaries 8 and 9. The five blue points indicate the rate-memory points $(R^{(0)}, M^{(0)})$, (R^{single}, M^{single}), $(R^{(1)}, M^{(1)})$, $(R^{(2)}, M^{(2)})$, and $(R^{(3)}, M^{(3)})$ for a zero-mean variance-P Gaussian distribution P_X. For comparison, the figure also shows the upper bound in Theorem 4 for a setup with uniform cache assignment $\frac{M}{K}$ across all receivers. We observe that a smart cache assignment provides substantial gains in the capacity-memory tradeoff.
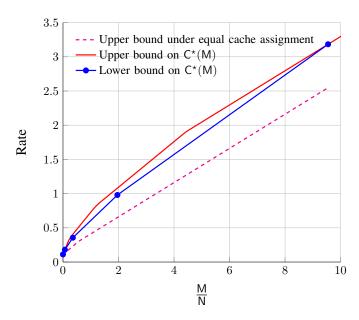
Fig. 5. Upper and lower bounds on $\mathsf{C}^\star(\mathsf{M})$ on a 4-receiver Gaussian BC with input power $P = 1$ and noise variances $\sigma_1 = 4$, $\sigma_2^2 = 1$, $\sigma_3^2 = 0.5$, and $\sigma_4^2 = 0.1$. We have $\mathsf{N} = 10$.

VIII. 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 memories with thresholds depending on the BC statistics.

For small cache memory sizes (characterized in (84)), the global capacity-memory tradeoff is achieved by assigning all the available cache memory to the weakest receiver. In this regime, $\mathsf{C}^\star(\mathsf{M})$ grows as $\frac{\mathsf{M}}{\mathsf{N}}$ which corresponds to a perfect global caching gain; i.e., all receivers can benefit from all the cache contents in the network. This performance is achieved by superposition piggyback coding which provides every receiver virtual access to the weakest receiver's cache content.

For the regime of moderate M, we proposed a generalized coded caching scheme that performs a particular cache assignment such that the weaker receivers are provided with larger cache sizes. It then simultaneously serves $t + 1$, receivers in each delivery, where $t \in \{1, \ldots, \mathsf{K}-1\}$ is the parameter of the scheme (similar to the scheme in [1]). We observed that the larger the total cache budget M, the larger the coded caching parameter t needs to be chosen. Hence, as M increases, cache memories will have to store more overlapping contents, and hence the caching gain decreases. In other words, the slope of the rate-memory tradeoff (achieved by the generalized coded caching) decreases as the total cache budget M increases. The same behavior is also observed from the upper bound.

In the regime of large M (characterized by the threshold in (85)), the caching gain is only local and $\mathsf{C}^\star(\mathsf{M})$ grows as $\frac{\mathsf{M}}{\mathsf{KN}}$. In particular, the memory threshold in (85)) corresponds to the extreme case where $t = \mathsf{K} - 1$. In this case, the generalized coded caching scheme and its corresponding cache assignment are optimal and achieve the global capacity-memory tradeoff. For larger cache memories, it is optimal to first allocate the

total cache memory as proposed by our scheme for $t = \mathsf{K}-1$, and then uniformly allocate all the remaining cache memory across all the receivers (and store the same content in those extra portions of the receivers' cache memories).

By examining several examples, we have demonstrated that assigning the total cache memory uniformly across all the receivers can be highly suboptimal over noisy BCs.

From a practical view point, one of the main technical challenges for implementing the proposed generalized coded caching scheme is its high level of subpacketization, i.e., the fact that messages need to be split into a very large number of smaller parts. Recent efforts focus on schemes with low subpacketization levels. See for example [42]–[44] for results on the noise-free bit-pipe model.

APPENDIX A
PROOF OF UPPER BOUND IN THEOREM 4

Fix the rate of communication

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

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 $\{f^{(n)}(\cdots, \mathbf{d})\}$, and K decoding functions $\{\varphi_k^{(n)}(\cdots, \mathbf{d})\}$ so that the probability of error $\mathsf{P}_\mathrm{e}^{(n)}$ tends to 0 as $n \to \infty$. Recall that $\mathsf{P}_\mathrm{e}^{(n)}$ is the average over all error probabilities $\mathsf{P}_\mathrm{e}^{(n)}(\mathbf{d})$, $\mathbf{d} \in \mathcal{N}^\mathsf{K}$. So $\mathsf{P}_\mathrm{e}^{(n)}$ goes to zero if and only if $\mathsf{P}_\mathrm{e}^{(n)}(\mathbf{d})$ goes to zero for all demand vectors $\mathbf{d}$. We note that this should hold for every demand vector $\mathbf{d}$ and, in particular, for $\mathbf{d}$ with all-different entries. To obtain the upper bound in Theorem 4, we consider only demand vectors $\mathbf{d}$ in which all entries are different.

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

$$V_k = g_k^{(n)}(W_1, \ldots, W_\mathsf{N}), \qquad k \in \{1, \ldots, \mathsf{K}\}, \quad (100)$$

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

$$X_\mathbf{d}^n = f^{(n)}(W_1, \ldots, W_\mathsf{N}, \mathbf{d}) \qquad (101)$$

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 $X_\mathbf{d}, Y_{1,\mathbf{d}}, \ldots, Y_{\mathsf{K},\mathbf{d}}$ and for each set $\mathcal{S}$ as in (18) random variables $\{U_{\mathcal{S},1,\mathbf{d}}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d}}\}$, so that given $X_\mathbf{d} = x \in \mathcal{X}$:

$$(Y_{1,\mathbf{d}}, Y_{2,\mathbf{d}}, \ldots, Y_{\mathsf{K},\mathbf{d}}) \sim \Gamma(\cdots | x); \qquad (102a)$$

and for each $\mathcal{S}$:

$$U_{\mathcal{S},1,\mathbf{d}} \to \cdots \to U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d}} \to X_\mathbf{d} \to Y_{\mathsf{K},\mathbf{d}} \to Y_{\mathsf{K}-1,\mathbf{d}} \cdots \to Y_{1,\mathbf{d}} \qquad (102b)$$

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

$$R - \epsilon \le \frac{1}{n} I\big(W_{d_{j_1}}; V_{j_1}\big) + I\big(U_{\mathcal{S},1,\mathbf{d}}; Y_{j_1,\mathbf{d}}\big), \qquad (103a)$$

$$R - \epsilon \leq \frac{1}{n}I\big(W_{d_{j_k}}; V_{j_1}, \ldots, V_{j_k} | W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}\big)$$
$$+ I\big(U_{\mathcal{S},k,\mathbf{d}}; Y_{j_k,\mathbf{d}} | U_{\mathcal{S},k-1,\mathbf{d}}\big),$$
$$\forall k \in \{2, \ldots, |\mathcal{S}| - 1\},$$
(103b)

$$R - \epsilon \leq \frac{1}{n}I\big(W_{d_{j_{|\mathcal{S}|}}}; V_{j_1}, \ldots, V_{j_{|\mathcal{S}|}} | W_{d_{j_1}}, \ldots, W_{d_{j_{|\mathcal{S}|-1}}}\big)$$
$$+ I\big(X_{\mathbf{d}}; Y_{j_{|\mathcal{S}|},\mathbf{d}} | U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d}}\big). \quad (103c)$$

Proof: The proof is inspired by the converse proof of the capacity of degraded BCs without caching [38]. Details are as follows. Since the worst case error probability is bounded by ϵ, using Fano's inequality we have

$$R - \epsilon \leq \frac{1}{n}I\big(W_{d_{j_1}}; Y_{j_1,\mathbf{d}}^n, V_{j_1}\big)$$
$$= \frac{1}{n}I\big(W_{d_{j_1}}; V_{j_1}\big) + \frac{1}{n}I\big(W_{d_{j_1}}; Y_{j_1,\mathbf{d}}^n | V_{j_1}\big), \quad (104a)$$

where the equality follows by the chain rule of mutual information. Similarly, for $k \in \{2, \ldots, \mathsf{K}\}$:

$$R - \epsilon \overset{(a)}{\leq} \frac{1}{n}I\big(W_{d_{j_k}}; Y_{j_k,\mathbf{d}}^n, V_{j_1}, \ldots, V_{j_k} | W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}\big)$$
$$= \frac{1}{n}I\big(W_{d_{j_k}}; V_{j_1}, \ldots, V_{j_k} | W_{d_1}, \ldots, W_{d_{j_{k-1}}}\big)$$
$$+ \frac{1}{n}I\big(W_{d_{j_k}}; Y_{j_k,\mathbf{d}}^n | V_1, \ldots, V_{j_k}, W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}\big), \quad (104b)$$

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 (104a) and (104b). For the second summand in (104a) we write

$$\frac{1}{n}I\big(W_{d_{j_1}}; Y_{j_1,\mathbf{d}}^n | V_{j_1}\big) = \frac{1}{n}\sum_{t=1}^n I\big(W_{d_{j_1}}; Y_{j_1,\mathbf{d},t} | V_{j_1}, Y_{j_1,\mathbf{d}}^{t-1}\big)$$
$$\overset{(a)}{\leq} \frac{1}{n}\sum_{t=1}^n I\big(W_{d_{j_1}}, Y_{j_1,\mathbf{d}}^{t-1}, V_{j_1}; Y_{j_1,\mathbf{d},t}\big)$$
$$= I\big(U_{\mathcal{S},1,\mathbf{d},T}; Y_{j_1,\mathbf{d},T} | T\big)$$
$$\leq I\big(U_{\mathcal{S},1,\mathbf{d}}; Y_{j_1,\mathbf{d}}\big) \quad (105)$$

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

$$U_{\mathcal{S},1,\mathbf{d},T} := (V_{j_1}, W_{d_{j_1}}, Y_{j_1,\mathbf{d}}^{T-1}), \quad (106)$$
$$U_{\mathcal{S},1,\mathbf{d}} := (U_{\mathcal{S},1,\mathbf{d},T}, T), \quad (107)$$
$$Y_{j_1,\mathbf{d}} := Y_{j_1,\mathbf{d},T}. \quad (108)$$

Step (a) holds because, for any triple of random variables (A, B, C), we have $I(A; B|C) \leq I(A, C; B)$ by the chain rule of mutual information and the nonnegativity of mutual information.

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

$$U_{\mathcal{S},k,\mathbf{d},T} := (U_{\mathcal{S},k-1,\mathbf{d},T}, V_{j_k}, W_{d_{j_k}}, Y_{j_k,\mathbf{d}}^{T-1}), \quad (109)$$
$$U_{\mathcal{S},k,\mathbf{d}} := (U_{\mathcal{S},k,\mathbf{d},T}, T), \quad (110)$$
$$Y_{j_k,\mathbf{d}} := Y_{j_k,\mathbf{d},T}, \quad (111)$$

and

$$Y_{j_{|\mathcal{S}|},\mathbf{d}} := Y_{j_{|\mathcal{S}|},\mathbf{d},T} \quad (112)$$
$$X_{\mathbf{d}} := X_{\mathbf{d},T}. \quad (113)$$

For $k \in \{2, \ldots, \mathsf{K} - 1\}$, we expand the second summand in (104b) as:

$$\frac{1}{n}I\big(W_{d_{j_k}}; Y_{j_k,\mathbf{d}}^n | V_{j_1}, \ldots, V_{j_k}, W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}\big)$$
$$= \frac{1}{n}\sum_{t=1}^n I\big(W_{d_{j_k}}; Y_{j_k,\mathbf{d},t} | V_{j_1}, \ldots, V_{j_k}, W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}, Y_{j_k,\mathbf{d}}^{t-1}\big)$$
$$\overset{(a)}{=} \frac{1}{n}\sum_{t=1}^n I\big(W_{d_{j_k}}; Y_{j_k,\mathbf{d},t} | V_{j_1}, \ldots, V_{j_k}, W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}},$$
$$Y_{j_1,\mathbf{d}}^{t-1}, \ldots, Y_{j_{k-1},\mathbf{d}}^{t-1}, Y_{j_k,\mathbf{d}}^{t-1}\big)$$
$$\leq \frac{1}{n}\sum_{t=1}^n I\big(W_{d_{j_k}}, Y_{j_k,\mathbf{d}}^{t-1}, V_{j_k}; Y_{j_k,\mathbf{d},t} | V_{j_1}, \ldots, V_{j_{k-1}},$$
$$W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}, Y_{j_1,\mathbf{d}}^{t-1}, \ldots, Y_{j_{k-1},\mathbf{d}}^{t-1}\big)$$
$$\overset{(b)}{=} I\big(U_{\mathcal{S},k,\mathbf{d},T}; Y_{j_k,\mathbf{d},T} | U_{\mathcal{S},k-1,\mathbf{d},T}, T\big)$$
$$\overset{(c)}{=} I\big(U_{\mathcal{S},k,\mathbf{d}}; Y_{j_k,\mathbf{d}} | U_{\mathcal{S},k-1,\mathbf{d}}\big), \quad (114)$$

where (a) follows from the degradedness of the outputs; (b) by (106) and the definition of T; and (c) by the definitions in (110), (112), and (113).

Similarly, we also have

$$\frac{1}{n}I\big(W_{d_{j_{|\mathcal{S}|}}}; Y_{j_{|\mathcal{S}|},\mathbf{d}}^n | V_{j_1}, \ldots, V_{j_{|\mathcal{S}|}}, W_{d_{j_1}}, \ldots, W_{d_{j_{|\mathcal{S}|-1}}}\big)$$
$$= \frac{1}{n}\sum_{t=1}^n I\big(W_{d_{j_{|\mathcal{S}|}}}; Y_{j_{|\mathcal{S}|},\mathbf{d},t} | V_{j_1}, \ldots, V_{j_{|\mathcal{S}|}}, W_{d_{j_1}}, \ldots, W_{d_{j_{|\mathcal{S}|-1}}},$$
$$Y_{j_1,\mathbf{d}}^{t-1}, \ldots, Y_{j_{|\mathcal{S}|-1},\mathbf{d}}^{t-1}, Y_{j_{|\mathcal{S}|},\mathbf{d}}^{t-1}\big)$$
$$\leq \frac{1}{n}\sum_{t=1}^n I\big(W_{d_{j_{|\mathcal{S}|}}}, Y_{j_{|\mathcal{S}|},\mathbf{d}}^{t-1}, V_{j_{|\mathcal{S}|}}; Y_{j_{|\mathcal{S}|},\mathbf{d},t} | V_{j_1}, \ldots, V_{j_{|\mathcal{S}|-1}},$$
$$W_{d_{j_1}}, \ldots, W_{d_{j_{|\mathcal{S}|-1}}}, Y_{j_1,\mathbf{d}}^{t-1}, \ldots, Y_{j_{|\mathcal{S}|-1},\mathbf{d}}^{t-1}\big)$$
$$\leq I\big(X_{\mathbf{d},T}; Y_{j_{|\mathcal{S}|},\mathbf{d},T} | U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d},T}, T\big)$$
$$= I\big(X_{\mathbf{d}}; Y_{j_{|\mathcal{S}|},\mathbf{d}} | U_{\mathcal{S},|\mathcal{S}|-1,\mathbf{d}}\big). \quad (115)$$

It can be verified that the defined random variables satisfy Conditions (102). Combining this observation with (104)–(115) concludes the proof. ∎

We average the bounds in (103) over demand vectors. Let $\mathcal{N}_{\mathsf{K}}^{\mathrm{dist}}$ be the set of all the $\binom{\mathsf{N}}{\mathsf{K}}\mathsf{K}!$ K-dimensional demand vectors with all distinct entries. Also, let $\mathbf{Q}$ be a uniform random variable over the elements of $\mathcal{N}_{\mathsf{K}}^{\mathrm{dist}}$ and independent of all other random variables. Define for each set $\mathcal{S}$ as in (18):

$$U_{\mathcal{S},1} := (U_{\mathcal{S},1,\mathbf{Q}}, \mathbf{Q}), \quad (116)$$
$$U_{\mathcal{S},k} := U_{\mathcal{S},k,\mathbf{Q}}, \quad k \in \{2, \ldots, |\mathcal{S}| - 1\}, \quad (117)$$
$$X := X_{\mathbf{Q}}, \quad (118)$$
$$Y_k := Y_{k,\mathbf{Q}}, \quad k \in \mathcal{K}. \quad (119)$$

Notice that the defined random variables defined in (116)–(119) satisfy conditions (15b) and (68) in the theorem. It remains to prove that they also satisfy (69). To this end, we

average inequalities (103) over all the demand vectors in $\mathcal{N}_K^{\text{dist}}$. Using standard arguments to take care of the averaging random variable $\mathbf{Q}$, and defining

$$\alpha_{\mathcal{S},1} := \frac{1}{\binom{N}{K}K!} \sum_{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}} \frac{1}{n} I(W_{d_{j_1}}; V_1), \tag{120a}$$

$$\alpha_{\mathcal{S},k} := \frac{1}{\binom{N}{K}K!} \sum_{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}} \frac{1}{n} I(W_{d_{j_k}}; V_1, \ldots, V_{j_k}|W_{d_{j_1}}, \ldots, W_{d_{j_{k-1}}}),$$
$$k \in \{2, \ldots, |\mathcal{S}|\}, \tag{120b}$$

we obtain for each $\mathcal{S}$ as in (18):

$$R - \epsilon \le I(U_{\mathcal{S},1}; Y_{j_1}) + \alpha_{\mathcal{S},1}, \tag{121a}$$

$$R - \epsilon \le I(U_{\mathcal{S},k}; Y_{j_k}|U_{\mathcal{S},k-1}) + \alpha_{\mathcal{S},k}, \ \forall k \in \{2, \ldots, |\mathcal{S}|-1\}, \tag{121b}$$

$$R - \epsilon \le I(X; Y_{j_{|\mathcal{S}|}}|U_{\mathcal{S},|\mathcal{S}|-1}) + \alpha_{\mathcal{S},|\mathcal{S}|}, \tag{121c}$$

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

$$0 \le \alpha_{\mathcal{S},k} \le \frac{\sum_{i=1}^{k} M_{j_i}}{N - k + 1}, \qquad k \in \{1, \ldots, |\mathcal{S}|\}, \tag{122a}$$

$$\alpha_{\mathcal{S},k'} \le \alpha_{\mathcal{S},k}, \qquad k, k' \in \{1, \ldots, |\mathcal{S}|\}, \ k' \le k, \tag{122b}$$

$$\sum_{k=1}^{|\mathcal{S}|} \alpha_{\mathcal{S},k} \le \frac{|\mathcal{S}|}{N} \sum_{k=1}^{|\mathcal{S}|} M_{j_k}. \tag{122c}$$

Proof: See Appendix B. ∎

By (121)–(122) 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 (18) random variables $\{U_{\mathcal{S},1}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1}\}$, so that (15b) and (68) hold, and for all $\mathcal{S}$:

$$C(M_1, \ldots, M_K) \le I(U_{\mathcal{S},1}; Y_{j_1}) + \alpha_{\mathcal{S},1}, \tag{123a}$$

$$C(M_1, \ldots, M_K) \le I(U_{\mathcal{S},k}; Y_{j_k}|U_{\mathcal{S},k-1}) + \alpha_{\mathcal{S},k},$$
$$\forall k \in \{2, \ldots, |\mathcal{S}|\}, \tag{123b}$$

for parameters $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$ satisfying (122).

By the following Lemma 17, because constraints (123) are increasing in $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$, and by constraint (122c), we conclude that the choice $\alpha_{\mathcal{S},k} = \alpha_{\mathcal{S},k}^\star$ in (67) makes the upper bound (123) 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:

- $\alpha_{\mathcal{S},k} = \frac{\sum_{i=1}^{k} M_{j_i}}{N-k+1}$; 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 proofs for the other sets are similar.

We first prove (122a). Every $\alpha_{\mathcal{K},k}$ is non-negative, because mutual information is non-negative. To prove the upper bound in (122a), we proceed as follows. Let $\mathcal{N}_K^{\text{dist}}$ again be the set of K-dimensional demand vectors that have K distinct entries in $\{1, \ldots, N\}$; 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:

$$\alpha_{\mathcal{K},k}$$
$$= \frac{1}{K!\binom{N}{K}} \sum_{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}} I(W_{d_k}; V_1, \ldots, V_k|W_{d_1}, \ldots, W_{d_{k-1}})$$

$$= \frac{1}{K!\binom{N}{K}} \sum_{\tilde{\mathbf{d}} \in \mathcal{N}_{k-1}^{\text{dist}}} \sum_{\substack{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}: \\ (d_1, \ldots, d_{k-1}) = \tilde{\mathbf{d}}}} I(W_{d_k}; V_1, \ldots, V_k|W_{\tilde{\mathbf{d}}})$$

$$\overset{(a)}{=} \frac{1}{K!\binom{N}{K}} \sum_{\tilde{\mathbf{d}} \in \mathcal{N}_{k-1}^{\text{dist}}} \sum_{j \in N \setminus \tilde{\mathbf{d}}} I(W_j; V_1, \ldots, V_k|W_{\tilde{\mathbf{d}}})$$
$$\cdot \binom{N-k}{K-k}(K-k)!$$

$$= \frac{1}{k!\binom{N}{k}} \sum_{\tilde{\mathbf{d}} \in \mathcal{N}_{k-1}^{\text{dist}}} \sum_{j \in N \setminus \tilde{\mathbf{d}}} I(W_j; V_1, \ldots, V_k|W_{\tilde{\mathbf{d}}})$$

$$\overset{(b)}{=} \frac{1}{k!\binom{N}{k}} \sum_{\tilde{\mathbf{d}} \in \mathcal{N}_{k-1}^{\text{dist}}} \Big[H(W_1, \ldots, W_N|W_{\tilde{\mathbf{d}}})$$
$$- \sum_{j \in N \setminus \tilde{\mathbf{d}}} H(W_j|V_1, \ldots, V_k, W_{\tilde{\mathbf{d}}}) \Big]$$

$$\overset{(c)}{\le} \frac{1}{k!\binom{N}{k}} \sum_{\tilde{\mathbf{d}} \in \mathcal{N}_{k-1}^{\text{dist}}} I(W_1, \ldots, W_N; V_1, \ldots, V_k|W_{\tilde{\mathbf{d}}})$$

$$\overset{(d)}{\le} \frac{(k-1)!\binom{N}{k-1}}{k!\binom{N}{k}} \sum_{i=1}^{k} M_i$$

$$= \frac{\sum_{i=1}^{k} M_i}{N - k + 1} \tag{124}$$

where (a) holds because for each value of K and j there are $\binom{N-k}{K-k}(K-k)!$ ordered demand vectors $\mathbf{d} \in \mathcal{N}_K^{\text{dist}}$ 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 $\sum_{l=1}^{L} H(A_l) \ge H(A_1, \ldots, A_L)$; and (d) holds because $I(W_1, \ldots, W_N; V_1, \ldots, V_k|W_{\tilde{\mathbf{d}}})$ cannot exceed $\sum_{i=1}^{k} M_i$. This concludes the proof of (122a).

To prove constraint (122b), we fix a K-dimensional demand vector $\mathbf{d} \in \mathcal{N}_K^{\text{dist}}$, 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,

$$d_k^{(\ell)} = d_{(k-\ell) \bmod K} \tag{125}$$

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

$$\xi \bmod K = \xi - bK \quad \text{for some positive integer } b. \tag{126}$$

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

$$
I(W_{d_1}; V_1) \overset{(a)}{=} I(W_{d_{k'}^{(k'-1)}}; V_1)
$$

$$
\overset{(b)}{\leq} I(W_{d_{k'}^{(k'-1)}}; V_1, \ldots, V_{k'} | W_{d_1^{(k'-1)}}, \ldots, W_{d_{k'-1}^{(k'-1)}})
$$

$$
\overset{(a)}{=} I(W_{d_k^{(k-1)}}; V_1, \ldots, V_{k'} | W_{d_{1+k-k'}^{(k-1)}}, \ldots, W_{d_{k-1}^{(k-1)}})
$$

$$
\overset{(b)}{\leq} I(W_{d_k^{(k-1)}}; V_1, \ldots, V_k | W_{d_1^{(k-1)}}, \ldots, W_{d_{k-1}^{(k-1)}})
\tag{127}
$$

where (a) follows by (125) and (b) is by the independence of messages.

Fix a demand vector $\mathbf{d} \in \mathcal{N}_K^{\text{dist}}$ and sum up the above inequality (127) over all K cyclic shifts $\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(K-1)}$ of $\mathbf{d}$ (where for simplicity we relabel the shifts) to obtain:

$$
\sum_{\ell=0}^{K-1} I(W_{d_1^{(\ell)}}; V_1)
$$

$$
\leq \sum_{\ell=0}^{K-1} I(W_{d_{k'}^{(\ell)}}; V_1, \ldots, V_{k'} | W_{d_1^{(\ell)}}, \ldots, W_{d_{k'-1}^{(\ell)}})
$$

$$
\leq \sum_{\ell=0}^{K-1} I(W_{d_k^{(\ell)}}; V_1, \ldots, V_k | W_{d_1^{(\ell)}}, \ldots, W_{d_{k-1}^{(\ell)}}).
\tag{128}
$$

Since the set $\mathcal{N}_K^{\text{dist}}$ can be partitioned into subsets of demand vectors that are cyclic shifts of each others and all cyclic shifts of a demand vector in $\mathcal{N}_K^{\text{dist}}$ are also in $\mathcal{N}_K^{\text{dist}}$, we conclude from (128):

$$
\sum_{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}} I(W_{d_1}; V_1)
$$

$$
\leq \sum_{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}} I(W_{d_{k'}}; V_1, \ldots, V_{k'} | W_{d_1}, \ldots, W_{d_{k'-1}})
$$

$$
\leq \sum_{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}} I(W_{d_k}; V_1, \ldots, V_k | W_{d_1}, \ldots, W_{d_{k-1}}).
\tag{129}
$$

This proves (122b).

We proceed to prove constraint (122c). For each $\mathbf{d} \in \mathcal{N}_K^{\text{dist}}$:

$$
I(W_{d_1}; V_1) + \sum_{k=2}^{K} I(W_{d_k}; V_1, \ldots, V_k | W_{d_1}, W_{d_2}, \ldots, W_{d_{k-1}})
$$

$$
\leq I(W_{d_1}, W_{d_2}, \ldots, W_{d_K}; V_1, \ldots, V_K).
\tag{130}
$$

So,

$$
\sum_{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}} \Bigg[I(W_{d_1}; V_1)
$$

$$
+ \sum_{k=2}^{K} I(W_{d_k}; V_1, \ldots, V_k | W_{d_1}, W_{d_2}, \ldots, W_{d_{k-1}}) \Bigg]
$$

$$
\leq \sum_{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}} I(W_{d_1}, W_{d_2}, \ldots, W_{d_K}; V_1 \ldots, V_K)
$$

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

$$
- H(W_{d_1}, \ldots, W_{d_K} | V_1, \ldots, V_K) \Big]
$$

$$
\overset{(b)}{=} \frac{K}{N} |\mathcal{N}_K^{\text{dist}}| H(W_1, \ldots, W_N)
$$

$$
- \sum_{\mathbf{d} \in \mathcal{N}_K^{\text{dist}}} H(W_{d_1}, \ldots, W_{d_K} | V_1, \ldots, V_K)
$$

$$
\overset{(c)}{\leq} \frac{K}{N} K! \binom{N}{K} H(W_1, \ldots, W_N)
$$

$$
- \frac{K}{N} K! \binom{N}{K} H(W_1, \ldots, W_N | V_1, \ldots, V_K)
$$

$$
\overset{(b)}{=} \frac{K}{N} K! \binom{N}{K} I(W_1, \ldots, W_N; V_1, \ldots, V_K)
$$

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

where (a) holds by the chain rule of mutual information, (b) by the independence and uniform rate of messages $W_1, \ldots, W_N$ and the definition of the set $\mathcal{N}_K^{\text{dist}}$, which is of size $\binom{N}{K} K!$, 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 $i \in \{1, \ldots, L\}$:

$$
\frac{1}{\binom{L}{i}} \sum_{\substack{\mathcal{J} \subseteq \{1, \ldots, L\}: \\ |\mathcal{J}| = i}} \frac{H(A_{\mathcal{J}})}{i} \geq \frac{1}{L} H(A_1, \ldots, A_L).
\tag{131}
$$

Proof: See [45, 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 (15b) and for each set $\mathcal{S}$ as in (18) a random tuple $U_{\mathcal{S},1}, U_{\mathcal{S},2}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1}$ satisfying (68) and real numbers $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$ satisfying (122).

Assume that for some set $\mathcal{S}$ as in (18) and some $\tilde{k} \in \{1, \ldots, |\mathcal{S}| - 1\}$:

$$
\alpha_{\mathcal{S}, \tilde{k}} \neq \alpha_{\mathcal{S}, \tilde{k}+1}
\tag{132}
$$

and

$$
\alpha_{\mathcal{S}, \tilde{k}} < \frac{\sum_{i=1}^{\tilde{k}} M_{j_i}}{N - \tilde{k} + 1}.
\tag{133}
$$

Let

$$
\gamma := \max \left\{ \frac{1}{2}, \frac{\alpha_{\mathcal{S}, \tilde{k}+1} - \frac{\sum_{i=1}^{\tilde{k}} M_{j_i}}{N - \tilde{k} + 1}}{\alpha_{\mathcal{S}, \tilde{k}+1} - \alpha_{\mathcal{S}, \tilde{k}}} \right\}.
\tag{134}
$$

Notice that by (133):

$$
\gamma \in \left[\frac{1}{2}, 1 \right).
\tag{135}
$$

Define the new parameters

$$
\bar{\alpha}_{\mathcal{S},k} := \alpha_{\mathcal{S},k}, \quad k \in \{1, \ldots, |\mathcal{S}|\} \setminus \{\tilde{k}, \tilde{k}+1\}
\tag{136a}
$$

$$
\bar{\alpha}_{\mathcal{S},\tilde{k}} := \gamma \alpha_{\mathcal{S},\tilde{k}} + (1-\gamma) \alpha_{\mathcal{S},\tilde{k}+1}
\tag{136b}
$$

$$
\bar{\alpha}_{\mathcal{S},\tilde{k}+1} := (1-\gamma) \alpha_{\mathcal{S},\tilde{k}} + \gamma \alpha_{\mathcal{S},\tilde{k}+1}.
\tag{136c}
$$

Notice that this new set of parameters satisfies constraints (122) 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,

$$\bar{\alpha}_{\mathcal{S},k} \leq \bar{\alpha}_{\mathcal{S},k+1}, \qquad k \in \{1, \ldots, |\mathcal{S}| - 1\}. \tag{137}$$

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 (68), and so that upper bound (103) is looser for these new auxiliaries 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

$$U_{\mathcal{S},|\mathcal{S}|} := X. \tag{138}$$

Notice that since $\alpha_{\mathcal{S},\tilde{k}} \neq \alpha_{\mathcal{S},\tilde{k}+1}$ and by (122b), the strict inequality

$$\alpha_{\mathcal{S},\tilde{k}} < \alpha_{\mathcal{S},\tilde{k}+1} \tag{139}$$

must hold. Choose

$$\bar{U}_{\mathcal{S},k} = U_{\mathcal{S},k}, \qquad k \in \{1, \ldots, |\mathcal{S}| - 1\} \backslash \{\tilde{k}\}, \tag{140}$$

and

$$\bar{U}_{\mathcal{S},|\mathcal{S}|} = U_{\mathcal{S},|\mathcal{S}|} = X. \tag{141}$$

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

$$I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}) \leq I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}), \tag{142a}$$

or

$$I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}) > I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}). \tag{142b}$$

If (142a) holds, choose

$$\bar{U}_{\mathcal{S},\tilde{k}} = U_{\mathcal{S},\tilde{k}}. \tag{143}$$

If (142b) holds, let $E \in \{0,1\}$ be a Bernoulli-β random variable independent of everything else, where

$$\beta := (1 - \gamma) - (1 - \gamma) \cdot \frac{I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}})}{I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1})}. \tag{144}$$

Choose

$$\bar{U}_{\mathcal{S},\tilde{k}} = \begin{cases} (U_{\mathcal{S},\tilde{k}}, E), & \text{if } E = 0 \\ (U_{\mathcal{S},\tilde{k}-1}, E), & \text{if } E = 1. \end{cases} \tag{145}$$

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 (103) is unchanged if we replace $(U_{\mathcal{S},1}, U_{\mathcal{S},2}, \ldots, U_{\mathcal{S},|\mathcal{S}|-1}, X)$ by $(\bar{U}_{\mathcal{S},1}, \bar{U}_{\mathcal{S},2}, \ldots, \bar{U}_{\mathcal{S},\mathsf{K}-1}, X)$ 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 (142a) holds, then the proposed replacement relaxes constraint (103) for $k = \tilde{k}$ (because $\bar{\alpha}_{\mathcal{S},\tilde{k}} > \alpha_{\mathcal{S},\tilde{k}}$) and it tightens it for $k = \tilde{k}+1$ (because $\bar{\alpha}_{\mathcal{S},\tilde{k}+1} < \alpha_{\mathcal{S},\tilde{k}+1}$). However, the new constraint for $k = \tilde{k}+1$ is less stringent than the original constraint for $k = \tilde{k}$:

$$\bar{\alpha}_{\mathcal{S},\tilde{k}+1} + I(\bar{U}_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | \bar{U}_{\mathcal{S},\tilde{k}})$$
$$\overset{(a)}{=} (1 - \gamma) \cdot \alpha_{\mathcal{S},\tilde{k}} + \gamma \cdot \alpha_{\mathcal{S},\tilde{k}+1} + I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}})$$
$$\overset{(b)}{>} \alpha_{\mathcal{S},\tilde{k}} + I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}})$$

$$\overset{(c)}{\geq} \alpha_{\mathcal{S},\tilde{k}} + I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}), \tag{146}$$

where (a) holds by (136c); (b) holds by (139); and (c) holds by holds by assumption (142a). We conclude that when (142a) holds, the upper bound on $\mathsf{C}(\mathsf{M}_1, \ldots, \mathsf{M}_\mathsf{K})$ in (103) 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 (142b) 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 (144) and (145) we have

$$I(\bar{U}_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | \bar{U}_{\mathcal{S},\tilde{k}-1})$$
$$= I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | \bar{U}_{\mathcal{S},\tilde{k}-1} E)$$
$$= (1 - \beta) \cdot I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1})$$
$$= \gamma \cdot I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1})$$
$$\quad + (1 - \gamma) \cdot I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}). \tag{147}$$

By (136b) and (147):

$$\bar{\alpha}_{\mathcal{S},\tilde{k}} + I(\bar{U}_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | \bar{U}_{\mathcal{S},\tilde{k}-1})$$
$$= (\gamma \alpha_{\mathcal{S},\tilde{k}} + (1 - \gamma) \alpha_{\mathcal{S},\tilde{k}+1})$$
$$\quad + \gamma I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}) + (1 - \gamma) I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}})$$
$$\geq \min \big\{ \alpha_{\mathcal{S},\tilde{k}} + I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}),$$
$$\qquad \alpha_{\mathcal{S},\tilde{k}+1} + I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) \big\}. \tag{148}$$

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

$$I(\bar{U}_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | \bar{U}_{\mathcal{S},\tilde{k}})$$
$$\overset{(a)}{=} (1 - \beta) I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}})$$
$$\quad + \beta I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}-1})$$
$$\overset{(b)}{=} (1 - \beta) I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}})$$
$$\quad + \beta I(U_{\mathcal{S},\tilde{k}+1}, U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}-1})$$
$$\overset{(c)}{=} I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) + \beta I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}-1})$$
$$\overset{(d)}{\geq} I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) + \beta I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1})$$
$$\overset{(e)}{=} \gamma I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) + (1 - \gamma) I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}), \tag{149}$$

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 (68); (c) by the chain rule of mutual information and Markov chain (68); (d) by the degradedness of the channel (15b); (e) by the definition of β in (144).

Therefore, by (136c):

$$\bar{\alpha}_{\mathcal{S},k+1} + I(\bar{U}_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | \bar{U}_{\mathcal{S},\tilde{k}})$$
$$\geq (1 - \gamma) \cdot \alpha_{\mathcal{S},\tilde{k}} + \gamma \cdot \alpha_{\mathcal{S},\tilde{k}+1}$$
$$\quad + (1 - \gamma) \cdot I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}) + \gamma \cdot I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}})$$
$$\geq \min \big\{ \alpha_{\mathcal{S},\tilde{k}} + I(U_{\mathcal{S},\tilde{k}}; Y_{\tilde{k}} | U_{\mathcal{S},\tilde{k}-1}),$$
$$\qquad \alpha_{\mathcal{S},\tilde{k}+1} + I(U_{\mathcal{S},\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\mathcal{S},\tilde{k}}) \big\}. \tag{150}$$

We thus conclude that also when (142b) holds, the upper bound on $C(M_1, \ldots, M_K)$ in (103) 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 4 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 4, Appendix A. Relax the upper bound in this lemma by replacing for $k = 2, \ldots, K$ constraint (122a) by

$$\alpha_{\mathcal{S},k} \geq 0. \tag{151}$$

Following similar steps as in the proof of Lemma 17, see also [22, 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

$$\alpha_{\mathcal{S},1} = \frac{M_1}{N} \qquad \text{or} \qquad \alpha_{\mathcal{S},1} = \alpha_{\mathcal{S},2}.$$

Since constraints (123) are increasing in $\alpha_{\mathcal{S},1}, \ldots, \alpha_{\mathcal{S},|\mathcal{S}|}$, by constraint (122c), we conclude that the relaxed upper bound is loosest for

$$\alpha_{\mathcal{S},1} = \frac{M_1}{N}$$
$$\alpha_{\mathcal{S},k} = \frac{|\mathcal{S}| \sum_{i=1}^{|\mathcal{S}|} M_{j_i} - M_{j_1}}{(|\mathcal{S}|-1)N}, \quad k \in \{2, \ldots, |\mathcal{S}|-1\},$$

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

We now prove that the bound in Theorem 4 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 (122a) by replacing it with $\alpha_{\mathcal{S},k} \geq 0$, for *all* $k = 1, \ldots, K$. Following the steps in [22, Lemma 12], it can be shown that the new constraints are loosest if each

$$\alpha_{\mathcal{S},k} = \alpha_{\mathcal{S},k}'. \tag{153}$$

This concludes the proof.

APPENDIX E
PROOF OF PROPOSITION 7

For $\Delta = 0$, achievability follows by specializing Theorem 3 to $t = K - 1$ and to the input distribution P_X that maximizes (74). In fact, for this input distribution:

$$R^{(K-1)} = KC_{\text{avg}} = C_{\text{avg}} + \frac{\sum_{k=1}^{K} M_k^{\star(K-1)}}{K \cdot N}.$$

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

The converse is proved as follows. Consider cache sizes $M_1^{*(K-1)}, \ldots, M_K^{*(K-1)}$ as given in (41). Apply Theorem 4, but consider only the constraints (69) 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 (15b) and so that

$$C(M_1^{(K-1)}, \ldots, M_K^{(K-1)}) \leq \frac{1}{K} \sum_{k \in \mathcal{K}} I(X; Y_k) + \frac{1}{K} \sum_{k \in \mathcal{K}} \frac{M_k^{(K-1)}}{N}. \tag{154}$$

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 4 by considering constraints (69) 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:

$$C(M_1, \ldots, M_K)$$
$$\leq \max \min \left\{ (1-\delta_1)\beta_1 + \frac{M_1}{N}, (1-\delta_2)\beta_2 + \frac{KM - M_1}{N \cdot (K-1)}, \right.$$
$$\left. \ldots, (1-\delta_K)\beta_K + \frac{KM - M_1}{N \cdot (K-1)} \right\}, \tag{155}$$

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

$$\sum_{k=1}^{K} \beta_k \leq 1. \tag{156}$$

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

$$C(M_1, \ldots, M_K)$$
$$\leq \max_{\beta_1 \in [0,1]} \min \left\{ \beta_1 C_1 + \frac{M_1}{N}, \right.$$
$$\left. (1-\beta_1)C_{\{2,\ldots,K\}} + \frac{KM - M_1}{(K-1) \cdot N} \right\}$$
$$= \max_{\beta_1 \in [0,1]} \min \left\{ \beta_1 C_1, (1-\beta_1)C_{\{2,\ldots,K\}} + \frac{K(M - M_1)}{(K-1) \cdot N} \right\}$$
$$+ \frac{M_1}{N}. \tag{157}$$

If

$$\frac{K(M - M_1)}{(K-1) \cdot N} \geq C_1,$$

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

$$C(M_1, \ldots, M_K) \leq C_1 + \frac{M_1}{N}. \tag{158}$$

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

$$\beta_1^\star := \frac{C_{\{2,\ldots,K\}} + \frac{K(M-M_1)}{(K-1) \cdot N}}{C_1 + C_{\{2,\ldots,K\}}}, \tag{159}$$

and the upper bound results in

$$C(M_1, \ldots, M_K) \leq C_{\mathcal{K}} + \frac{K(M - M_1)}{(K-1) \cdot N} \cdot \frac{C_1}{C_1 + C_{\{2,\ldots,K\}}} + \frac{M_1}{N}$$
$$= C_{\mathcal{K}} + \frac{K(M - M_1)}{(K-1) \cdot N} \cdot \frac{C_{\mathcal{K}}}{C_{\{2,\ldots,K\}}} + \frac{M_1}{N} \quad (160)$$

where we used that for erasure BCs

$$C_{\mathcal{K}} = \frac{C_1 \cdot C_{\{2,\ldots,K\}}}{C_1 + C_{\{2,\ldots,K\}}}. \quad (161)$$

APPENDIX G
PROOF OF COROLLARY 13

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

$$(r_1 \ldots, r_t) \in \mathcal{C}_{\mathcal{S}} \quad \Longleftrightarrow \quad \sum_{k=1}^{t} r_k \leq 1. \quad (162)$$

The upper bound in Corollary 5 thus states that for this noise-free BC a rate-memory tuple $(R, M_1, \ldots, M_K)$ is achievable only if

$$tR - \sum_{k=1}^{t} \alpha_{\mathcal{S},k}^{\star} \leq 1. \quad (163)$$

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

$$C(M_1, \ldots, M_K) \leq \frac{1}{t}\left(1 + \sum_{k=1}^{t} \alpha_{\mathcal{S},k}^{\star}\right). \quad (164)$$

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

$$\sum_{k=1}^{t} \alpha_{\mathcal{S},k}^{\star} = \frac{t \sum_{k=1}^{t} M_k}{N} \quad (165a)$$

or

$$\sum_{k=1}^{t} \alpha_{\mathcal{S},k}^{\star} = \sum_{k=1}^{t} \frac{\sum_{i=1}^{k} M_i}{N - k + 1}. \quad (165b)$$

Combining (164) with (165), applying the correspondence $\rho = R^{-1}$ and $m_k = \frac{M_k}{R}$, and setting $m_1 = m_2 = \ldots = m_k = m$ yields,

$$1 \leq \frac{1}{t}\left(\rho + m \cdot \min\left\{\frac{t^2}{N}, \sum_{k=1}^{t} \frac{k}{N-k+1}\right\}\right), \quad (166)$$

which is equivalent to the bound in the corollary.

REFERENCES

[1] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. 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 users with small buffers," *IET Commun.*, vol. 10, no. 17, pp. 2315–2318, 2016.

[3] K. Wan, D. Tuninetti, and P. Piantanida, "On caching with more users than files," in *IEEE Int. Symp. Inf. Theory*, July 2016, pp. 135–139.

[4] S. Sahraei and M. Gastpar, "K users caching two files: An improved achievable rate," in *2016 Annual Conf. Inf. Science and Syst.*, March 2016, pp. 620–624.

[5] C. Tian and J. Chen, "Caching and delivery via interference elimination," *IEEE Trans. Inf. Theory*, vol. 64, no. 3, pp. 1548–1560, March 2018.

[6] M. M. Amiri, Q. Yang, and D. Gündüz, "Coded caching for a large number of users," in *IEEE Inf. Theory Workshop*, Sept 2016, pp. 171–175.

[7] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1281–1296, Feb 2018.

[8] J. Gómez-Vilardebó, "Fundamental limits of caching: Improved bounds with coded prefetching," *CoRR*, vol. abs/1612.09071, 2016. [Online]. Available: http://arxiv.org/abs/1612.09071

[9] C. Tian, "On the fundamental limits of coded caching and exact-repair regenerating codes," in *2015 International Symposium on Network Coding (NetCod)*, June 2015, pp. 56–60.

[10] A. Sengupta, R. Tandon, and T. C. Clancy, "Improved approximation of storage-rate tradeoff for caching via new outer bounds," in *IEEE Int. Symp. Inf. Theory*, June 2015, pp. 1691–1695.

[11] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4388–4413, July 2017.

[12] C. Y. Wang, S. H. Lim, and M. Gastpar, "A new converse bound for coded caching," in *IEEE Inf. Theory and Applications Workshop*, Jan 2016, pp. 1–6.

[13] C. Tian, "Symmetry, demand types and outer bounds in caching systems," in *IEEE Int. Symp. Inf. Theory*, July 2016, pp. 825–829.

[14] M. M. Amiri, Q. Yang, and D. Gündüz, "Decentralized caching and coded delivery with distinct cache capacities," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4657–4669, Nov 2017.

[15] S. Wang, W. Li, X. Tian, and H. Liu, "Fundamental limits of heterogenous cache," *CoRR*, vol. abs/1504.01123, 2015. [Online]. Available: http://arxiv.org/abs/1504.01123

[16] A. M. Ibrahim, A. A. Zewail, and A. Yener, "Centralized coded caching with heterogeneous cache sizes," in *IEEE Wireless Commun. Networking Conf. (WCNC)*, March 2017, pp. 1–6.

[17] ——, "Optimization of heterogeneous caching systems with rate limited links," in *ICC*, May 2017, pp. 1–6.

[18] W. Huang, S. Wang, L. Ding, F. Yang, and W. Zhang, "The performance analysis of coded cache in wireless fading channel," *CoRR*, vol. abs/1504.01452, 2015. [Online]. Available: http://arxiv.org/abs/1504.01452

[19] M. Ji and R. R. Chen, "Caching and coded multicasting in slow fading environment," in *IEEE Wireless Commun. Networking Conf. (WCNC)*, March 2017, pp. 1–6.

[20] J. Zhang and P. Elia, "Wireless coded caching: A topological perspective," in *IEEE Int. Symp. Inf. Theory*, June 2017, pp. 401–405.

[21] R. Timo and M. Wigger, "Joint cache-channel coding over erasure broadcast channels," in *Int. Symp. Wireless Commun. Syst.*, Aug 2015, pp. 201–205.

[22] S. S. Bidokhti, M. A. Wigger, and R. Timo, "Noisy broadcast networks with receiver caching," *IEEE Trans. Inf. Theory*, 2018, to appear.

[23] M. M. Amiri and D. Gündüz, "Cache-aided data delivery over erasure broadcast channels," in *IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.

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

[25] P. Hassanzadeh, E. Erkip, J. Llorca, and A. Tulino, "Distortion-memory tradeoffs in cache-aided wireless video delivery," in *Allerton Conf. on Commun., Control, and Computing*, Sept 2015, pp. 1150–1157.

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

[27] J. Zhang and P. Elia, "Fundamental limits of cache-aided wireless BC: Interplay of coded-caching and CSIT feedback," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3142–3160, May 2017.

[28] E. Piovano, H. Joudeh, and B. Clerckx, "On coded caching in the overloaded MISO broadcast channel," in *IEEE Int. Symp. Inf. Theory (ISIT)*, June 2017, pp. 2795–2799.

[29] S. Yang, K. H. Ngo, and M. Kobayashi, "Content delivery with coded caching and massive MIMO in 5G," in *Int. Symp. Turbo Codes and Iterative Inf. Process.*, Sept 2016, pp. 370–374.

[30] S. P. Shariatpanahi, G. Caire, and B. Hossein Khalaj, "Physical-layer schemes for wireless coded caching," *IEEE Trans. Inf. Theory*, vol. 65, no. 5, pp. 2792–2807, May 2019.

[31] P. Hassanzadeh, A. M. Tulino, J. Llorca, and E. Erkip, "Broadcast caching networks with two receivers and multiple correlated sources," in *Asilomar Conf. Signals, Systems, and Computers*, Oct 2017, pp. 1449–1454.

[32] Q. Yang, P. Hassanzadeh, D. Gündüz, and E. Erkip, "Centralized caching and delivery of correlated contents over a Gaussian broadcast channel," in *Int. Symp. on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2018, pp. 1–6.

[33] A. Ibrahim, A. Zewail, and A. Yener, "On coded caching with heterogeneous distortion requirements," in *IEEE Inf. Theory and Applications Workshop*, Feb 2018.

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

[35] S. S. Bidokhti, M. Wigger, and R. Timo, "An upper bound on the capacity-memory tradeoff of degraded broadcast channels," in *Int. Symp. Turbo Codes and Iterative Inf. Process.*, Sept 2016, pp. 350–354.

[36] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "Characterizing the rate-memory tradeoff in cache networks within a factor of 2," in *IEEE Int. Symp. Inf. Theory*, June 2017, pp. 386–390.

[37] S. S. C. Y. Wang and M. W. R. Timo, "Improved converses and gap-results for coded caching," *IEEE Trans. Inf. Theory*, 2018, to appear.

[38] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Peredachi Inf.*, pp. 185–193, 1974.

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

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

[41] M. Wigger, R. Timo, and S. Shamai, "Complete interference mitigation through receiver-caching in wyner's networks," in *IEEE Inf. Theory Workshop*, Sept 2016, pp. 335–339.

[42] Q. Yan, M. Cheng, X. Tang, and Q. Chen, "On the placement delivery array design for centralized coded caching scheme," *IEEE Trans. Inf. Theory*, vol. 63, no. 9, pp. 5821–5833, Sept 2017.

[43] K. Shanmugam, A. M. Tulino, and A. G. Dimakis, "Coded caching with linear subpacketization is possible using Ruzsa-Szemérédi graphs," in *ISIT*, June 2017, pp. 1237–1241.

[44] L. Tang and A. Ramamoorthy, "Coded caching schemes with reduced subpacketization from linear block codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, pp. 3099–3120, April 2018.

[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

Shirin Saeedi Bidokhti (S'09-M'12) is a research assistant professor in the Department of Electrical and Systems Engineering at the University of Pennsylvania (UPenn). She received the B.Sc. degree in Electrical Engineering from University of Tehran in 2005, and the M.Sc and Ph.D. degrees in Communication Systems from École Polytechnique Fédérale de Lausanne (EPFL) in 2007 and 2012, respectively. Before joining UPenn in 2017, she was a postdoctoral fellow with the Institute for Communications Engineering at the Technische Universität München (2013-2015), a postdoctoral fellow with the Department of Electrical Engineering at Stanford University (2015-2017), and a postdoctoral research scholar with the Department of Electrical Engineering at the Pennsylvania State University (2017). She is a recipient of the 2019 NSF-CRII Research Initiative award and the prospective researcher and advanced postdoctoral fellowships from the Swiss National Science Foundation. Her research interests include network information theory, network coding, and data compression.

Michèle Wigger (S'05, M'09, SM'14) received the M.Sc. degree in electrical engineering, with distinction, and the Ph.D. degree in electrical engineering both from ETH Zurich in 2003 and 2008, respectively. In 2009, she was first a post-doctoral fellow at the University of California, San Diego, USA, and then joined Telecom Paris Tech, Paris, France, where she is currently a Full Professor. Dr. Wigger has held visiting professor appointments at the Technion-Israel Institute of Technology and ETH Zurich. Dr. Wigger has previously served as an Associate Editor of the IEEE Communication Letters, and is now Associate Editor for Shannon Theory of the IEEE Transactions on Information Theory. She is currently also serving on the Board of Governors of the IEEE Information Theory Society. Dr. Wigger's current research interests are in distributed source coding, distributed detection systems, free-space optical communication, caching systems, and networks with mixed-delay constraints.

Aylin Yener (S'91–M'01–SM'14–F'15) received the B.Sc. degree in electrical and electronics engineering and the B.Sc. degree in physics from Bogazici University, Istanbul, Turkey, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Wireless Information Network Laboratory (WINLAB), Rutgers University, New Brunswick, NJ, USA. She is a Distinguished Professor of Electrical Engineering at The Pennsylvania State University, University Park, PA, USA, where she joined the faculty as an assistant professor in 2002. Since 2017, she is also a Dean's Fellow in the College of Engineering at The Pennsylvania State University. She was a visiting professor of Electrical Engineering at Stanford University in 2016-2018 and a visiting associate professor in the same department in 2008-2009. Her current research interests are in information security, green communications, caching systems, and more generally in the fields of information theory, communication theory and networked systems. She received the NSF CAREER Award in 2003, the Best Paper Award in Communication Theory from the IEEE International Conference on Communications in 2010, the Penn State Engineering Alumni Society (PSEAS) Outstanding Research Award in 2010, the IEEE Marconi Prize Paper Award in 2014, the PSEAS Premier Research Award in 2014, the Leonard A. Doggett Award for Outstanding Writing in Electrical Engineering at Penn State in 2014, the IEEE Women in Communications Engineering Outstanding Achievement Award in 2018, and the IEEE Communications Society Best Tutorial Paper Award in 2019. She is a distinguished lecturer for the IEEE Information Theory Society (2019-2020), the IEEE Communications Society (2018-2020) and the IEEE Vehicular Technology Society (2017-2019). Dr. Yener is serving as the vice president of the IEEE Information Theory Society in 2019. Previously she was the second vice president (2018), member of the Board of Governors (2015-2018) and the treasurer (2012-2014) of the IEEE Information Theory Society. She served as the Student Committee Chair for the IEEE Information Theory Society (2007-2011), and was the co-Founder of the Annual School of Information Theory in North America in 2008. She was a Technical (Co)-Chair for various symposia/tracks at the IEEE ICC, PIMRC, VTC, WCNC, and Asilomar in 2005, 2008-2014 and 2018. Previously, she served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS (2009-2012), an Editor for the IEEE TRANSACTIONS ON MOBILE COMPUTING (2017-2018), and an Editor and an Editorial Advisory Board Member for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2001-2012). She also served a Guest Editor for the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY in 2011, and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS in 2015. Currently, she serves as a Senior Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and is on the inaugural Senior Editorial Board of the IEEE JOURNAL ON SELECTED AREAS IN INFORMATION THEORY.