Noisy Broadcast Networks with Receiver Caching

Shirin Saeedi Bidokhti, Michèle Wigger and Roy Timo

***Abstract*—An erasure broadcast network is considered with two disjoint sets of receivers: a set of weak receivers with all-equal erasure probabilities and equal cache sizes, and a set of strong receivers with all-equal erasure probabilities and no cache memories. Lower and upper bounds are presented on the *capacity-memory tradeoff* of this network (the largest rate at which messages can be reliably communicated for given cache sizes). The lower bound is achieved by means of a *joint cache-channel coding scheme* and significantly improves over traditional schemes based on *separate cache-channel coding*. In particular, it is shown that joint cache-channel coding offers new global caching gains that scale with the number of strong receivers in the network. The upper bound uses bounding techniques from degraded broadcast channels and introduces an averaging argument to capture the fact that the contents of the cache memories are designed before knowing users' demands. The derived upper bound is valid for all stochastically degraded broadcast channels. The lower and upper bounds match for a single weak receiver (and any number of strong receivers) when the cache size does not exceed a certain threshold. Improved bounds are presented for the special case of a single weak and a single strong receiver with two files and the bounds are shown to match over a large range of cache sizes.**

I. Introduction

We address a one-to-many broadcast communication problem where many users demand files from a single server during *peak-traffic* times (periods of high network congestion). To improve network performance, the server can pre-place information in local cache memories near users during off-peak times when the communication rate is not a limiting network resource. The server typically does not know in advance which files the users will demand, so it will try to place information that is likely to be useful for many users during periods of peak-traffic. Machine-learning techniques can be used to predict user behavior and identify files that are more likely to be selected in peak-traffic [1].

The above caching problem is particularly relevant to video- and music-streaming services in mobile networks: Content providers pre-place information in clients' caches (or, on servers close to the clients) to improve latency and throughput during peak-traffic times. The content provider however does not know in advance which movies or songs the clients will request, and thus the cached information cannot depend on the clients' specific demands. It somehow needs to be generic and fit the demands of a large group of users. For example, music-providers automatically cache popular songs on the region's "top 50" chart.

The information-theoretic aspects of cache-aided communications have received significant attention in recent years. Maddah-Ali and Niesen [2] considered a one-to-many communications problem where the receivers have independent caches of equal sizes and the *delivery phase* (the peak-traffic communication) takes place over a noiseless bit-pipe that is connected to all receivers. They showed that a smart design of the cache contents enables the server to send coded (XOR-ed) data during the delivery phase and simultaneously meet the demands of multiple receivers. This *coded caching scheme* allows the server to reduce the delivery rate beyond the obvious *local caching gain*, i.e., the data rate that each receiver can locally retrieve from its cache. Intuitively, the performance improvement occurs because the receivers can profit from other receivers' caches and the gain scales with the number of receivers in the networks. It was thus termed [2] a *global caching gain*. Several works [2]–[15] have presented upper and lower bounds on the minimum delivery rate as a function of the cache sizes. The problem has also been studied in the framework of lossless and lossy source coding in [16]–[19].

More realistically, the channel for the delivery phase can be modeled as a *noisy* broadcast channel (BC). This modeling approach was followed for example in [20]–[35] and is also the basis of this current paper. While the works in [21]–[23], [25]–[33], [35] focus on the interplays of caching with feedback, channel state information and massive MIMO, in this paper, we consider a simple erasure BC and show that further global caching gains can be achieved by *joint cache-channel coding*. In joint cache-channel coding, cache contents not only determine *what* to transmit but also *how* to transmit it. Previous works have adopted *separate cache-channel coding architectures* with encoder/decoders consisting of a cache encoder/decoder and a channel encoder/decoder specifically designed for the cache contents and the BC statistics, respectively. Notice that by recasting the cache-contents as sources available at the receiver, joint cache-channel coding becomes an instance of *joint source-channel coding*. Joint source-channel coding schemes for BCs without cache memories but with receiver side-information have previously been presented in [36]–[38], [40]–[47]. Tuncel [36], for example, provided sufficient and necessary conditions when a memoryless source can be transmitted losslessly over a BC with receiver side-information. Particularly related to the caching model here is the scenario where the receivers' side-information is also available at the transmitter [46], [47], a scenario that also relates to the BC with *feedback* because the fed back chan-

S. Saeedi Bidokhti is with the Department of Electrical Engineering at Stanford University, saeedi@stanford.edu. S. Saeedi Bidokhti is supported by the Swiss National Science Foundation fellowship no. 158487.

M. Wigger is with LTCI, Telecom ParisTech, Université Paris-Saclay, 75013 Paris, michele.wigger@telecom-paristech.fr. The work of M. Wigger was supported by the ERC under grant agreement 715111.

R. Timo is with Ericsson Research, Stockholm, Sweden, roy.timo@ericsson.com.

The material in this paper has been presented, in part, at the *IEEE International Symposium on Wireless Communications Systems (ISWCS)*, Bruxelles, Belgium, August 2015 [20], the *2016 IEEE International Symposium on Information Theory*, Barcelona, Spain, and the *International Symposium on Turbo Codes & Iterative Information Processing*, Brest, France, September 2016.

nel outputs can be viewed as such a side-information. The feedback schemes in [44], [48]–[57] exactly exploit this side-information to improve over the no-feedback capacity region of the considered BCs. Joint source-channel coding is however only used in [44], [48].

The main interest of this paper is on the fundamental *capacity-memory tradeoff* (i.e., the largest rate at which messages can be reliably communicated for given cache sizes) of the K-receiver erasure BC illustrated in Figure 1. In this BC the receivers are partitioned into two sets:

- A set of K_w *weak receivers* with equal "large" BC erasure probabilities $\delta_\mathrm{w} \geq 0$. These receivers are each equipped with an individual cache of size nM bits.
- A set of $K_\mathrm{s} := K - K_\mathrm{w}$ *strong receivers* with equal "small" BC erasure probabilities $\delta_\mathrm{s} \geq 0$ with $\delta_\mathrm{s} \leq \delta_\mathrm{w}$. These receivers are not provided with caches.

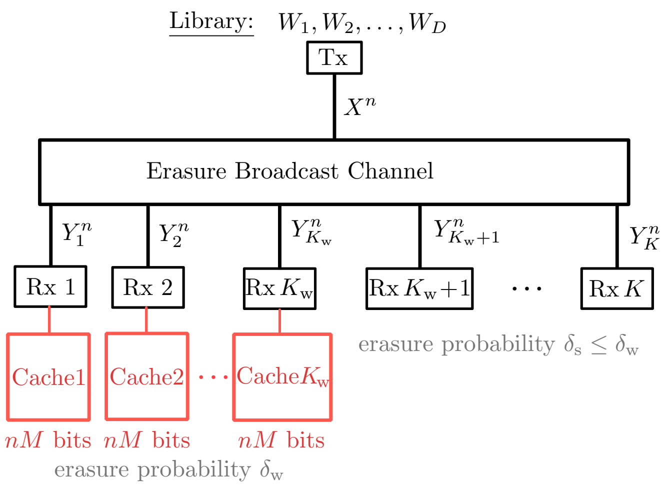

Fig. 1: K user erasure BC with K_w weak and K_s strong receivers and where the weak receivers have cache memories.

This scenario is motivated by previous studies [20], [21] that showed the benefit of prioritizing cache placements near weaker receivers. In practical systems, this means that telecommunications operators with a limited number of caches might first place caches at houses that are further away from an optical fiber access point. Or, they might place caches at pico or femto base stations in heterogenous networks that are located in areas with notoriously bad throughput. Another motivation for this work is the idea of using caching techniques to extend the coverage of cellular networks in remote rural areas. For example, the 3GPP technical report [58, page 16] outlines a single-cell *extreme long distance coverage in low density areas* deployment scenario in which network providers aim to provide opportunistic coverage to relatively few users located up to 100km from the basestation. The performance of these extreme cell-edge users could be improved, for example, by using *customer-premise equipment (CPE)* with caching. Specifically, the network provider could use caching to enable the weak extreme cell-edge users to be co-scheduled with strong users on the same time-frequency resources, and, thus, allocate a greater portion of the time-frequency resources to the few extreme cell-edge users.

In this work, we find lower and upper bounds on the capacity-memory tradeoff $C(M)$ of the broadcast network shown in Figure 1. The proposed lower bounds are achieved by a joint cache-channel code design that builds on the *piggyback coding* idea in [20], [48]. The basic idea of piggyback coding is to carry messages to strong receivers on the back of messages to the weak receivers. These messages can be carried for "free" if the server pre-places appropriate *message side-information* in the weak receivers' caches. A secrecy-version of the piggyback coding idea has previously been adopted in [39]. We remark that by recasting the cached contents as message side-information, one could also use Tuncel coding [36] (instead of piggyback coding)[1]. Our joint cache-channel coding exploits the heterogeneity of the users in their cache sizes. A similar technique has concurrently been developed to exploit heterogeneity in rate [24]. In a dual manner, more recently, joint source-channel coding is beneficial for lossy transmission over cache-aided networks when there is heterogeneity in the distortion criteria [19].

The new lower bounds substantially improve over standard separate cache-channel coding schemes that combine stand-alone codes for coded caching [2] and for reliable transmission over the erasure BC. For example, when M is sufficiently small (depending on certain problem parameters), the joint cache-channel coding scheme achieves the following lower bound on the capacity-memory tradeoff:

$$C(M) \geq R_0 + \frac{K_\mathrm{w}(1-\delta_\mathrm{s})}{K_\mathrm{w}(1-\delta_\mathrm{s}) + K_\mathrm{s}(1-\delta_\mathrm{w})} \cdot \frac{1+K_\mathrm{w}}{2} \cdot \gamma_{\text{joint}} \cdot \frac{M}{D}. \tag{1}$$

Here, R_0 represents the largest symmetric rate that is achievable over the erasure BC in Figure 1 when neither strong nor weak receivers have cache memories; and the constant

$$\gamma_{\text{joint}} := 1 + \frac{2K_\mathrm{w}}{1+K_\mathrm{w}} \cdot \frac{K_\mathrm{s}(1-\delta_\mathrm{w})}{K_\mathrm{w}(1-\delta_\mathrm{s})} \geq 1 \tag{2}$$

describes this scheme's gain over separate cache-channel coding. This means that separate cache-channel coding can only achieve the lower bound in (1) when γ_{joint} is replaced by 1. Inequalities (1) and (2) show that the improvement of our joint cache-channel coding scheme (over separate cache-channel coding schemes) is not bounded for small cache sizes. In particular, it is strictly increasing in the number of strong receivers K_s when the other problem parameters $\delta_\mathrm{w}, \delta_\mathrm{s}$ and K_w are fixed.

We also present a general upper bound (converse) on the capacity-memory tradeoff of cache-aided broadcast networks, when the broadcast channel is stochastically degraded. Note that although the underlying broadcast channel is assumed to be stochastically degraded, the overall cache-aided network

[1] There is a subtle difference between piggyback coding and Tuncel coding in this special case. In both coding schemes the receivers decode the transmitted sources (messages) by means of a joint typicality test involving their observed channel outputs and a *subset* (called bin) of the channel codewords. Piggyback coding uses a standard binning structure where the channel codebook is explicitly partitioned into such bins. In Tuncel coding, the bins are not explicit, do not necessarily form a partition, and can be different over the various receivers. The bin of interest is therefore created on-the-fly (through another joint typicality test between source sequences and side-information) by each receiver just before the actual decoding operation.

is not stochastically degraded because the caches can be used to improve the decoding capabilities of the weaker receivers. Consequently, one cannot directly apply known converse bounds for degraded broadcast channels (e.g. [60]). The derived upper bound holds for any cache sizes.

The lower and upper bounds match for the network in Figure 1 when there is only a single weak receiver with a small cache memory:

$$K_{\mathrm{w}} = 1$$

and

$$M \leq D \frac{(1-\delta_{\mathrm{s}})(\delta_{\mathrm{w}}-\delta_{\mathrm{s}})}{K_{\mathrm{s}}(1-\delta_{\mathrm{w}}) + (1-\delta_{\mathrm{s}})}, \tag{3}$$

where D denotes the number of files in the system.

For the special case $K_{\mathrm{w}} = K_{\mathrm{s}} = 1$ and $D = 2$, a second, improved, set of lower and upper bounds on $C(M)$ is presented. They coincide when the cache memory is either small as in (3) or large:

$$M \geq ((1-\delta_{\mathrm{s}}) + (\delta_{\mathrm{w}} - \delta_{\mathrm{s}})),$$

and also for general cache memories $M \geq 0$ when both receivers are equally strong:

$$\delta_{\mathrm{w}} = \delta_{\mathrm{s}}.$$

To facilitate comparison with previous works, the lower and upper bounds on the capacity-memory tradeoff $C(M)$ are also translated into equivalent bounds on the minimum delivery rate-memory tradeoff, as considered in [2].

As already mentioned, the present paper shows the benefits of joint cache-channel coding over schemes that apply cache-channel separation. A different kind of separation was addressed in [63]: Cache-placement and creation of coded (XOR) messages are separated from the delivery of these coded messages over the network using a stand-alone code that ignores the caches. It is proved that this separation-based architecture is near-optimal for any network (i.e., achieves the optimal delivery rate-memory tradeoff up to a constant factor) when restricting to separate cache-channel codes.

A. Paper Organization

The paper is organized as follows. In Section II, we state the problem setup. Section III presents a separate-cache channel coding scheme and Section IV describes our joint cache-channel coding scheme. Section V gives a fundamental converse result (upper bound) for the capacity-memory tradeoff of general degraded BCs with receiver cache-memories. Section VI states lower and upper bounds on the capacity memory tradeoff of the erasure BC in Figure 1. Section VII restates the obtained general upper and lower bounds on the capacity-memory tradeoff as lower and upper bounds on the delivery rate-memory tradeoff. Finally, Section VIII concludes the paper.

II. Problem Definition

A. Notation

Random variables are denoted by uppercase letters, e.g. A, their alphabets by matching calligraphic font, e.g. $\mathcal{A}$, and their realizations by lowercase letters, e.g. a. We also use uppercase letters for deterministic quantities like rate R, capacity C, number of users K, cache size M, and the number of files in the library D. 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 $\mathcal{A}^n$. The shorthand notation A^n is used for the tuple $(A_1, \dots, A_n)$.

LHS and RHS stand for left-hand side and right-hand side, and IID stands for independentely and identically distributed.

Finally, the notation $W_1 \oplus W_2$ denotes the bitwise XOR of the binary strings that correspond to the messages W_1 and W_2, where these strings are assumed to be of equal length.

B. Message and Channel Models

Consider the one-to-K BC in Figure 1. We have two sets of receivers: K_{w} weak receivers with *bad* channels and $K_{\mathrm{s}} = K - K_{\mathrm{w}}$ strong receivers with *good* channels. (The meaning of good and bad channels will be explained shortly.) For convenience of notation, we assume that the first K_{w} receivers are weak and the subsequent K_{s} receivers are strong, and we define the following sets accordingly:

$$\begin{aligned} \mathcal{K}_{\mathrm{w}} &:= \{1, \dots, K_{\mathrm{w}}\}, \\ \mathcal{K}_{\mathrm{s}} &:= \{K_{\mathrm{w}}+1, \dots, K\}, \\ \mathcal{K} &:= \{1, \dots, K\}. \end{aligned}$$

We model the channel from the transmitter to the receivers by a memoryless *erasure BC*[2] with input alphabet

$$\mathcal{X} := \{0, 1\}$$

and equal output alphabet at all receivers

$$\mathcal{Y} := \mathcal{X} \cup \{\Delta\}.$$

The output erasure symbol Δ models loss of a bit at a given receiver. Receiver $k \in \mathcal{K}$ observes the erasure symbol Δ with a given probability $\delta_k \geq 0$, and it observes an output y_k equal to the input, $y_k = x$, with probability $1 - \delta_k$. The marginal transition laws[3] of the memoryless BC are thus described by

$$\mathbb{P}[Y_k = y_k | X = x] = \begin{cases} 1-\delta_k & \text{if } y_k = x \\ \delta_k & \text{if } y_k = \Delta \end{cases} \quad \forall\, k \in \mathcal{K}. \tag{4}$$

We will assume throughout that

$$\delta_i = \begin{cases} \delta_{\mathrm{w}} & \text{if } i \in \mathcal{K}_{\mathrm{w}} \\ \delta_{\mathrm{s}} & \text{if } i \in \mathcal{K}_{\mathrm{s}} \end{cases} \tag{5}$$

[2]The results in this paper extend readily to packet-erasure BCs. It suffices to scale the message rate R and the cache size M defined in the following by the packet size.

[3]We need only specify the individual marginal probability channel laws $\mathbb{P}[Y_k = y_k | X = x]$ for $k \in \mathcal{K}$ (and not the entire joint channel law $\mathbb{P}[Y_1 = y_1, Y_2 = y_2, ..., Y_K = y_K | X = x]$) for the optimal cache-memory tradeoff problem defined shortly. That is, it will become clear in the following discussion that any two channels with the same marginal channel laws, but different joint laws, exhibit the same fundamental cache-memory tradeoffs.

for fixed erasure probabilities[4]

$$0 \leq \delta_{\mathrm{s}} \leq \delta_w < 1. \tag{6}$$

Since $\delta_{\mathrm{s}} \leq \delta_w$, the weak receivers have statistically worse channels than the strong receivers, hence the distinction between good and bad channels. In the sequel, we will assume that each weak receiver is provided with a cache memory of size nM bits. The strong receivers are not provided with cache memories.

C. Message Library and Receiver Demands

The transmitter has access to a library with $D \geq K$ independent messages

$$W_1, \ldots, W_D. \tag{7}$$

Each message is uniformly distributed over the message set $\{1, \ldots, \lfloor 2^{nR} \rfloor\}$, where $R \geq 0$ is the rate of the message and n is the blocklength of transmission.

Each receiver demands (i.e., requests and downloads) exactly one of the messages. Let

$$\mathcal{D} := \{1, \ldots, D\}.$$

We denote the demand of receiver 1 by $d_1 \in \mathcal{D}$, the demand of receiver 2 by $d_2 \in \mathcal{D}$, etc., to indicate that receiver 1 desires message W_{d_1}, receiver 2 desires message W_{d_2}, and so on. We assume throughout that the resulting demand vector

$$\mathbf{d} := (d_1, \ldots, d_K) \tag{8}$$

can take on any value in

$$\mathcal{D}^K = \mathcal{D} \times \cdots \times \mathcal{D} \quad (K\text{-fold Cartesian product}). \tag{9}$$

Communication takes place in two phases: A first *placement phase* where information is stored in the weak receivers' cache memories and a subsequent *delivery phase* where the demanded messages are delivered to all the receivers. The next two subsections detail these two communication phases.

D. Placement Phase

During the placement phase, the transmitter sends caching information V_i to each weak receiver $i \in \mathcal{K}_{\mathrm{w}}$. This information is stored in the local cache memory of the user. The strong receivers do not take part in the placement phase. Note that the users' demand vector $\mathbf{d}$ is unknown to the transmitter and receivers during the placement phase, and, therefore, the cached information V_i cannot depend on $\mathbf{d}$. Instead, V_i is a function of the entire library only:

$$V_i := g_i(W_1, \ldots, W_D), \qquad i \in \mathcal{K}_{\mathrm{w}},$$

for some function

$$g_i \colon \left\{1, \ldots, \lfloor 2^{nR} \rfloor\right\}^D \to \mathcal{V}, \qquad i \in \mathcal{K}_{\mathrm{w}}, \tag{10}$$

where

$$\mathcal{V} := \{1, \ldots, \lfloor 2^{nM} \rfloor\}.$$

[4]Although we are technically allowing $\delta_s = \delta_w$, our main interest will be $\delta_{\mathrm{s}} < \delta_{\mathrm{w}}$.

The placement phase occurs during a low-congestion period. We therefore assume that any transmission errors are corrected using, for example, retransmissions. Each weak receiver $i \in \mathcal{K}_{\mathrm{w}}$ can thus store V_i in its cache memory.

E. Delivery Phase

In the delivery phase, users' demands are announced. The transmitter is provided with the demand vector $\mathbf{d}$, and it communicates the corresponding messages $W_{d_1}, \ldots, W_{d_K}$ over the erasure BC. The entire demand vector $\mathbf{d}$ is assumed to be known to the transmitter and all receivers[5].

The transmitter chooses an encoding function that corresponds to the specific demand vector $\mathbf{d}$:

$$f_{\mathbf{d}} \colon \{1, \ldots, \lfloor 2^{nR} \rfloor\}^D \to \mathcal{X}^n, \tag{11}$$

and it sends

$$X^n = f_{\mathbf{d}}(W_1, \ldots, W_D) \tag{12}$$

over the erasure BC.

Each receiver $k \in \mathcal{K}$ observes Y_k^n according to the memoryless transition law in (4). The weak receivers attempt to reconstruct their desired messages from their channel outputs, local cache contents, and the demand vector $\mathbf{d}$. Similarly, the strong receivers attempt to reconstruct their desired messages from their channel outputs and the demand vector $\mathbf{d}$. More formally,

$$\hat{W}_i := \begin{cases} \varphi_{i,\mathbf{d}}(Y_i^n, V_i) & \text{if } i \in \mathcal{K}_{\mathrm{w}} \\ \varphi_{i,\mathbf{d}}(Y_i^n) & \text{if } i \in \mathcal{K}_{\mathrm{s}}, \end{cases} \tag{13a}$$

where

$$\varphi_{i,\mathbf{d}} \colon \mathcal{Y}^n \times \mathcal{V} \to \{1, \ldots, \lfloor 2^{nR} \rfloor\}, \qquad i \in \mathcal{K}_{\mathrm{w}}, \tag{13b}$$

and

$$\varphi_{i,\mathbf{d}} \colon \mathcal{Y}^n \to \{1, \ldots, \lfloor 2^{nR} \rfloor\}, \qquad i \in \mathcal{K}_{\mathrm{s}}. \tag{13c}$$

F. Capacity-Memory Tradeoff

An error occurs whenever

$$\hat{W}_k \neq W_{d_k} \quad \text{for some} \quad k \in \mathcal{K}. \tag{14}$$

For a given demand vector $\mathbf{d}$, the probability of error is

$$\mathsf{P}_{\mathrm{e}}(\mathbf{d}) := \mathbb{P}\left[\bigcup_{k=1}^{K} \hat{W}_k \neq W_{d_k} \right]. \tag{15}$$

We consider a worst-case probability of error over all feasible demand vectors:

$$\mathsf{P}_{\mathrm{e}}^{\text{worst}} := \max_{\mathbf{d} \in \mathcal{D}^K} \mathsf{P}_{\mathrm{e}}(\mathbf{d}). \tag{16}$$

In Definitions (10)–(16), we sometimes add a superscript (n) to emphasise the dependency on the blocklength n.

[5]It takes only $\lceil \log(D) \rceil$ bits to describe the demand vector $\mathbf{d}$. The demand vector can thus be revealed to all terminals using zero-transmission rate (in the usual Shannon sense).

We say that a rate-memory pair (R, M) is *achievable* if, for every $\epsilon > 0$, there exists a sufficiently large blocklength n and placement, encoding and decoding functions as in (10), (11) and (13) such that $\mathsf{P}_e^{\text{worst}} < \epsilon$. The main problem of interest in this paper is to determine the following capacity versus cache-memory tradeoff.

Definition 1: Given the cache memory size M, we define the *capacity-memory tradeoff* $C(M)$ as the supremum of all rates R such that the rate-memory pair (R, M) is achievable.

G. Trivial and Non-Trivial Cache Sizes

When the cache size $M = 0$, the capacity-memory tradeoff equals the symmetric capacity R_0 of a standard erasure BC [59]:

$$C(M = 0) = R_0, \tag{17}$$

where

$$R_0 := \left(\frac{K_\text{w}}{1-\delta_\text{w}} + \frac{K_\text{s}}{1-\delta_\text{s}} \right)^{-1}. \tag{18}$$

Since the strong receivers do not have cache memories, the capacity-memory tradeoff cannot exceed the capacity to the strong receivers, irrespective of the cache size at the weak receivers. Thus,

$$C(M) \leq \frac{1-\delta_\text{s}}{K_\text{s}}, \quad \forall\, M \geq 0. \tag{19}$$

When $M \geq D(1-\delta_\text{s})/K_\text{s}$, the weak receivers can store the entire library in their caches and the transmitter thus needs to only serve the strong receivers during the delivery phase. Therefore,

$$C(M) = \frac{1-\delta_\text{s}}{K_\text{s}}, \qquad \forall\, M \geq D \cdot \frac{1-\delta_\text{s}}{K_\text{s}}. \tag{20}$$

We henceforth restrict attention to nontrivial cache memories

$$M \in \left(0, \ D \cdot \frac{1-\delta_\text{s}}{K_\text{s}}\right).$$

III. A Separate Cache-Channel Coding Scheme

As first step, consider the following separate cache-channel coding scheme that is built on stand-alone capacity-achieving codes for the erasure BC and the coded caching scheme of Maddah-Ali and Niesen [2]. The scheme is described in detail using the following coded-caching methods, which will serve also in later sections:

- Method `Ca` describes the placement operations.
- Method `En` describes the delivery-phase encoding.
- Methods $\{\texttt{De}_i;\ i = 1, 2, \ldots, K_\text{w}\}$ describe the delivery-phase decodings.

1) Preliminaries: The scheme has parameter

$$\tilde{t} \in \{0, \ldots, K_\text{w}\}.$$

If $\tilde{t} \neq 0$, let

$$\mathcal{G}_1, \ldots, \mathcal{G}_{\binom{K_\text{w}}{\tilde{t}}}$$

denote the $\binom{K_\text{w}}{\tilde{t}}$ subsets of $\mathcal{K}_\text{w}$ that have size $\tilde{t}$. Split each message W_d, $d = 1, \ldots, D$, into $\binom{K_\text{w}}{\tilde{t}}$ independent submessages

$$W_d = \left\{ W_{d,\mathcal{G}_\ell} \colon\ \ell = 1, \ldots, \binom{K_\text{w}}{\tilde{t}} \right\},$$

each with equal rate

$$R_\text{sub} := R \binom{K_\text{w}}{\tilde{t}}^{-1}. \tag{21}$$

2) Placement: For $\tilde{t} = 0$, no content is stored in the caches. For $\tilde{t} \neq 0$, cache placement is performed using Method `Ca` as follows:

> Method `Ca` takes as input the entire library $W_1, \ldots, W_D$ and outputs for each $i \in \mathcal{K}_\text{w}$ the cache content
>
> $$V_i = \left\{ W_{d,\mathcal{G}_\ell} \colon\ \ d \in \mathcal{D} \text{ and } \ell \in \left\{1, \ldots, \binom{K_w}{\tilde{t}}\right\},\ \ i \in \mathcal{G}_\ell \right\}. \tag{22}$$

In other words, when $\tilde{t} \neq 0$, then at any given weak Receiver $i \in \mathcal{K}_\text{w}$, the procedure stores all the tuples

$$\big(W_{1,\mathcal{G}_\ell},\ W_{2,\mathcal{G}_\ell},\ \ldots,\ W_{D,\mathcal{G}_\ell} \big)$$

for which $i \in \mathcal{G}_\ell$.

3) Delivery-Encoding: If $\tilde{t} = 0$, no contents have been stored in the cache memories, and the transmitter simply sends messages

$$W_{d_1}, W_{d_2}, \ldots, W_{d_K}$$

to the intended receivers using a capacity-achieving scheme for the erasure BC.

If $\tilde{t} = K_\text{w}$, the weak receivers can directly retrieve their desired messages from their cache memories. The transmitter thus only needs to send Messages

$$W_{d_{K_\text{w}+1}},\ W_{d_{K_\text{w}+2}},\ \ldots,\ W_{d_K} \tag{23}$$

to the strong receivers using a capacity-achieving code.

If $0 < \tilde{t} < K_\text{w}$, the transmitter first applies the following Method `En`:

> Method `En` takes as inputs the weak receivers' demand vector $\mathbf{d}_\text{w} := (d_1, \ldots, d_{K_\text{w}})$ and the messages $W_1, \ldots, W_D$. It produces the outputs
>
> $$\big\{ W_{\text{XOR},\mathcal{S}} \colon \quad \mathcal{S} \subseteq \mathcal{K}_\text{w} \text{ and } |\mathcal{S}| = \tilde{t} + 1 \big\}, \tag{24}$$
>
> where
>
> $$W_{\text{XOR},\mathcal{S}} := \bigoplus_{k \in \mathcal{S}} W_{d_k, \mathcal{S}\setminus\{k\}}. \tag{25}$$

The transmitter then uses a capacity-achieving scheme for erasure BCs to send the messages in (25) to all[6] weak receivers in $\mathcal{K}_\text{w}$ and the messages in (23) to all strong receivers $\mathcal{K}_\text{s}$.

[6] Since they have equal channel statistics, all weak receivers can decode the same messages. A similar observation applies for the strong receivers.

4) Delivery-Decoding: The strong receivers decode their intended messages in (23) using a capacity-achieving decoder for the erasure BC.

If $\tilde{t} = 0$, the weak receivers decode in the same way as the strong receivers.

If $\tilde{t} = K_\mathrm{w}$, the weak receivers can directly retrieve their desired messages from their cache memories.

If $1 \leq \tilde{t} \leq K_\mathrm{w} - 1$, the weak receivers first decode the XOR-messages $\left\{\hat{W}_{\mathrm{XOR},\mathcal{S}} \colon \mathcal{S} \subseteq \mathcal{K}_\mathrm{w},\ |\mathcal{S}| = \tilde{t}+1\right\}$ in (24), using a capacity-achieving decoder for the erasure BC. Each receiver $i \in \mathcal{K}_\mathrm{w}$ then applies the following Method $\mathtt{De}_i$:

Method $\mathtt{De}_i$ takes as inputs the demand vector $\mathbf{d}_\mathrm{w}$, the decoded messages $\{\hat{W}_{\mathrm{XOR},\mathcal{S}} \colon i \in \mathcal{S}\}$, see (24), and the cache content V_i. It outputs the reconstruction

$$\hat{W}_i := \left(\hat{W}_{d_i,\mathcal{G}_1}, \ldots, \hat{W}_{d_i,\mathcal{G}_{\binom{K_\mathrm{w}}{\tilde{t}}}}\right), \tag{26}$$

where

$$\hat{W}_{d_i,\mathcal{G}_\ell} = \begin{cases} \left(\bigoplus_{s\in\mathcal{G}_\ell} W_{d_s,\mathcal{G}_\ell\cup\{i\}\setminus\{s\}}\right) \\ \oplus \hat{W}_{\mathrm{XOR},\mathcal{G}_\ell\cup\{i\}} & \text{if } i \notin \mathcal{G}_\ell, \\ W_{d_i,\mathcal{G}_\ell} & \text{if } i \in \mathcal{G}_\ell. \end{cases} \tag{27}$$

5) Analysis: By standard arguments, for a given parameter $\tilde{t} \in \mathcal{K}_\mathrm{w}$, the described separate cache-channel coding scheme allows for vanishing probability of error whenever the rate does not exceed

$$R_{\tilde{t},\mathrm{sep}} := \left(\frac{K_\mathrm{w} - \tilde{t}}{(\tilde{t}+1)(1-\delta_\mathrm{w})} + \frac{K_\mathrm{s}}{(1-\delta_\mathrm{s})}\right)^{-1}. \tag{28a}$$

Moreover, the scheme requires the weak receivers to have cache memories of size

$$M_{\tilde{t},\mathrm{sep}} := D\frac{\tilde{t}}{K_\mathrm{w}} R_{\tilde{t},\mathrm{sep}}. \tag{28b}$$

By time- and memory-sharing arguments, the following proposition holds.

Proposition 1: The upper convex hull of the rate-memory pairs in (28) is achievable:

$$C(M) \geq \text{upp hull}\left(\left\{(R_{\tilde{t},\mathrm{sep}}, M_{\tilde{t},\mathrm{sep}});\quad \tilde{t} \in \{0,\ldots,K_\mathrm{w}\}\right\}\right). \tag{29}$$

IV. A Joint Cache-Channel Coding Scheme

This section describes a joint cache-channel coding scheme for the broadcast network in Figure 1. The general scheme is parameterized by a positive integer $t \in \mathcal{K}_\mathrm{w}$ where $t+1$ will be the number of weak receivers that can be simultaneously served by each transmission.

A. A Simple Example

To better illustrate the ideas, we start with an example. Consider the scenario in Figure 2 with $K_\mathrm{w} = 3$ weak receivers and $K_\mathrm{s} = 1$ strong receivers. We describe the scheme that corresponds to $t = 2$.

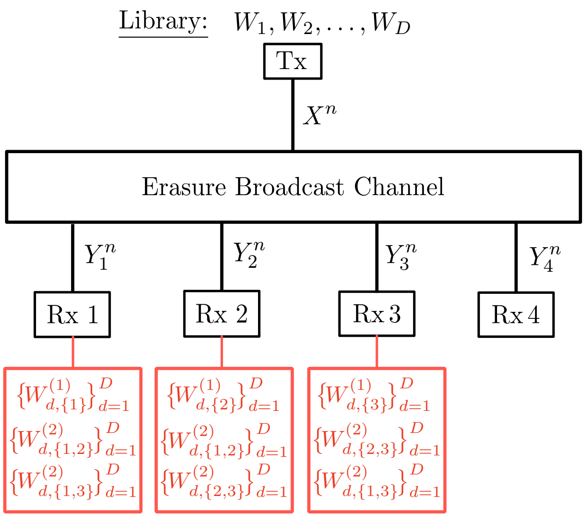

Fig. 2: An example network with 3 weak and a single strong receiver. The figure illustrates the contents cached in the proposed joint cache-channel coding scheme when $t = 2$.

1) Scheme: Define

$$R^{(1)} := \frac{1-\delta_\mathrm{w}}{1-\delta_\mathrm{s}} R, \tag{30a}$$

$$R^{(2)} := \frac{\delta_\mathrm{w}-\delta_\mathrm{s}}{1-\delta_\mathrm{s}} R, \tag{30b}$$

where, as before, R denotes the common rate of the messages. Note that the ratio of $R^{(1)}$ to $R^{(2)}$ simplifies to

$$\frac{R^{(1)}}{R^{(2)}} = \frac{1-\delta_\mathrm{w}}{\delta_\mathrm{w}-\delta_\mathrm{s}}. \tag{31}$$

Split each message W_d into two submessages

$$W_d = \left(W_d^{(1)},\ W_d^{(2)}\right)$$

of rates $R^{(1)}$ and $R^{(2)}$. Further split each submessage $W_d^{(1)}$ into 3 parts:

$$W_d^{(1)} = \left(W^{(1)}_{d,\{1\}}, W^{(1)}_{d,\{2\}}, W^{(1)}_{d,\{3\}}\right)$$

of equal rates $R^{(1)}/3$, and each submessage $W_d^{(2)}$ into 3 parts

$$W_d^{(2)} = \left(W^{(2)}_{d,\{1,2\}}, W^{(2)}_{d,\{1,3\}}, W^{(2)}_{d,\{2,3\}}\right)$$

of equal rates $R^{(2)}/3$.

Placement Phase: Cache all messages $\left\{W^{(1)}_{d,\{i\}}\right\}_{d=1}^D$ at receiver i, for $i \in \{1,2,3\}$, and all messages $\left\{W^{(2)}_{d,\{i,j\}}\right\}_{d=1}^D$ at receivers i and j, for $i,j \in \{1,2,3\}$ with $i \neq j$. The cache contents are shown in Figure 2.

The placement phase is a two-fold application of the coded-caching method $\mathtt{Ca}$ given in the previous Section III: First apply method $\mathtt{Ca}$ with parameter $\tilde{t} = 1$ to messages $W_1^{(1)}, \ldots, W_D^{(1)}$, and then apply the same method with parameter $\tilde{t} = 2$ to messages $W_1^{(2)}, \ldots, W_D^{(2)}$.

Delivery Phase: Delivery transmission takes place in Subphases 1–3 of lengths $n_1, n_2, n_3 \geq 0$ that sum up to the entire blocklength n.[7] Table I shows the messages that are transmitted in the various subphases. Notice that given the cache contents in Figure 2, each receiver can recover its desired message without error, if the submessages in Table I are correctly decoded by the appropriate receivers.

We now explain the transmissions in the three subphases in detail.

Subphase 1 is dedicated solely to the weak receivers. The transmitter sends the XOR-message

$$W_{\text{XOR},\{1,2,3\}} := W^{(2)}_{d_3,\{1,2\}} \oplus W^{(2)}_{d_2,\{1,3\}} \oplus W^{(2)}_{d_1,\{2,3\}}$$

to receivers 1–3 using a capacity-achieving scheme for the erasure BC to these three receivers. At the end of this first subphase, receiver 1 decodes an estimate for the XOR-message $W_{\text{XOR},\{1,2,3\}}$, denoted by $\hat{W}_{\text{XOR},\{1,2,3\}}$. It then retrieves messages $W^{(2)}_{d_3,\{1,2\}}$ and $W^{(2)}_{d_2,\{1,3\}}$ from its cache memory and produces:

$$\hat{W}^{(2)}_{d_1,\{2,3\}} := W^{(2)}_{d_3,\{1,2\}} \oplus W^{(2)}_{d_2,\{1,3\}} \oplus \hat{W}_{\text{XOR},\{1,2,3\}}. \tag{32}$$

Receivers 2 and 3 produce $\hat{W}^{(2)}_{d_2,\{1,3\}}$ and $\hat{W}^{(2)}_{d_3,\{1,2\}}$, following similar steps.

Subphase 2: The second subphase is the most interesting one, and is where we utilize joint cache-channel coding (namely in the decoding at the weak receivers). It is divided into three length-$\lfloor n_2/3 \rfloor$ periods, which we index by $\{1,2\}, \{1,3\}, \{2,3\}$ (i.e. by the subsets of $\{1,2,3\}$ of size 2).

In period $\{i,j\}$, the XOR-message

$$W_{\text{XOR},\{i,j\}} := W^{(1)}_{d_j,\{i\}} \oplus W^{(1)}_{d_i,\{j\}} \tag{33a}$$

is sent as a common message to the weak receivers i and j, and at the same time Message

$$W^{(2)}_{d_4,\{i,j\}} \tag{33b}$$

is sent to the only strong receiver 4. Notice that this latter message $W^{(2)}_{d_4,\{i,j\}}$ is stored in the cache memories of both weak receivers i and j.

For the transmission of the messages in (33), a codebook $\mathcal{C}_{i,j}$ with $\lfloor 2^{nR^{(2)}/3} \rfloor \times \lfloor 2^{nR^{(1)}/3} \rfloor$ codewords of length $n_{2,\text{per}} := \lfloor n_2/3 \rfloor$ is generated by randomly and independently drawing each entry according to a Bernoulli-$1/2$ distribution.

The codewords of $\mathcal{C}_{i,j}$ are arranged in an array with $\lfloor 2^{nR^{(2)}/3} \rfloor$ rows and $\lfloor 2^{nR^{(1)}/3} \rfloor$ columns, as depicted in Figure 3, where each dot illustrates a codeword. We refer to the codeword in row w_{row} and column w_{column} as $x^{n_{2,\text{per}}}_{i,j}(w_{\text{row}},\ w_{\text{column}})$. The codebook $\mathcal{C}_{i,j}$ is revealed to all parties.

The transmitter sends the codeword

$$x^{n_{2,\text{per}}}_{i,j}\left(W^{(2)}_{d_4,\{i,j\}},\ W_{\text{XOR},\{i,j\}}\right)$$

over the channel.

[7]Specifically, throughout the following we take n_1, n_2, and n_3 to be functions of n such that $n_1(n) + n_2(n) + n_3(n) = n$. To simplify notation we will not explicitly write the dependence of n_1, n_2, and n_3 on n.

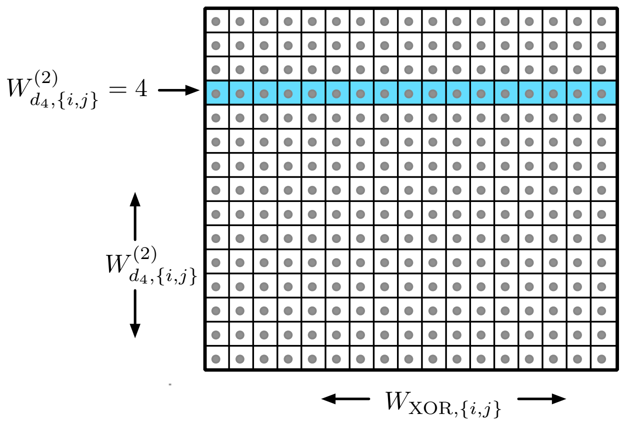

Fig. 3: Codebook used for piggyback coding in period $\{i,j\}$ of Subphase 2.

The strong receiver 4 decodes both messages $W_{\text{XOR},\{i,j\}}$ and $W^{(2)}_{d_4,\{i,j\}}$ using a standard decoder and in particular produces the estimate $\hat{W}^{(2)}_{d_4,\{i,j\}}$.

The weak receiver i, however, decodes in three steps. It first retrieves $W^{(2)}_{d_4,\{i,j\}}$ from its cache memory and extracts the *row-codebook* $\mathcal{C}_{i,j,\text{row}}\big(W^{(2)}_{d_4,\{i,j\}}\big)$ from $\mathcal{C}_{i,j}$:

$$\mathcal{C}_{i,j,\text{row}}\left(W^{(2)}_{d_4,\{i,j\}}\right) := \left\{x^{n_{2,\text{per}}}_{i,j}\left(W^{(2)}_{d_4,\{i,j\}},\ w\right)\right\}_{w=1}^{\lfloor 2^{nR^{(1)}/3} \rfloor}. \tag{34}$$

(For example, the blue row in Figure 3 indicates the row-codebook $\mathcal{C}_{i,j,\text{row}}\big(W^{(2)}_{d_4,\{i,j\}}\big)$ to consider when $W^{(2)}_{d_4,\{i,j\}} = 4$.) Receiver i then decodes the XOR-message $W_{\text{XOR},\{i,j\}}$ by restricting attention to the codewords in $\mathcal{C}_{i,j,\text{row}}\big(W^{(2)}_{d_4,\{i,j\}}\big)$, and uses its estimate $\hat{W}_{\text{XOR},\{i,j\}}$ to form

$$\hat{W}^{(1)}_{d_i,\{j\}} := W^{(1)}_{d_j,\{i\}} \oplus \hat{W}_{\text{XOR},\{i,j\}}.$$

Receiver j produces $\hat{W}^{(1)}_{d_j,\{i\}}$ following similar steps.

Subphase 3: Message $W^{(1)}_{d_4}$ is sent to receiver 4 using a capacity-achieving scheme for this receiver. At the end of Subphase 3, receiver 4 applies a standard decoder and produces the estimate $\hat{W}^{(1)}_{d_4}$.

Final decoding: Receivers 1–4 finally declare, respectively:

$$\hat{W}_1 := \Big(W^{(1)}_{d_1,\{1\}}, \hat{W}^{(1)}_{d_1,\{2\}}, \hat{W}^{(1)}_{d_1,\{3\}}, W^{(2)}_{d_1,\{1,2\}}, W^{(2)}_{d_1,\{1,3\}}, \hat{W}^{(2)}_{d_1,\{2,3\}}\Big); \tag{35}$$

$$\hat{W}_2 := \Big(\hat{W}^{(1)}_{d_2,\{1\}}, W^{(1)}_{d_2,\{2\}}, \hat{W}^{(1)}_{d_2,\{3\}}, W^{(2)}_{d_2,\{1,2\}}, \hat{W}^{(2)}_{d_2,\{1,3\}}, W^{(2)}_{d_2,\{2,3\}}\Big); \tag{36}$$

$$\hat{W}_3 := \Big(\hat{W}^{(1)}_{d_3,\{1\}}, \hat{W}^{(1)}_{d_3,\{2\}}, W^{(1)}_{d_3,\{3\}}, \hat{W}^{(2)}_{d_3,\{1,2\}}, W^{(2)}_{d_3,\{1,3\}}, W^{(2)}_{d_3,\{2,3\}}\Big); \tag{37}$$

$$\hat{W}_4 := \Big(\hat{W}^{(1)}_{d_4}, \hat{W}^{(2)}_{d_4,\{1,2\}}, \hat{W}^{(2)}_{d_4,\{1,3\}}, \hat{W}^{(2)}_{d_4,\{2,3\}}\Big). \tag{38}$$

	Subphase 1	Subphase 2	Subphase 3
Messages sent to Rxs $1, 2, 3$	$W^{(2)}_{d_3,\{1,2\}} \oplus W^{(2)}_{d_2,\{1,3\}} \oplus W^{(2)}_{d_1,\{2,3\}}$	$W^{(1)}_{d_1,\{2\}} \oplus W^{(1)}_{d_2,\{1\}}$ $W^{(1)}_{d_1,\{3\}} \oplus W^{(1)}_{d_3,\{1\}}$ $W^{(1)}_{d_2,\{3\}} \oplus W^{(1)}_{d_3,\{2\}}$	
Messages sent to Rx 4		$W^{(2)}_{d_4}$	$W^{(1)}_{d_4}$

TABLE I: Table indicating the messages sent in the three subphases of the delivery phase.

2) Analysis: The cache memory required for this scheme is:

$$M = D\left(\frac{1}{3}R^{(1)} + \frac{2}{3}R^{(2)}\right) = D \cdot \frac{1-\delta_{\text{s}}+\delta_{\text{w}}-\delta_{\text{s}}}{3(1-\delta_{\text{s}})}R \tag{39}$$

The probability of error in Subphase 1can be made arbitrarily small by choosing n sufficiently large, because

$$\frac{n}{n_1} \cdot \frac{1}{3}R^{(2)} < 1-\delta_{\text{w}}. \tag{40}$$

The probability of error of each period in Subphase 2 can be made arbitrarily small by choosing n sufficiently large, because

$$\frac{n}{n_2/3} \cdot \frac{1}{3}R^{(1)} < 1-\delta_{\text{w}}, \tag{41}$$

and

$$\frac{n}{n_2/3} \cdot \frac{1}{3}\big(R^{(1)}+R^{(2)}\big) < 1-\delta_{\text{s}}. \tag{42}$$

Here, because weak receivers decode their desired messages based on a row-codebook containing only $\lfloor 2^{nR^{(1)}/3} \rfloor$ codewords, Inequality (41) ensures that the probability of decoding error at the weak receivers can be made arbitrarily small. Inequality (42) ensures that the probability of decoding error at the strong receiver can be made arbitrarily small. By (31), Inequalites (41) and (42) are equivalent, and we drop (42) in the following.

The probability of error in Subphase 3 can be made arbitrarily small by choosing n sufficiently large, because

$$\frac{n}{n_3}R^{(1)} < 1-\delta_{\text{s}}. \tag{43}$$

Recall that $n_1+n_2+n_3 = n$. Therefore, when

$$\frac{R^{(2)}}{3(1-\delta_{\text{w}})} + \frac{R^{(1)}}{1-\delta_{\text{w}}} + \frac{R^{(1)}}{1-\delta_s} < 1, \tag{44}$$

there exist appropriate choices of the lengths n_1, n_2, n_3 (as a function of the total blocklength n) so that (39)–(43) are satisfied, and as a consequence the probability of decoding error can be made arbitrarily small by choosing the blocklength n sufficiently large.

Finally, notice that by (30), Inequality (44) is equivalent to

$$R \le (1-\delta_{\text{s}})\left(\frac{\delta_{\text{w}}-\delta_{\text{s}}}{3(1-\delta_{\text{w}})} + 1 + \frac{1-\delta_{\text{w}}}{1-\delta_{\text{s}}}\right)^{-1}, \tag{45}$$

and the described scheme achieves any rate $R > 0$ satisfying (45).

3) Discussion: Thanks to the weak receivers' cache information, in Subphase 2, messages

$$W^{(2)}_{d_4,\{1,2\}}, W^{(2)}_{d_4,\{1,3\}}, W^{(2)}_{d_4,\{2,3\}} \tag{46}$$

can be piggybacked on the communications of the XOR messages

$$W^{(2)}_{\text{XOR},\{1,2\}}, W^{(2)}_{\text{XOR},\{1,3\}}, W^{(2)}_{\text{XOR},\{2,3\}} \tag{47}$$

without harming the performance of the latter. In fact, by our choice in (31), the probability of error in Subphase 2 can be arbitrarily small whenever Inequality (41) holds, which coincides with the required condition when solely the messages in (47) are transmitted but not the messages in (46).

We remark that if in Subphase 2, the weak receivers apply separate cache-channel decoding, the performance of the scheme would be degraded. Specifically, the additional summand

$$\frac{R^{(2)}}{1-\delta_{\text{s}}}$$

would appear on the LHS of (44) (because we would have to communicate the messages in (46) separately), or equivalently, the additional summand

$$\frac{\delta_{\text{w}}-\delta_{\text{s}}}{1-\delta_{\text{s}}}$$

would appear on the RHS of (45). This additional summand comes from the communication to receiver 4. In this sense, joint cache-channel coding is beneficial also for the strong receivers (without cache memories). This explains why we term the new caching gain that is offered by our joint cache channel coding scheme as "global".

B. General Scheme

The general scheme is parameterized by a positive integer

$$t \in \mathcal{K}_{\text{w}} \tag{48}$$

and is described in the following. We show in Appendix A that, for every parameter t, this scheme achieves the rate-memory pair

$$R_t := \frac{(1-\delta_{\text{w}})\left(1+\frac{K_{\text{w}}-t+1}{tK_{\text{s}}}\,\frac{\delta_{\text{w}}-\delta_{\text{s}}}{1-\delta_{\text{w}}}\right)}{\frac{K_{\text{w}}-t+1}{t}\left(1+\frac{K_{\text{w}}-t}{(t+1)K_{\text{s}}}\,\frac{\delta_{\text{w}}-\delta_{\text{s}}}{1-\delta_{\text{w}}}\right)+K_{\text{s}}\frac{1-\delta_w}{1-\delta_{\text{s}}}}, \tag{49a}$$

$$M_t := R_t\ \frac{D}{K_\text{w}}\left(t - \left(1 + \frac{K_\text{w} - t + 1}{tK_\text{s}}\ \frac{\delta_\text{w} - \delta_\text{s}}{1 - \delta_\text{w}}\right)^{-1}\right). \quad (49b)$$

1) Preliminaries: For each $d \in \mathcal{D}$, split message W_d into two parts:

$$W_d = \left(W_d^{(t-1)},\ W_d^{(t)}\right) \quad (50)$$

of rates

$$R^{(t-1)} = R \cdot \frac{tK_\text{s}(1-\delta_\text{w})}{(K_\text{w} - t + 1)(\delta_\text{w} - \delta_\text{s}) + tK_\text{s}(1-\delta_\text{w})}, \quad (51a)$$

$$R^{(t)} = R \cdot \frac{(K_\text{w} - t + 1)(\delta_\text{w} - \delta_\text{s})}{(K_\text{w} - t + 1)(\delta_\text{w} - \delta_\text{s}) + tK_\text{s}(1-\delta_\text{w})}. \quad (51b)$$

Notice that $R^{(t-1)} + R^{(t)} = R$. We refer to $W_d^{(t-1)}$ as the "$t-1$ part" of message W_d and to $W_d^{(t)}$ as its "t part". We will see that each submessage $W_d^{(t)}$ is stored in the cache memories of t receivers and each submessage $W_d^{(t-1)}$ is stored in the cache memories of $t-1$ receivers.

2) Placement phase: First, apply the coded-caching Method `Ca` (described in Section III-2) with parameter $\tilde{t} = t$ to messages $W_1^{(t)}, \ldots, W_D^{(t)}$ to produce for each $i \in \mathcal{K}_\text{w}$ the cache content

$$V_i^{(t)} = \left\{ W_{d,\mathcal{G}_\ell^{(t)}}^{(t)} :\ d \in \mathcal{D} \text{ and } \ell \in \left\{1, \ldots, \binom{K_w}{t}\right\},\ i \in \mathcal{G}_\ell^{(t-1)} \right\}.$$

Here, $\mathcal{G}_1^{(t)}, \ldots, \mathcal{G}_{\binom{K_\text{w}}{t}}^{(t)}$ denote the $\binom{K_w}{t}$ subsets of $\mathcal{K}_\text{w}$ that have size t, and each message $W_{d,\mathcal{G}_1^{(t)}}^{(t)}, \ldots, W_{d,\mathcal{G}_{\binom{K_\text{w}}{t}}^{(t)}}^{(t)}$ are of equal rate

$$R_\text{sub}^{(t)} = R^{(t)} \cdot \binom{K_\text{w}}{t}^{-1}. \quad (52a)$$

Then, apply Method `Ca` with parameter $\tilde{t} = t-1$ to messages $W_1^{(t-1)}, \ldots, W_D^{(t-1)}$ to produce for each $i \in \mathcal{K}_\text{w}$ the cache content

$$V_i^{(t-1)} = \left\{ W_{d,\mathcal{G}_\ell^{(t-1)}}^{(t-1)} :\ d \in \mathcal{D} \text{ and } \ell \in \left\{1, \ldots, \binom{K_w}{t-1}\right\},\ i \in \mathcal{G}_\ell^{(t-1)} \right\}.$$

Here, $\mathcal{G}_1^{(t-1)}, \ldots, \mathcal{G}_{\binom{K_\text{w}}{t-1}}^{(t-1)}$ denote the $\binom{K_\text{w}}{t-1}$ subsets of $\mathcal{K}_\text{w}$ that have size $t-1$, and messages $W_{d,\mathcal{G}_\ell^{(t-1)}}^{(t-1)}$ are of rate

$$R_\text{sub}^{(t-1)} = R^{(t-1)} \cdot \binom{K_\text{w}}{t-1}^{-1}. \quad (52b)$$

For each $i \in \mathcal{K}_\text{w}$, the transmitter stores the content

$$V_i = V_i^{(t)} \cup V_i^{(t-1)} \quad (53)$$

in the cache memory of receiver i.

3) Delivery Phase: The delivery phase takes place in three subphases of lengths $n_1, n_2, n_3 \geq 0$ that sum up to the entire blocklength n.

Delivery Subphase 1: This subphase exists only if $t < K_\text{w}$. In the first subphase, the "t parts"

$$W_{d_1}^{(t)},\ W_{d_2}^{(t)},\ \ldots, W_{d_{K_\text{w}}}^{(t)}, \quad (54)$$

are communicated using a separate cache-channel coding scheme (as described in Section III), but assuming that there are no strong receivers. In fact, the strong receivers are completely ignored in this subphase. Let $\hat{W}_i^{(t)}$ denote the guess produced by the weak receiver $i \in \mathcal{K}_\text{w}$ at the end of Subphase 1.

Delivery Subphase 2: In Subphase 2, the "t parts"

$$W_{d_{K_w+1}}^{(t)},\ W_{d_{K_\text{w}+2}}^{(t)},\ \ldots, W_{d_K}^{(t)}, \quad (55)$$

are sent to the strong receivers, and the "$t-1$ parts"

$$W_{d_1}^{(t-1)},\ W_{d_2}^{(t-1)},\ \ldots, W_{d_{K_\text{w}}}^{(t-1)}, \quad (56)$$

to the weak receivers. Both communications will be done simultaneously by means of piggyback-coding. Details are as follows. The transmitter first applies the coded-caching Method `En` (see Section III-3) with parameter $\tilde{t} = t-1$ to the (weak receivers') demand vector $\mathbf{d}_\text{w} = (d_1, \ldots, d_{K_\text{w}})$ and to the messages

$$\left\{ W_{d_i}^{(t-1)} :\ i \in \mathcal{K}_\text{w} \right\}.$$

This produces the XOR-messages

$$\left\{ W_{\text{XOR},\mathcal{G}_\ell^{(t)}}^{(t-1)} :\ \ell = 1, \ldots, \binom{K_\text{w}}{t} \right\}, \quad (57)$$

which are of rate $R_\text{sub}^{(t-1)}$ (see (52b)).

Transmission takes place over $\binom{K_\text{w}}{t}$ equally-long periods. Consider period $\ell \in \left\{1, \ldots, \binom{K_\text{w}}{t}\right\}$; the other periods are similar. In period ℓ, the XOR message

$$W_{\text{XOR},\mathcal{G}_\ell^{(t)}}^{(t-1)} \quad (58a)$$

is conveyed to all the weak receivers in $\mathcal{G}_\ell^{(t)}$ and the message tuple

$$\mathbf{W}_{\ell,\text{strong}}^{(t)} := \left(W_{d_{K_\text{w}+1},\mathcal{G}_\ell^{(t)}}^{(t)},\ \ldots,\ W_{d_K,\mathcal{G}_\ell^{(t)}}^{(t)} \right) \quad (58b)$$

is conveyed to all strong receivers $\mathcal{K}_\text{s}$. Note that the entire message tuple $\mathbf{W}_{\ell,\text{strong}}^{(t)}$ is known at all the weak receivers $i \in \mathcal{G}_\ell^{(t)}$ by the way caching was done in the placement phase.

For the communication of these messages, we generate a codebook $\mathcal{C}_\ell$ with $\lfloor 2^{nK_\text{s}R_\text{sub}^{(t)}} \rfloor \times \lfloor 2^{nR_\text{sub}^{(t-1)}} \rfloor$ codewords of length $n_{2,\text{per}} := \lfloor n_2 / \binom{K_\text{w}}{t} \rfloor$ by randomly and independently drawing each entry according to a Bernoulli-1/2 distribution. Arrange the codewords in an array with $\lfloor 2^{nK_\text{s}R_\text{sub}^{(t)}} \rfloor$ rows and $\lfloor 2^{nR_\text{sub}^{(t-1)}} \rfloor$ columns, and denote the codeword in row w_row and column w_column by

$$x_\ell^{n_{2,\text{per}}}(w_\text{row},\ w_\text{column}). \quad (59)$$

Reveal the codebook $\mathcal{C}_\ell$ to all parties.

The transmitter sends the codeword

$$x_\ell^{n_{2,\text{per}}}\left(\mathbf{W}_{\ell,\text{strong}}^{(t)}, W_{\text{XOR},\mathcal{G}_\ell^{(t)}}^{(t-1)}\right)$$

over the channel.

Each strong receiver $j \in \mathcal{K}_{\mathrm{s}}$ decodes the message tuple $\mathbf{W}^{(t)}_{\ell,\text{strong}}$ as well as the message $W^{(t-1)}_{\text{XOR},\mathcal{G}^{(t)}_\ell}$, but it will further only use the estimate $\hat{W}^{(t)}_{d_j,\mathcal{G}^{(t)}_\ell}$ (its intended message), i.e., the $j-K_{\mathrm{w}}$-th component of its message estimate $\hat{\mathbf{W}}^{(t)}_{\ell,\text{strong}}$. At the end of the last period $\binom{K_{\mathrm{w}}}{t}$, the strong receiver $j \in \mathcal{K}_{\mathrm{s}}$ produces

$$\hat{W}^{(t)}_j := \left(\hat{W}^{(t)}_{d_j,\mathcal{G}^{(t)}_1}, \ldots, \hat{W}^{(t)}_{d_j,\mathcal{G}^{(t)}_{\binom{K_{\mathrm{w}}}{t}}}\right). \tag{60}$$

Each weak receiver $i \in \mathcal{G}^{(t)}_\ell$ retrieves the message tuple $\mathbf{W}^{(t)}_{\ell,\text{strong}}$ from its cache memory and constructs the corresponding row-codebook $\mathcal{C}_{\ell,\text{row}}(\mathbf{W}^{(t)}_{\ell,\text{strong}})$ from $\mathcal{C}_\ell$:

$$\begin{aligned}&\mathcal{C}_{\ell,\text{row}}\left(\mathbf{W}^{(t)}_{\ell,\text{strong}}\right)\\&:= \left\{x^{n_{2,\text{per}}}_\ell\left(\mathbf{W}^{(t)}_{\ell,\text{strong}}, W^{(t-1)}_{\text{XOR},\mathcal{G}^{(t)}_\ell}\right)\right\}_{w=1}^{\left\lfloor 2^{nR^{(t-1)}_{\text{sub}}}\right\rfloor}.\end{aligned} \tag{61}$$

It then decodes the XOR-message $W^{(t-1)}_{\text{XOR},\mathcal{G}^{(t)}_\ell}$ from its period-ℓ outputs using an optimal decoder for $\mathcal{C}_{\ell,\text{row}}(\mathbf{W}^{(t)}_{\ell,\text{strong}})$.

After the last period $\binom{K_{\mathrm{w}}}{t}$, each receiver $i \in \mathcal{K}_{\mathrm{w}}$ applies the coded-caching method $\mathtt{De}_i$ (see Section III-4) to the demand vector $\mathbf{d}_{\mathrm{w}}$, the decoded messages

$$\left\{W^{(t-1)}_{\text{XOR},\mathcal{G}^{(t)}_\ell} \colon\; i \in \mathcal{G}^{(t)}_\ell,\; \ell = 1, \ldots, \binom{K_{\mathrm{w}}}{t}\right\},$$

and the cache content $V^{(t-1)}_i$. This method outputs the desired estimates $\hat{W}^{(t-1)}_i$.

Delivery Subphase 3: The transmitter sends the "$(t-1)$ parts"

$$W^{(t-1)}_{d_{K_w+1}},\; W^{(t-1)}_{d_{K_{\mathrm{w}}+2}},\; \ldots, W^{(t-1)}_{d_K}, \tag{62}$$

to the strong receivers using a capacity-achieving code for the erasure BC. The receivers produce the estimates

$$\hat{W}^{(t-1)}_j,\; j \in \mathcal{K}_{\mathrm{s}}. \tag{63}$$

Final Decoding: At the end of the entire transmission, each receiver $k \in \mathcal{K}$ declares the following message:

$$\hat{W}_k = \left(\hat{W}^{(t-1)}_k, \hat{W}^{(t)}_k\right). \tag{64}$$

V. A Converse for General Degraded BCs

In this section, we present an upper bound on the capacity-memory tradeoff of a more general class of cache-aided broadcast networks. Specifically, we assume that each receiver $k \in \mathcal{K}$ has a cache of size nM_k bits, the BC is discrete and memoryless with input alphabet $\mathcal{X}$, output alphabets $\mathcal{Y}_1, \ldots, \mathcal{Y}_K$, and an arbitrary (stochastically) degraded channel transition law $P_{Y_1Y_2\cdots Y_K|X}(y_1, \ldots, y_K|x)$. For simplicity of exposition, we assume that the BC is *physically degraded*, i.e., the transition law satisfies the Markov chain

$$X \to Y_K \to Y_{K-1} \to \cdots \to Y_1. \tag{65}$$

The extension to stochastically degraded BCs follows trivially, because the capacity-memory tradeoff only depends on the marginal BC transition laws, see footnote 3.

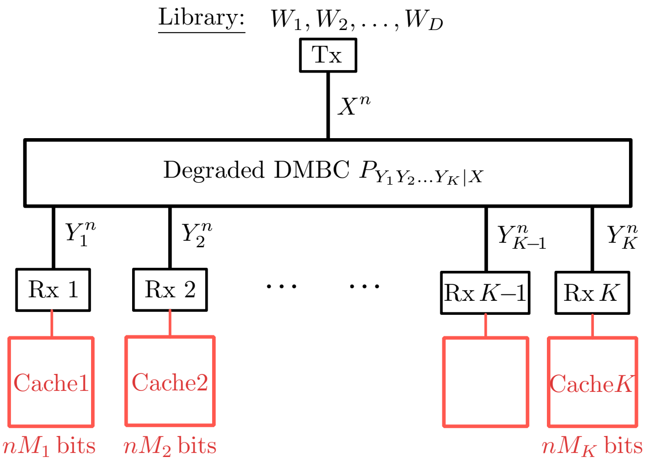

Fig. 4: Degraded K-user BC $P_{Y_1Y_2\cdots Y_K|X}$ where each receiver $k \in \mathcal{K}$ has cache memory of size nM_k bits.

The library and the probability of worst-case error $\mathsf{P}_{\mathrm{e}}^{\text{worst}}$ are defined as before. A rate-memory tuple $(R,\; M_1, \ldots, M_K)$ is said *achievable* if for every $\epsilon > 0$ there exists a sufficiently large blocklength n and placement, encoding and decoding functions as in (10)–(13) such that $\mathsf{P}_{\mathrm{e}}^{\text{worst}} < \epsilon$. The capacity-memory tradeoff $C(M_1, \ldots, M_K)$ is defined as the supremum over all rates $R > 0$ such that $(R, M_1, \ldots, M_K)$ are achievable.

For each ordered subset $\mathcal{S} = \{j_1, \ldots, j_{|\mathcal{S}|}\} \subseteq \mathcal{K}$, where

$$j_1 \le j_2 \le \ldots \le j_{|\mathcal{S}|}, \tag{66}$$

define

$$\begin{aligned}&R_{\text{sym},\mathcal{S}}\\&:= \max\min\left\{I(U_1;Y_{j_1}),\; I(U_2;Y_{j_2}|U_1),\; \ldots,\right.\\&\qquad\left. I\left(U_{|\mathcal{S}|};Y_{j_{|\mathcal{S}|-1}}\middle|U_{|\mathcal{S}|-2}\right),\; I\left(X;Y_{j_{|\mathcal{S}|}}\middle|U_{|\mathcal{S}|-1}\right)\right\},\end{aligned} \tag{67}$$

where the maximization is over all choices of the auxiliary random variables $U_1, \ldots, U_{|\mathcal{S}|-1}, X$ forming the Markov chain

$$U_1 \to U_2 \to \cdots \to U_{|\mathcal{S}|-1} \to X \to \left(Y_{j_1}, \ldots, Y_{j_{|\mathcal{S}|}}\right). \tag{68}$$

Notice that $R_{\text{sym},\mathcal{S}}$ is the largest symmetric rate that is achievable over the BC to receivers in $\mathcal{S}$ when there are no cache memories [60].

Theorem 2: The capacity-memory tradeoff $C(M_1, \ldots, M_K)$ of a degraded BC is upper bounded as:

$$C(M_1, \ldots, M_K) \le \min_{\mathcal{S}\subseteq\mathcal{K}} \left(R_{\text{sym},\mathcal{S}} + \frac{M_{\mathcal{S}}}{D}\right), \tag{69}$$

where $M_{\mathcal{S}}$ is the total cache size at receivers in $\mathcal{S}$:

$$M_{\mathcal{S}} = \sum_{k\in\mathcal{S}} M_k. \tag{70}$$

Proof: See Appendix B. ■

Remark 1: Theorem 2 also holds for stochastically degraded BCs because the capacity-memory tradeoff only depends on the marginal channel laws. Note that the erasure BC is stochastically degraded.

VI. Main Results for the Erasure Network in Figure 1

A. General Lower Bound on $C(M)$

Let

$$R_0 = \left(\frac{K_\mathrm{w}}{1-\delta_\mathrm{w}} + \frac{K_\mathrm{s}}{1-\delta_\mathrm{s}}\right)^{-1}, \qquad M_0 := 0; \tag{71}$$

and

$$R_{K_\mathrm{w}+1} := \frac{1-\delta_\mathrm{s}}{K_\mathrm{s}}, \qquad M_{K_\mathrm{w}+1} := D\,\frac{1-\delta_\mathrm{s}}{K_\mathrm{s}}; \tag{72}$$

and recall for each $t \in \mathcal{K}_\mathrm{w}$ the rate-memory pair

$$R_t = \frac{(1-\delta_\mathrm{w})\left(1 + \dfrac{K_\mathrm{w}-t+1}{tK_\mathrm{s}}\,\dfrac{\delta_\mathrm{w}-\delta_\mathrm{s}}{1-\delta_\mathrm{w}}\right)}{\dfrac{K_\mathrm{w}-t+1}{t}\left(1+\dfrac{K_\mathrm{w}-t}{(t+1)K_\mathrm{s}}\,\dfrac{\delta_\mathrm{w}-\delta_\mathrm{s}}{1-\delta_\mathrm{w}}\right) + K_\mathrm{s}\dfrac{1-\delta_w}{1-\delta_\mathrm{s}}}, \tag{73a}$$

$$M_t = R_t\,\frac{D}{K_\mathrm{w}}\left(t - \left(1+\frac{K_\mathrm{w}-t+1}{tK_\mathrm{s}}\,\frac{\delta_\mathrm{w}-\delta_\mathrm{s}}{1-\delta_\mathrm{w}}\right)^{-1}\right). \tag{73b}$$

Theorem 3: The upper convex hull of the $K_\mathrm{w}+2$ rate-memory pairs in (71)–(73) forms a lower bound on the capacity-memory tradeoff:

$$C(M) \geq \text{upper hull}\big\{(R_t, M_t)\colon\ t = 0,\ldots,K_\mathrm{w}+1\big\}. \tag{74}$$

Proof outline: The pair $(R_0,\ M_0 = 0)$ corresponds to the case without caches, and the achievability follows from (17). The achievability of the pair $(R_{K_\mathrm{w}+1},\ M_{K_\mathrm{w}+1})$ follows from (20). The pairs $(R_1, M_1), \ldots, (R_{K_\mathrm{w}}, M_{K_\mathrm{w}})$ are achieved by the joint cache-channel coding scheme in Section IV. Finally, the upper convex hull of $\{(R_t, M_t); t = 0, 1, \ldots, K_\mathrm{w}+1\}$ is achieved by time- and memory-sharing. ■

The lower bound is piece-wise linear, where the slope of the lower bound decreases from one interval to the other. The caching gain achieved by our joint cache-channel coding scheme is thus largest in the regime of small cache memories $M \in [0, M_1]$, where M_1 is defined through (73b) and equals

$$M_1 = D\cdot\frac{(\delta_\mathrm{w}-\delta_\mathrm{s})K_\mathrm{s}^{-1}}{K_\mathrm{w} + \frac{K_\mathrm{w}-1}{2}\cdot\frac{K_\mathrm{w}(\delta_\mathrm{w}-\delta_\mathrm{s})}{K_\mathrm{s}(1-\delta_\mathrm{w})} + K_\mathrm{s}\frac{1-\delta_\mathrm{w}}{1-\delta_\mathrm{s}}}. \tag{75}$$

In this regime $M \leq M_1$, Theorem 3 specializes to:

$$C(M) \geq R_0 + \frac{M}{D}\cdot\frac{K_\mathrm{w}(1-\delta_\mathrm{s})}{K_\mathrm{w}(1-\delta_\mathrm{s}) + K_\mathrm{s}(1-\delta_\mathrm{w})}\cdot\frac{1+K_\mathrm{w}}{2}\cdot\gamma_\mathrm{joint}, \tag{76}$$

where

$$\gamma_\mathrm{joint} := 1 + \frac{2K_\mathrm{w}}{1+K_\mathrm{w}}\cdot\frac{K_\mathrm{s}(1-\delta_\mathrm{w})}{K_\mathrm{w}(1-\delta_\mathrm{s})}. \tag{77}$$

Note that if we replace γ_joint by 1 in the lower bound of (76), we recover the lower bound of Proposition 1. The factor γ_joint thus represents the gain of our joint cache-channel coding scheme compared to the simple separate cache-channel coding scheme of Section III. Note that γ_joint is unbounded in the number of strong receivers K_s (when K_w and the erasure probabilities δ_s and δ_w are constant). More generally, γ_joint is increasing in the ratio $\frac{K_\mathrm{s}(1-\delta_\mathrm{w})}{K_\mathrm{w}(1-\delta_\mathrm{s})}$ when $K_\mathrm{w} \geq 1$.

B. A general Upper Bound on $C(M)$

Define for each $k_\mathrm{w} \in \{0, \ldots, K_\mathrm{w}\}$:

$$R_{k_\mathrm{w}}(M) := \left(\frac{k_\mathrm{w}}{1-\delta_\mathrm{w}} + \frac{K_\mathrm{s}}{1-\delta_\mathrm{s}}\right)^{-1} + \frac{k_\mathrm{w}M}{D}.$$

Theorem 4: The capacity-memory tradeoff $C(M)$ is upper bounded as

$$C(M) \leq \min_{k_\mathrm{w}\in\{0,\ldots,K_\mathrm{w}\}} R_{k_\mathrm{w}}(M). \tag{78}$$

Proof: We specialize the upper bound in Theorem 2 to erasure BCs. Note that the strong receivers do not have cache memories and the symmetric capacity $R_{\mathrm{sym},\mathcal{S}}$ decreases as the receiver set $\mathcal{S}$ increases. So to compute (69), it suffices to only consider the bounds that correspond to subsets $\mathcal{S} \subseteq \mathcal{K}$ containing all the strong receivers (i.e., receivers $K_\mathrm{w}+1, \ldots, K$). ■

The choice of k_w in (78) that leads to the tightest upper bound depends on the cache size M. For small values of M, the tightest bound is attained for $k_\mathrm{w} = K_\mathrm{w}$ and for larger cache sizes, smaller values of k_w lead to tighter bounds.

The upper and lower bounds on $C(M)$ in Theorems 3 and 4 are illustrated in Figure 5.

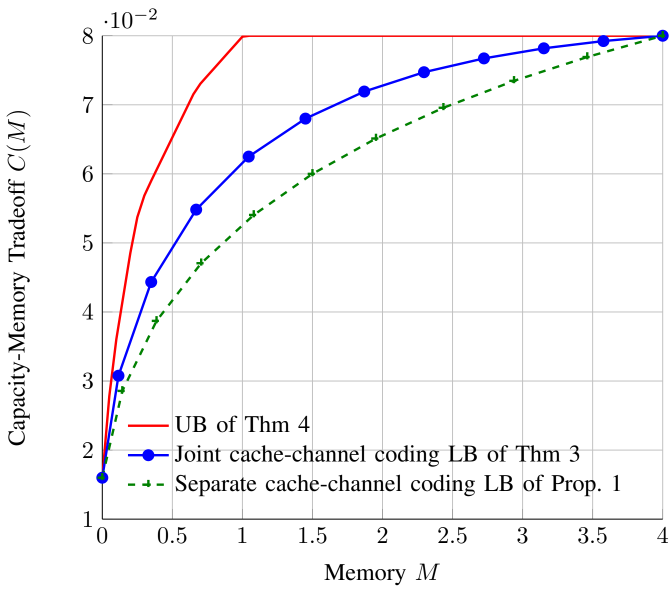

Fig. 5: Bounds on the capacity-memory tradeoff $C(M)$ for $K_\mathrm{w} = K_\mathrm{s} = 10$, $D = 50$, $\delta_\mathrm{w} = 0.8$, $\delta_\mathrm{s} = 0.2$.

C. Special Case of $K_\mathrm{w} = 1$

We evaluate our bounds for a setup with a single weak receiver and any number of strong receivers. Let

$$\Gamma_1 := \frac{(1-\delta_\mathrm{s})}{K_\mathrm{s}}\cdot\frac{(\delta_w-\delta_\mathrm{s})}{(K_\mathrm{s}(1-\delta_w) + (1-\delta_\mathrm{s}))}, \tag{79}$$

$$\Gamma_2 := \frac{(1-\delta_\mathrm{s})}{K_\mathrm{s}}\cdot\frac{(1-\delta_\mathrm{s})}{(K_\mathrm{s}(1-\delta_w) + (1-\delta_\mathrm{s}))}, \tag{80}$$

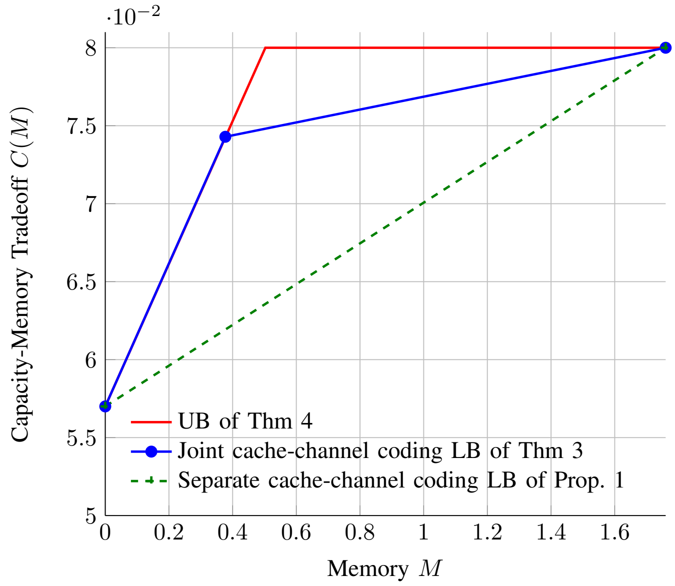

Fig. 6: Bounds on the capacity-memory tradeoff for $K_w = 1$, $K_s = 10$, $D = 22$, $\delta_w = 0.8$, $\delta_s = 0.2$.

$$\Gamma_3 := \frac{(1-\delta_s)}{K_s}. \tag{81}$$

Notice that $0 \le \Gamma_1 \le \Gamma_2 \le \Gamma_3$. From Theorems 3 and 4, we obtain the following corollary.

Corollary 4.1: If $K_w = 1$ the capacity-memory tradeoff is lower bounded by

$$C(M) \ge \begin{cases} \frac{(1-\delta_w)(1-\delta_s)}{K_s(1-\delta_w)+(1-\delta_s)} + \frac{M}{D}, & \text{if } \frac{M}{D} \in [0, \Gamma_1] \\ \frac{(1-\delta_s)}{1+K_s} + \frac{M}{(1+K_s)D}, & \text{if } \frac{M}{D} \in (\Gamma_1, \Gamma_3], \end{cases} \tag{82}$$

and upper bounded by

$$C(M) \le \begin{cases} \frac{(1-\delta_w)(1-\delta_s)}{K_s(1-\delta_w)+(1-\delta_s)} + \frac{M}{D}, & \text{if } \frac{M}{D} \in [0, \Gamma_2] \\ \frac{(1-\delta_s)}{K_s} & \text{if } \frac{M}{D} \in (\Gamma_2, \Gamma_3]. \end{cases} \tag{83}$$

Figure 6 shows these two bounds and the bound in Proposition 1 for $K_w = 1$, $K_s = 10$, $D = 22$, $D = 10$, $\delta_w = 0.8$, $\delta_s = 0.2$.

We identify two regimes of interest. First, in the regime $0 \le \frac{M}{D} \le \Gamma_1$, our lower and upper bounds match and show that the rate R scales with $\frac{M}{D}$ by the slope 1. This is achievable by our joint cache-channel coding scheme and corresponds to the performance when all the K_s strong receivers directly access receiver 1's cache contents. In the second regime, $\Gamma_1 < \frac{M}{D} \le \Gamma_3$, the joint cache-channel coding scheme still profits from an increasing cache size, but the gain is less significant: the rate only increases as $\frac{1-\delta_s}{K_s} \cdot \frac{M}{D}$.

We specialize this setup further to $K_s = 1$ and $D = 2$. So, assume for the rest of this section that

$$K_w = K_s = 1 \quad \text{and} \quad D = 2. \tag{84}$$

For this special case, we present tighter upper and lower bounds on $C(M)$. These new bounds meet for a larger range of cache sizes M. Let

$$\tilde{\Gamma}_1 := \frac{(1-\delta_s)^2 - (1-\delta_w)(\delta_w - \delta_s)}{(1-\delta_w) + (1-\delta_s)}, \tag{85}$$

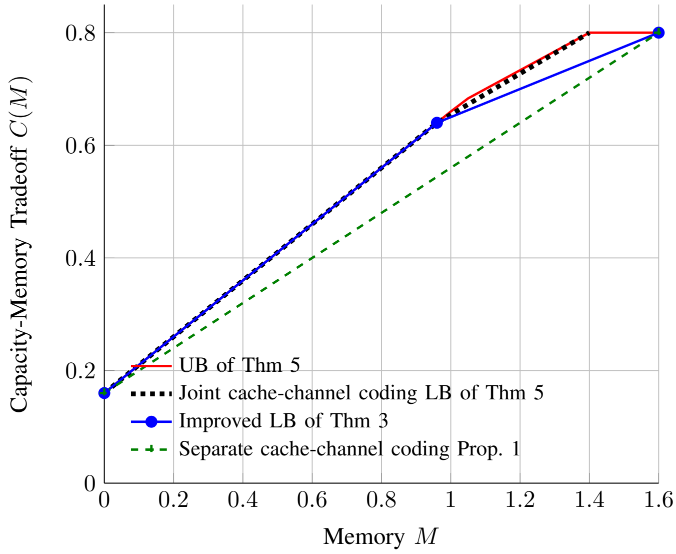

Fig. 7: Bounds on the capacity-memory tradeoff for $K_w = 1$, $K_s = 1$, $D = 2$, $\delta_w = 0.8$, $\delta_s = 0.2$.

$$\tilde{\Gamma}_2 := \frac{1}{2}\left((1-\delta_s) + (\delta_w - \delta_s)\right). \tag{86}$$

Notice that $0 \le \tilde{\Gamma}_1 \le \tilde{\Gamma}_2 < \Gamma_3$.

Theorem 5: If $K_w = K_s = 1$ and $D = 2$, the capacity-memory tradeoff is upper bounded as:

$$C(M) \le \begin{cases} \frac{(1-\delta_w)(1-\delta_s)}{(1-\delta_w)+(1-\delta_s)} + \frac{M}{2}, & \text{if } \frac{M}{2} \in [0, \tilde{\Gamma}_1] \\ \frac{1}{3}(2 - \delta_s - \delta_w) + \frac{M}{3}, & \text{if } \frac{M}{2} \in (\tilde{\Gamma}_1, \tilde{\Gamma}_2] \\ 1 - \delta_s & \text{if } \frac{M}{2} \in (\tilde{\Gamma}_2, \Gamma_3]. \end{cases} \tag{87}$$

and lower bounded as:

$$C(M) \ge \begin{cases} \frac{(1-\delta_w)(1-\delta_s)}{(1-\delta_w)+(1-\delta_s)} + \frac{M}{2}, & \frac{M}{2} \in [0, \Gamma_1] \\ \frac{(1-\delta_s)((1-\delta_s)+M)}{3(1-\delta_s)-(1-\delta_w)}, & \frac{M}{2} \in (\Gamma_1, \tilde{\Gamma}_2] \\ 1 - \delta_s & \frac{M}{2} \in (\tilde{\Gamma}_2, \Gamma_3]. \end{cases} \tag{88}$$

Proof: The lower bound in (88) coincides with the upper convex hull of the three rate-memory pairs: (R_0, M_0) in (71); (R_1, M_1) in (49); and $((1-\delta_s), 2\Gamma_3)$. Achievability of the former two pairs follows from Theorem 3. Achievability of the last pair follows from a joint cache-channel coding scheme that caches coded content and is described in Appendix F. The upper bound is proved in Appendix G. ■

Figure 7 shows the bounds of Theorem 5 for $\delta_w = 0.8$ and $\delta_s = 0.2$. The upper and lower bounds of Theorem 5 coincide for $0 \le M \le \Gamma_1$ and for $M \ge \tilde{\Gamma}_2$.

Using Theorem 5, we conclude that the minimum cache size M for which communication is possible at the maximum rate $(1-\delta_s)$ is $M = 2\tilde{\Gamma}_2$. Notice also that upper and lower bounds in Theorem 5 coincide for all values of M when $\delta_w = \delta_s$.

VII. Equivalent Results on Minimum Delivery Rate

The capacity-memory tradeoff considered thus far was formulated and presented using the typical nomenclature of multi-user information theory. This presentation, however, differs slightly to many previous works on caching (e.g., [2]). In this section, we will connect the two setups. Let us temporarily suppose that Messages $W_1, \ldots, W_D$ are F-bit packets and the weak receivers have mF-bit cache memories, for some positive integer F and some positive real number $m \in [0, D)$. Additionally, suppose that the delivery-phase communication takes place over ρF uses of the BC, where $\rho > 0$ is called the *delivery rate*.

A *delivery rate-memory pair* (ρ, m) is achievable in this new setup if there exist placement, encoding, and decoding functions such that the probability of decoding error vanishes as the packet size $F \to \infty$. Given a cache size m, the minimum delivery rate ρ for which (ρ, m) is achievable is called the *delivery rate-memory tradeoff* and is denoted by $\rho^\star(m)$.

There is a simple relation between the delivery rate-memory pairs (ρ, m) that are achievable in this new setup and the (message) rate-memory pairs (R, M) achievable in our original setup:

$$(R,\ M) \text{ achievable in original setup}$$
$$\Longleftrightarrow$$
$$\left(\rho = \frac{1}{R},\ m = \frac{M}{R}\right) \text{ achievable in new setup.}$$

Using this relation, we can now restate the rate-memory pairs achieved by the separate and the joint cache-channel coding schemes in terms of the delivery rate ρ and the cache size m. For each $\tilde{t} \in \mathcal{K}_\text{w}$, the separate cache-channel coding scheme in Section III achieves the delivery rate-memory pair

$$\rho_{\tilde{t},\text{sep}} := \frac{K_\text{w} - \tilde{t}}{(\tilde{t}+1)(1-\delta_\text{w})} + \frac{K_\text{s}}{(1-\delta_\text{s})}, \tag{89a}$$

$$m_{\tilde{t},\text{sep}} := D\frac{\tilde{t}}{K_\text{w}}. \tag{89b}$$

For each $t \in \mathcal{K}_\text{w}$, the joint cache-channel coding scheme in Section IV-B achieves the delivery rate-memory pair:

$$\rho_t := \nu_t \frac{K_\text{w} - t}{(t+1)(1-\delta_\text{w})} + (1-\nu_t)\frac{K_\text{w} - t + 1}{t(1-\delta_\text{w})} + (1-\nu_t)\frac{K_\text{s}}{1-\delta_\text{s}}, \tag{90a}$$

$$m_t := \nu_t D \frac{t}{K_\text{w}} + (1-\nu_t) D \frac{t-1}{K_\text{w}}, \tag{90b}$$

where

$$\nu_t := \frac{(K_\text{w} - t + 1)(\delta_\text{w} - \delta_\text{s})}{(K_\text{w} - t + 1)(\delta_\text{w} - \delta_\text{s}) + tK_\text{s}(1-\delta_\text{w})}. \tag{91}$$

The lower convex hull of the delivery rate-memory pairs in (89) and (90) upper bounds the delivery rate-memory tradeoff $\rho^\star(m)$.

The upper bound on $C(M)$ in Theorem 4 leads to the following lower bound on $\rho^\star(m)$:

$$\rho^\star(m) \geq \max_{k_\text{w} \in \{0,1,\ldots,K_\text{w}\}} \left[\left(\frac{k_\text{w}}{1-\delta_\text{w}} + \frac{K_\text{s}}{1-\delta_\text{s}}\right)\left(1 - \frac{k_\text{w} m}{D}\right)\right]. \tag{92}$$

Figures 8 and 9 present upper and lower bounds on $\rho^\star(m)$ when $K_\text{w} = K_\text{s} = 10$, $D = 50$, $\delta_\text{w} = 0.8$, $\delta_\text{s} = 0.2$ and when $K_\text{w} = 10$, $K_\text{s} = 1000$, $D = 5000$, $\delta_\text{w} = 0.8$, $\delta_\text{s} = 0.2$. The lower bound is given by (92) and the upper bounds depict the convex hulls of rate-memory pairs $\{(\rho_{\tilde{t},\text{sep}}, m_{\tilde{t},\text{sep}})\}_{\tilde{t}=0}^{K_\text{w}}$ and rate-memory pairs $\{(\rho_{0,\text{sep}}, m_{0,\text{sep}}),\ \{(\rho_t, m_t)\}_{t=1}^{K_\text{w}},\ (\rho_{K_\text{w},\text{sep}}, m_{K_\text{w},\text{sep}})\}$.

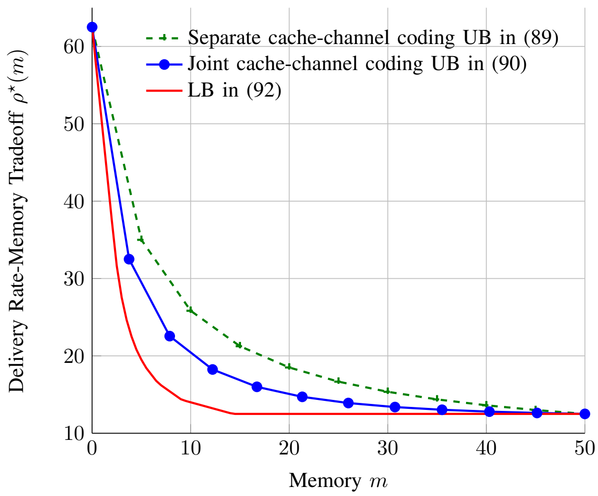

Fig. 8: Bounds on $\rho^\star(m)$ for $K_\text{w} = K_\text{s} = 10$, $D = 50$, $\delta_\text{w} = 0.8$, $\delta_\text{s} = 0.2$.

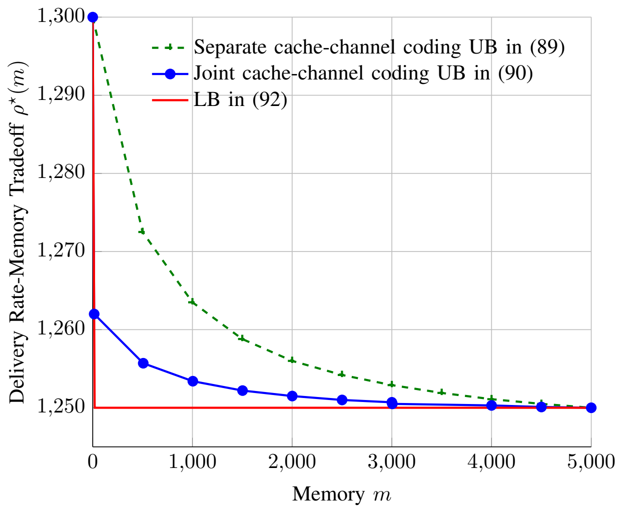

Fig. 9: Bounds on $\rho^\star(m)$ for $K_\text{w} = 10$ and $K_\text{s} = 1000$, $D = 5000$, $\delta_\text{w} = 0.8$, $\delta_\text{s} = 0.2$.

VIII. Summary and Concluding Remarks

In this paper, we considered an erasure broadcast network with a set of weak receivers with equal cache sizes M and a set of strong receivers with no cache memories. We derived upper and lower bounds on the capacity-memory tradeoff and discussed scenarios where the bounds match. In particular, the bounds match when there is a single weak receiver with a small cache size (and any number of strong receivers). A small cache size corresponds to a receiving device with limited storage space.

The derived upper bound holds more generally for any stochastically degraded BC. (An improved upper bound has more recently been proposed in [65].) The lower bound is obtained by means of joint cache-channel coding and significantly improves over a separate cache-channel coding scheme that combines coded caching with a capacity-achieving scheme for erasure BCs. In the regime of small cache memories, the improvement is even unbounded in the number of strong receivers. (An further improved upper bound that refines our coding scheme has recently been proposed in [64].)

To facilitate comparison with previous works that mostly focused on the delivery rate-memory tradeoff, we expressed our main results also in terms of the delivery rate-memory tradeoff. When specialized to the network with no strong receivers and with zero erasure probability at the weak receivers, the bounds presented in this paper coincide with the results by Maddah-Ali and Niesen [2].

For the setup with only one weak receiver and one strong receiver, we proposed improved upper and lower bounds that match over a wide regime of channel parameters and memory sizes. The lower bound is achieved by pre-placing coded content and using joint cache-channel coding for the delivery phase.

In the considered cache-aided BC model where weak and strong receivers are served at the same rate, performance is improved when larger cache memories are assigned to weak receivers as opposed to the traditional uniform cache assignment. In this work, we illustrated that applying joint cache-channel coding in such asymmetric cache configurations leads to further caching gains that cannot be attained with standard separate cache-channel coding schemes.

Acknowledgments

The authors thank the anonymous reviewers and the associate editor for their valuable comments that helped improve this manuscript.

Appendix A
Analysis of Joint Cache-Channel Coding Scheme in Section IV

Fix $t \in \mathcal{K}_{\text{w}}$, and

$$\beta_\imath := \frac{n_\imath}{n}, \qquad \imath \in \{1,2,3\}.$$

1) Placement Phase: By Proposition 1, the applied placement strategy requires a cache size of

$$\begin{aligned} M &= R^{(t)} \cdot D\frac{t}{K_{\text{w}}} + R^{(t-1)} \cdot D\frac{t-1}{K_{\text{w}}} \\ &= R \cdot \frac{D}{K_{\text{w}}}\left(t - \left(1 + \frac{K_{\text{w}} - t + 1}{tK_{\text{s}}} \cdot \frac{\delta_{\text{w}} - \delta_{\text{s}}}{1 - \delta_{\text{w}}}\right)^{-1}\right). \end{aligned} \quad (93)$$

2) Delivery Subphase 1: Notice that in Subphase 1 the separate cache-channel coding scheme of Section III is applied without strong receivers. Thus, by Proposition 1, the probability of decoding error can be made arbitrarily small by choosing n sufficiently large, because

$$\frac{R^{(t)} \cdot \frac{K_{\text{w}}-t}{t+1}}{1 - \delta_{\text{w}}} < \beta_1. \quad (94)$$

3) Delivery Subphase 2: Consider period ℓ with the transmission of messages in (58). The probability that the strong receivers make a decoding error can be made arbitrarily small by choosing n sufficiently large, because

$$\frac{R_{\text{sub}}^{(t-1)} + K_{\text{s}} R_{\text{sub}}^{(t)}}{1 - \delta_{\text{s}}} < \frac{\beta_2}{\binom{K_{\text{w}}}{t}}. \quad (95)$$

Weak receivers restrict their decoding to a row-codebook containing only $\lfloor 2^{nR_{\text{sub}}^{(t-1)}} \rfloor$ codewords. The probability that the weak receivers produce a wrong XOR message can be made arbitrarily small by choosing n sufficiently large, because

$$\frac{R_{\text{sub}}^{(t-1)}}{(1 - \delta_{\text{w}})} < \frac{\beta_2}{\binom{K_{\text{w}}}{t}}. \quad (96)$$

Notice that when the weak receivers decode their desired XOR messages correctly, then they also produce correct estimates for messages $W_{d_1}^{(t-1)}, \ldots, W_{d_{K_{\text{w}}}}^{(t-1)}$.

By our choice of the rates $R^{(t-1)}$ and $R^{(t)}$ in (51), the two constraints (95) and (96) coincide. We ignore (95) in the following.

4) Delivery Subphase 3: The probability that the strong receivers err in their decoding can be made arbitrarily small by choosing n sufficiently large, because

$$\frac{K_{\text{s}} R^{(t-1)}}{(1 - \delta_{\text{s}})} < \beta_3. \quad (97)$$

5) Overall Scheme: Combining (94), (96), and (97) and using the definitions of $R_{\text{sub}}^{(t-1)}$ and $R_{\text{sub}}^{(t-1)}$ in (52), we conclude that the probability of decoding can be made arbitrarily small by choosing n sufficiently large, because

$$\frac{R^{(t)} \frac{K_{\text{w}}-t}{t+1}}{1 - \delta_{\text{w}}} + \frac{R^{(t-1)} \cdot \frac{K_{\text{w}}-t+1}{t}}{1 - \delta_{\text{w}}} + \frac{K_{\text{s}} R^{(t-1)}}{1 - \delta_{\text{s}}} < 1. \quad (98)$$

Using the definitions of $R^{(t-1)}$ and $R^{(t)}$ in (51), one obtains that the probability of decoding error can be made arbitrarily small, if the total rate R satisfies

$$R < (1 - \delta_{\text{w}}) \cdot \frac{1 + \frac{K_{\text{w}}-t+1}{tK_{\text{s}}} \cdot \frac{\delta_{\text{w}}-\delta_{\text{s}}}{1-\delta_{\text{w}}}}{\frac{K_{\text{w}}-t+1}{t}\left(1 + \frac{K_{\text{w}}-t}{(t+1)K_{\text{s}}} \cdot \frac{\delta_{\text{w}}-\delta_{\text{s}}}{1-\delta_{\text{w}}}\right) + K_{\text{s}}\frac{1-\delta_{\text{w}}}{1-\delta_{\text{s}}}}.$$

Together with (93), this proves achievability of the rate-memory pair $(R_t,\ M_t)$ in (49).

Appendix B
Proof of Theorem 2

For ease of exposition, we only prove the bound corresponding to $\mathcal{S} = \mathcal{K}$:

$$C(M_1, \ldots, M_K) \le R_{\text{sym},\mathcal{K}} + \frac{1}{D}\sum_{k=1}^{K} M_k. \quad (99)$$

The bounds corresponding to other subsets $\mathcal{S}$ can be proved in a similar way. It suffices to ignore a subset of the receivers and their cache memories.

Fix the rate of communication

$$R < C(M_1, \ldots, M_K).$$

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

$$V_k^{(n)} = g_k^{(n)}(W_1, \ldots, W_D), \qquad k \in \mathcal{K},$$

denote the cache content for the chosen placement function.

Lemma 6: For any $\epsilon > 0$, any demand vector $\mathbf{d} = (d_1, \ldots, d_K)$ with all different entries,[8] and any blocklength n that is sufficiently large (depending on ϵ), there exist random variables $(U_{1,\mathbf{d}}, \ldots, U_{K-1,\mathbf{d}}, X_{\mathbf{d}}, Y_{1,\mathbf{d}}, \ldots, Y_{K,\mathbf{d}})$ such that

$$U_{1,\mathbf{d}} - U_{2,\mathbf{d}} - \cdots - U_{K-1,\mathbf{d}} - X_{\mathbf{d}} - Y_{K,\mathbf{d}} - Y_{K-1,\mathbf{d}} \cdots - Y_{1,\mathbf{d}} \tag{100a}$$

forms a Markov chain, such that given $X_{\mathbf{d}} = x \in \mathcal{X}$:

$$(Y_{1,\mathbf{d}}, Y_{2,\mathbf{d}}, \ldots, Y_{K,\mathbf{d}}) \sim P_{Y_1 \cdots Y_K|X}(\cdots|x), \tag{100b}$$

and such that the following K inequalities hold:

$$\begin{aligned} R - \epsilon &\leq \frac{1}{n} I\big(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}\big) + I\big(U_{1,\mathbf{d}}; Y_{1,\mathbf{d}}\big), \quad \text{(101a)} \\ R - \epsilon &\leq \frac{1}{n} I\big(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{k-1}}\big) \\ &\qquad + I\big(U_{k,\mathbf{d}}; Y_{k,\mathbf{d}} | U_{k-1,\mathbf{d}}\big), \quad \forall k \in \{2, \ldots, K-1\}, \\ R - \epsilon &\leq \frac{1}{n} I\big(W_{d_K}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{K-1}}\big) \\ &\qquad + I\big(X_{\mathbf{d}}; Y_{K,\mathbf{d}} | U_{K-1,\mathbf{d}}\big). \quad \text{(101b)} \end{aligned}$$

Proof: The proof is similar to the converse proof of the capacity of degraded BCs without caching [61, Theorem 5.2]. It is deferred to Appendix C. ∎

Fix $\epsilon > 0$ and a blocklength n (depending on this ϵ) so that Lemma 6 holds for all demand vectors $\mathbf{d}$ that have all different entries. We average the bound obtained in (101) over different demand vectors. Let $\mathcal{Q}$ be the set of all $\binom{D}{K} K!$ demand vectors whose K entries are all different. Also, let Q be a uniform random variable over the elements of $\mathcal{Q}$ that is independent of all previously defined random variables. Define: $U_1 := (U_{1,Q}, Q)$; $U_k := U_{k,Q}$, for $k \in \{2, \ldots, K-1\}$; $X := X_Q$; and $Y_k := Y_{k,Q}$ for $k \in \mathcal{K}$. Notice that they form the Markov chain

$$U_1 \to U_2 \to \cdots \to U_{K-1} \to X \to (Y_1, \ldots, Y_K), \tag{102}$$

and given $X = x \in \mathcal{X}$ satisfy

$$(Y_1, Y_2, \ldots, Y_K) \sim P_{Y_1 \cdots Y_K|X}(\cdots|x). \tag{103}$$

[8] Here we use the assumption that there are more messages than users, $D \geq K$, and therefore such a demand vector exists. A similar lemma can be obtained in the case where $D < K$.

Averaging each inequality in (101) over the demand vectors in $\mathcal{Q}$, and by using standard arguments to take care of the time-sharing random variable Q, we obtain:

$$R - \epsilon \leq \alpha_1 + I\big(U_1; Y_1\big), \tag{104a}$$
$$R - \epsilon \leq \alpha_k + I\big(U_k; Y_k | U_{k-1}\big), \quad \forall k \in \{2, \ldots, K-1\}, \tag{104b}$$
$$R - \epsilon \leq \alpha_k + I\big(X; Y_K | U_{K-1}\big), \tag{104c}$$

where $\alpha_1, \ldots, \alpha_K$ are defined as

$$\alpha_1 := \frac{1}{\binom{D}{K} K!} \sum_{\mathbf{d} \in \mathcal{Q}} \frac{1}{n} I(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}), \tag{105a}$$

$$\alpha_k := \frac{1}{\binom{D}{K} K!} \sum_{\mathbf{d} \in \mathcal{Q}} \frac{1}{n} I(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{k-1}}), \quad k \in \{2, \ldots, K\}. \tag{105b}$$

Lemma 7: Parameters $\alpha_1, \ldots, \alpha_K$, defined in (105), satisfy the following constraints:

$$\alpha_k \geq 0, \qquad k \in \mathcal{K}, \tag{106a}$$
$$\alpha_{k'} \leq \alpha_k, \qquad k, k' \in \mathcal{K},\ k' \leq k, \tag{106b}$$
$$\sum_{k \in \mathcal{K}} \alpha_k \leq \frac{K}{D} \sum_{k \in \mathcal{K}} M_k. \tag{106c}$$

Proof: See Appendix D. ∎

We now take $\epsilon \to 0$ and use Lemma 7 to conclude that the capacity-memory tradeoff $C(M_1, \ldots, M_K)$ is upper bounded by the following K inequalities:

$$C(M_1, \ldots, M_K) \leq \alpha_1 + I\big(U_1; Y_1\big), \tag{107a}$$
$$C(M_1, \ldots, M_K) \leq \alpha_k + I\big(U_k; Y_k \big| U_{k-1}\big), \quad \forall k \in \{2, \ldots, K-1\}, \tag{107b}$$
$$C(M_1, \ldots, M_K) \leq \alpha_K + I\big(X; Y_K \big| U_{K-1}\big), \tag{107c}$$

for some $\alpha_1, \ldots, \alpha_K$ that satisfy (106) and some $U_1, \ldots, U_{K-1}, X, Y_1, \ldots, Y_K$ that satisfy (102) and (103).

Lemma 8: Replacing each and every real number $\alpha_1, \ldots, \alpha_K$ in (107) by $\frac{1}{D} \sum_{k \in \mathcal{K}} M_k$ does not change the upper bound on $C(M_1, \ldots, M_K)$.

Proof: See Appendix E. ∎

Thus,

$$C(M_1, \ldots, M_K) - \frac{\sum_{k \in \mathcal{K}} M_k}{D} \leq I\big(U_1; Y_1\big), \tag{108a}$$

for all $k \in \{2, \ldots, K-1\}$:

$$C(M_1, \ldots, M_K) - \frac{\sum_{k \in \mathcal{K}} M_k}{D} \leq I\big(U_k; Y_k \big| U_{k-1}\big), \tag{108b}$$

and

$$C(M_1, \ldots, M_K) - \frac{\sum_{k \in \mathcal{K}} M_k}{D} \leq I\big(X; Y_K \big| U_{K-1}\big), \tag{108c}$$

for some $U_1, \ldots, U_K, X, Y_1, \ldots, Y_K$ satisfying (102) and (103).

All K constraints in (108) have the same LHS, and their RHSs coincide with the rate-constraints that determine the capacity region of a degraded BC without caches. Therefore, the choice of the random variables $(U_1,\ldots,U_{K-1},X)$ that leads to the most relaxed constraint on $C(M_1,\ldots,M_K)$ coincides with the choice of auxiliaries that determines the largest symmetric rate-point in the capacity region of a degraded BC without caches. This establishes the equivalence of (108) with the desired bound in (99), and thus concludes the proof.

Appendix C
Proof of Lemma 6

Fix a small $\epsilon>0$ and a demand vector $\mathbf{d}$ with all different entries. Then, let the blocklength n be sufficiently large as will become clear in the following. Also, let

$$V_k^{(n)}=g_k^{(n)}(W_1,\ldots,W_D),\qquad k\in\mathcal{K},\tag{109}$$
$$X_{\mathbf{d}}^n=f_{\mathbf{d}}^{(n)}(W_1,\ldots,W_D)\tag{110}$$

denote cache contents and the input of the degraded BC for demand vector $\mathbf{d}\in\mathcal{D}^K$ and for the above chosen placement and encoding functions. We denote by $Y_{k,\mathbf{d}}^n$ the corresponding channel outputs at receiver $k\in\mathcal{K}$.

By Fano's inequality, the independence of the messages $W_1,\ldots,W_D$, and because the placement, encoding, and decoding functions have been chosen so that the worst case probability of error tends to 0 as $n\to\infty$, we obtain that for any $\epsilon>0$ there exists a sufficiently large $n'\geq 0$ such that the following K inequalities hold for all $n\geq n'$:

$$\begin{aligned}R-\epsilon &\leq \frac{1}{n}I\big(W_{d_1};Y_{1,\mathbf{d}}^n,V_1^{(n)},\ldots,V_K^{(n)}\big)\\ &=\frac{1}{n}I\big(W_{d_1};V_1^{(n)},\ldots,V_K^{(n)}\big)\\ &\quad+\frac{1}{n}I\big(W_{d_1};Y_{1,\mathbf{d}}^n\big|V_1^{(n)},\ldots,V_K^{(n)}\big),\end{aligned}\tag{111a}$$

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

$$\begin{aligned}R-\epsilon_n &\leq\frac{1}{n}I\big(W_{d_k};Y_{k,\mathbf{d}}^n,V_1^{(n)},\ldots,V_K^{(n)}\big|W_{d_1},\ldots,W_{d_{k-1}}\big)\\ &=\frac{1}{n}I\big(W_{d_k};V_1^{(n)},\ldots,V_K^{(n)}\big|W_{d_1},\ldots,W_{d_{k-1}}\big)\\ &\quad+\frac{1}{n}I\big(W_{d_k};Y_{k,\mathbf{d}}^n\big|V_1^{(n)},\ldots,V_K^{(n)},W_{d_1},\ldots,W_{d_{k-1}}\big).\end{aligned}\tag{111b}$$

We further develop the second summands in (111a) and (111b). For the second summand in (111a) we obtain

$$\begin{aligned}&\frac{1}{n}I\big(W_{d_1};Y_{1,\mathbf{d}}^n\big|V_1^{(n)},\ldots,V_K^{(n)}\big)\\ &=\frac{1}{n}\sum_{t=1}^n I\big(W_{d_1};Y_{1,\mathbf{d},t}\big|V_1^{(n)},\ldots,V_K^{(n)},Y_{1,\mathbf{d}}^{t-1}\big)\\ &\leq\frac{1}{n}\sum_{t=1}^n I\big(W_{d_1},V_1^{(n)},\ldots,V_K^{(n)},Y_{1,\mathbf{d}}^{t-1};Y_{1,\mathbf{d},t}\big)\\ &=I\big(U_{1,\mathbf{d},T};Y_{1,\mathbf{d},T}\big|T\big)\\ &\leq I(U_{1,\mathbf{d}};Y_{1,\mathbf{d}}),\end{aligned}\tag{112}$$

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

$$\begin{aligned}U_{1,\mathbf{d},T}&:=\big(W_{d_1},V_1^{(n)}\ldots,V_K^{(n)},Y_{1,\mathbf{d}}^{T-1}\big),\\ U_{1,\mathbf{d}}&:=(U_{1,\mathbf{d},T},T),\\ Y_{1,\mathbf{d}}&:=Y_{1,\mathbf{d},T}.\end{aligned}$$

We also define for $k\in\{2,\ldots,K-1\}$:

$$\begin{aligned}U_{k,\mathbf{d},T}&:=(V_1^{(n)}\ldots,V_K^{(n)},W_{d_1},W_{d_2},\ldots,W_{d_k},\\ &\qquad\qquad Y_{1,\mathbf{d}}^{T-1},\ldots,Y_{k,\mathbf{d}}^{T-1}),\\ U_{k,\mathbf{d}}&:=(U_{k,\mathbf{d},T},T),\\ Y_{k,\mathbf{d}}&:=Y_{k,\mathbf{d},T},\end{aligned}$$

in order to expand the second summand in (111b) as:

$$\begin{aligned}&\frac{1}{n}I\big(W_{d_k};Y_{k,\mathbf{d}}^n\big|V_1^{(n)},\ldots,V_K^{(n)},W_{d_1},\ldots,W_{d_{k-1}}\big)\\ &=\frac{1}{n}\sum_{t=1}^n I\big(W_{d_k};Y_{k,\mathbf{d},t}\big|V_1^{(n)},\ldots,V_K^{(n)},\\ &\qquad\qquad W_{d_1},\ldots,W_{d_{k-1}},Y_{k,\mathbf{d}}^{t-1}\big)\\ &=\frac{1}{n}\sum_{t=1}^n I\big(W_{d_k};Y_{k,\mathbf{d},t}\big|V_1^{(n)},\ldots,V_K^{(n)},\\ &\qquad W_{d_1},\ldots,W_{d_{k-1}},Y_{1,\mathbf{d}}^{t-1},\ldots,Y_{k-1,\mathbf{d}}^{t-1},Y_{k,\mathbf{d}}^{t-1}\big)\\ &\leq\frac{1}{n}\sum_{t=1}^n I\big(W_{d_k},Y_{k,\mathbf{d}}^{t-1}\ ;\ Y_{k,\mathbf{d},t}\big|V_1^{(n)},\ldots,V_K^{(n)},\\ &\qquad\qquad W_{d_1},\ldots,W_{d_{k-1}},Y_{1,\mathbf{d}}^{t-1},\ldots,Y_{k-1,\mathbf{d}}^{t-1}\big)\\ &=I\big(U_{k,\mathbf{d},T};Y_{k,\mathbf{d},T}\big|U_{k-1,\mathbf{d},T},T\big)\\ &=I\big(U_{k,\mathbf{d}};Y_{k,\mathbf{d}}\big|U_{k-1,\mathbf{d}}\big)\end{aligned}\tag{113}$$

where the second equality follows from the degradedness of the outputs, see (65).

Similarly, for $k=K$:

$$\begin{aligned}&\frac{1}{n}I\big(W_{d_K};Y_{K,\mathbf{d}}^n\big|V_1^{(n)},\ldots,V_K^{(n)},W_{d_1},\ldots,W_{d_{K-1}}\big)\\ &\quad\leq I(X_{\mathbf{d}};Y_{k,\mathbf{d}}|U_{K-1,\mathbf{d}}),\end{aligned}\tag{114}$$

where

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

Since the defined random variables satisfy (100), Inequalities (111)–(114) conclude the proof.

Appendix D
Proof of Lemma 7

Constraint (106a) follows by the nonnegativity of mutual information. To prove Constraint (106b), we fix a demand vector $\mathbf{d}\in\mathcal{Q}$, and consider the cyclic shifts of this vector. For $\ell\in\{0,\ldots,K-1\}$, let $\overrightarrow{\mathbf{d}}^{(\ell)}$ be the vector obtained from $\overrightarrow{\mathbf{d}}$ when the elements are cyclically shifted ℓ positions to the right. E.g., if $\mathbf{d}=(1,2,3)$ then $\overrightarrow{\mathbf{d}}^{(2)}=(2,3,1)$. For each $\ell\in\{0,\ldots,K-1\}$ and $k\in\mathcal{K}$, let $\overrightarrow{d}_k^{(\ell)}$ denote the k-th index of demand vector $\overrightarrow{\mathbf{d}}^{(\ell)}$. Thus,

$$\overrightarrow{d}_k^{(\ell)}=d_{(k-\ell)\bmod K}\tag{115}$$

where for each positive integer ξ the term $(\xi\mod K)$ takes value in $\mathcal{K}$ so that

$$\xi\mod K=\xi-bK\quad\text{for some positive integer } b.\tag{116}$$

For each $\ell \in \{1, \ldots, K-1\}$ and $k, k' \in \{2, \ldots, K\}$ with $k' < k$:

$$\begin{aligned}
&I(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}) \stackrel{(a)}{=} I(W_{\vec{d}_{k'}^{(k'-1)}}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\stackrel{(b)}{\leq} I(W_{\vec{d}_{k'}^{(k'-1)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\vec{d}_1^{(k'-1)}}, \ldots, W_{\vec{d}_{k'-1}^{(k'-1)}}) \\
&\stackrel{(a)}{=} I(W_{\vec{d}_k^{(k-1)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\vec{d}_{k-k'+1}^{(k-1)}}, \ldots, W_{\vec{d}_{k-1}^{(k-1)}}) \\
&\stackrel{(b)}{\leq} I(W_{\vec{d}_k^{(k-1)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\vec{d}_1^{(k-1)}}, \ldots, W_{\vec{d}_{k-1}^{(k-1)}}),
\end{aligned} \tag{117}$$

where (a) follows by (115), and (b) follows by the independence of the messages, the fact that conditioning does not increase entropy, and because $k - k' + 1 \geq 2$.

Fix a demand vector $\mathbf{d} \in \mathcal{Q}$ and sum the above inequality (117) over all K cyclic shifts $\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(K-1)}$ of $\mathbf{d}$ to obtain:

$$\begin{aligned}
&\sum_{\ell=0}^{K-1} I(W_{\vec{d}_1^{(\ell)}}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\quad \leq \sum_{\ell=0}^{K-1} I(W_{\vec{d}_{k'}^{(\ell)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\vec{d}_1^{(\ell)}}, \ldots, W_{\vec{d}_{k'-1}^{(\ell)}}) \\
&\quad \leq \sum_{\ell=0}^{K-1} I(W_{\vec{d}_k^{(\ell)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\vec{d}_1^{(\ell)}}, \ldots, W_{\vec{d}_{k-1}^{(\ell)}}).
\end{aligned} \tag{118}$$

Since the set $\mathcal{Q}$ can be partitioned into subsets of demand vectors that are cyclic shifts of each other and all cyclic shifts of a demand vector in $\mathcal{Q}$ are also in $\mathcal{Q}$, we conclude from (118):

$$\begin{aligned}
&\sum_{\mathbf{d} \in \mathcal{Q}} I(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\quad \leq \sum_{\mathbf{d} \in \mathcal{Q}} I(W_{d_{k'}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{k'-1}}) \\
&\quad \leq \sum_{\mathbf{d} \in \mathcal{Q}} I(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{k-1}}).
\end{aligned} \tag{119}$$

This proves (106b) by showing that $\alpha_1 \leq \alpha_{k'} \leq \alpha_k$.

We proceed to prove Constraint (106c). For each $\mathbf{d} \in \mathcal{Q}$:

$$\begin{aligned}
&I(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\quad + \sum_{k=2}^{K} I(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, W_{d_2}, \ldots, W_{d_{k-1}}) \\
&\quad\quad = I(W_{d_1}, W_{d_2}, \ldots, W_{d_K}; V_1^{(n)}, \ldots, V_K^{(n)}).
\end{aligned} \tag{120}$$

Thus,

$$\begin{aligned}
&\sum_{\mathbf{d} \in \mathcal{Q}} \Bigg[I(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\quad + \sum_{k=2}^{K} I(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, W_{d_2}, \ldots, W_{d_{k-1}}) \Bigg] \\
&= \sum_{\mathbf{d} \in \mathcal{Q}} I(W_{d_1}, W_{d_2}, \ldots, W_{d_K}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\stackrel{(a)}{=} \sum_{\mathbf{d} \in \mathcal{Q}} \Big[H(W_{d_1}) + H(W_{d_2}) + \ldots + H(W_{d_K}) \\
&\qquad - H(W_{d_1}, \ldots, W_{d_K} | V_1^{(n)}, \ldots, V_K^{(n)}) \Big] \\
&\stackrel{(b)}{=} \frac{K}{D} |\mathcal{Q}| H(W_1, \ldots, W_D) \\
&\quad - \sum_{\mathbf{d} \in \mathcal{Q}} H(W_{d_1}, \ldots, W_{d_K} | V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\stackrel{(c)}{\leq} \frac{K}{D} K! \binom{D}{K} H(W_1, \ldots, W_D) \\
&\quad - \frac{K}{D} K! \binom{D}{K} H(W_1, \ldots, W_D | V_1^{(n)}, \ldots, V_K^{(n)}) \\
&= \frac{K}{D} K! \binom{D}{K} I(W_1, \ldots, W_D; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\leq \frac{K}{D} K! \binom{D}{K} n \sum_{k=1}^{K} M_k,
\end{aligned}$$

where (a) holds by the chain rule of mutual information, (b) by the independence and uniform rate of messages $W_1, \ldots, W_D$ and the definition of the set $\mathcal{Q}$, which is of size $\binom{D}{K} K!$, and (c) by the generalized Han-Inequality (the following Proposition 9).

Proposition 9: Let L be a positive integer and $A_1, \ldots, A_L$ be a finite random L-tuple. Denote by $A_{\mathcal{S}}$ the subset $\{A_\ell, \ \ell \in \mathcal{S}\}$. For every $\ell \in \{1, \ldots, L\}$:

$$\frac{1}{\binom{L}{\ell}} \sum_{\mathcal{S} \subseteq \{1, \ldots, L\} : |\mathcal{S}| = \ell} \frac{H(A_{\mathcal{S}})}{\ell} \geq \frac{1}{L} H(A_1, \ldots, A_L). \tag{121}$$

Proof: See [62, Theorem 17.6.1]. ■

Appendix E
Proof of Lemma 8

Fix random variables $U_1, U_2, \ldots, U_{K-1}, X$ satisfying the Markov chain (102) and real numbers $\alpha_1, \ldots, \alpha_K$ satisfying (106). We will show that if $\alpha_{\tilde{k}} \neq \alpha_{\tilde{k}+1}$ for some $\tilde{k} \in \{1, \ldots, K-1\}$, then we can find new random variables $\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_{K-1}$ satisfying the Markov chain

$$\bar{U}_1 \to \bar{U}_2 \to \ldots \to \bar{U}_{K-1} \to X \to (Y_1, \ldots, Y_K), \tag{122}$$

and real numbers $\bar{\alpha}_1, \ldots, \bar{\alpha}_K$ satisfying (106) so that the upper bound on $C(M_1, \ldots, M_K)$ in (107) is relaxed if we replace

$$(U_1, U_2, \ldots, U_{K-1}) \quad \text{and} \quad (\alpha_1, \ldots, \alpha_K)$$

by

$$(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_{K-1}) \quad \text{and} \quad (\bar{\alpha}_1, \ldots, \bar{\alpha}_K).$$

This proves that the upper bound on $C(M_1, \ldots, M_K)$ in (107) remains unchanged if we replace all numbers $\alpha_1, \ldots, \alpha_K$ by the same number α. By (106c) this number $\alpha \leq \frac{1}{D} \sum_{k \in \mathcal{K}} M_k$, and by the monotonicity of the RHSs of (107) in $\alpha_1, \ldots, \alpha_K$ the choice $\alpha = \frac{1}{D} \sum_{k \in \mathcal{K}} M_k$ leads to the most relaxed upper bound. This will conclude the proof.

Assume that $\alpha_{\tilde{k}} \neq \alpha_{\tilde{k}+1}$ for some $\tilde{k} \in \{1, \ldots, K-1\}$. By (106b), the strict inequality

$$\alpha_{\tilde{k}} < \alpha_{\tilde{k}+1} \tag{123}$$

must hold. Choose

$$\bar{\alpha}_k = \alpha_k, \qquad k \in \mathcal{K},\ k \notin \{\tilde{k}, \tilde{k}+1\}, \tag{124}$$

$$\bar{\alpha}_{\tilde{k}} = \bar{\alpha}_{\tilde{k}+1} = \frac{1}{2}(\alpha_{\tilde{k}} + \alpha_{\tilde{k}+1}), \tag{125}$$

$$\bar{U}_k = U_k, \qquad k \in \{1, \ldots, K-1\},\ k \neq \tilde{k}. \tag{126}$$

For convenience, define

$$\bar{U}_K := U_K := X. \tag{127}$$

The choice of $\bar{U}_{\tilde{k}}$ depends on whether.

$$I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}) \leq I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}), \tag{128a}$$

or

$$I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}) > I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}), \tag{128b}$$

where for $\tilde{k} = 1$ the random variable $U_{\tilde{k}-1}$ is defined as a constant.

If (128a) holds, choose

$$\bar{U}_{\tilde{k}} = U_{\tilde{k}}. \tag{129}$$

If (128b) holds, let $E \in \{0,1\}$ be a Bernoulli-β random variable independent of everything else, where

$$\beta := \frac{1}{2} + \frac{1}{2} \frac{I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}})}{I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1})}, \tag{130}$$

and choose

$$\bar{U}_{\tilde{k}} = \begin{cases} (U_{\tilde{k}}, E), & \text{if } E = 1 \\ (U_{\tilde{k}-1}, E), & \text{if } E = 0. \end{cases} \tag{131}$$

The proposed choice satisfies the Markov chain (122).

Trivially, for $k \notin \{\tilde{k}, \tilde{k}+1\}$, Constraint (107) is unchanged if we replace $(U_1, U_2, \ldots, U_{K-1})$ by $(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_{K-1})$ and $(\alpha_1, \ldots, \alpha_K)$ by $(\bar{\alpha}_1, \ldots, \bar{\alpha}_K)$.

If (128a) holds, then the proposed replacement relaxes Constraint (107) for $k = \tilde{k}$ and it tightens it for $k = \tilde{k}+1$. However, the new constraint for $k = \tilde{k}+1$ is less stringent than the original constraint for $k = \tilde{k}$. We conclude that when (128a) holds, the upper bound on $C(M_1, \ldots, M_K)$ in (107) remains unchanged if everywhere one replaces $(U_1, U_2, \ldots, U_{K-1})$ and $(\alpha_1, \ldots, \alpha_K)$ by $(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_{K-1})$ and $(\bar{\alpha}_1, \ldots, \bar{\alpha}_K)$.

Assume now that (128b) holds. Then, by (130) and (131), because $\bar{U}_{\tilde{k}-1} = U_{\tilde{k}-1}$, and because E is independent of the pair $(Y_{\tilde{k}}, \bar{U}_{\tilde{k}-1})$:

$$\begin{aligned} & I(\bar{U}_{\tilde{k}}; Y_{\tilde{k}} | \bar{U}_{\tilde{k}-1}) \\ & = I(\bar{U}_{\tilde{k}}; Y_{\tilde{k}} | \bar{U}_{\tilde{k}-1}, E) \\ & = \beta \cdot I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}, E = 1) \\ & = \frac{1}{2} \left(I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}) + I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}) \right). \end{aligned} \tag{132}$$

By (125) and (132), the new constraint obtained for $k = \tilde{k}$ coincides with the average of the two original constraints for $k = \tilde{k}$ and for $k = \tilde{k}+1$. This average constraint cannot be more stringent than the most stringent of the two original constraints. The new constraint obtained for $k = \tilde{k}+1$ is more relaxed than the new constraint obtained for $k = \tilde{k}$, because of (125) and because

$$\begin{aligned} & I(\bar{U}_{\tilde{k}+1}; Y_{\tilde{k}+1} | \bar{U}_{\tilde{k}}) \\ & \overset{(a)}{=} \beta I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}) + (1-\beta) I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}-1}) \\ & \overset{(b)}{=} \beta I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}) + (1-\beta) I(U_{\tilde{k}+1}, U_{\tilde{k}}; Y_{\tilde{k}+1} | U_{\tilde{k}-1}) \\ & \overset{(c)}{=} I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}) + (1-\beta) I(U_{\tilde{k}}; Y_{\tilde{k}+1} | U_{\tilde{k}-1}) \\ & \overset{(d)}{\geq} I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}) + (1-\beta) I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}) \\ & \overset{(e)}{=} \frac{1}{2} I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}) + \frac{1}{2} I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}) \\ & \overset{(f)}{=} I(\bar{U}_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}), \end{aligned} \tag{133}$$

where (a) follows by the definition of $\bar{U}_{\tilde{k}}$ and $\bar{U}_{\tilde{k}+1}$; (b) by the Markov chain (102); (c) by the chain rule of mutual information and the Markov chain (102); (d) by the degradedness of the channel (65); (e) by the definition of β in (130); and (f) by (132).

We can thus conclude that also when (128b) holds, the upper bound on $C(M_1, \ldots, M_K)$ in (107) remains unchanged if one replaces $(U_1, U_2, \ldots, U_{K-1})$ and $(\alpha_1, \ldots, \alpha_K)$ by $(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_{K-1})$ and $(\bar{\alpha}_1, \ldots, \bar{\alpha}_K)$.

Appendix F
Achievability Proof for Rate-Memory Pair $(1 - \delta_\text{s},\ 2\tilde{\Gamma}_2)$

Assume $K_\text{w} = K_\text{s} = 1$. The following scheme achieves the rate-memory pair

$$R = (1 - \delta_\text{s}) \quad \text{and} \quad M = 2\tilde{\Gamma}_2. \tag{134}$$

Split messages W_1 and W_2 into two independent submessages

$$W_d = (W_d^{(1)}, W_d^{(2)}), \quad d \in \{1, 2\},$$

of rates

$$R^{(1)} := \delta_\text{w} - \delta_\text{s}, \tag{135a}$$

$$R^{(2)} := 1 - \delta_\text{w} - \epsilon, \tag{135b}$$

for an arbitrarily small $\epsilon > 0$.

Placement Phase: Cache the triple

$$V_1 := \left(W_1^{(1)}, W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)} \right) \tag{136}$$

in the weak receiver's cache.

Delivery Phase: The strong receiver, receiver 2, has to learn $W_{d_2}^{(1)}$ and $W_{d_2}^{(2)}$. The weak receiver, receiver 1, only needs to learn $W_{d_1}^{(2)}$, because it has already stored $W_{d_1}^{(1)}$ in its cache memory. Since receiver 1 has also stored $W_1^{(2)} \oplus W_2^{(2)}$ in its cache memory, in our scheme, we convey $W_{d_2}^{(2)}$ to it. From this message part and the content in its cache memory, receiver 1 can then find $W_{d_1}^{(2)}$. We use the piggyback coding idea from Section IV to send $W_{d_2}^{(1)}$—which is cached at the weak receiver—to the strong receiver and to send $W_{d_2}^{(2)}$ to both receivers. For this purpose, construct a random codebook with $\lfloor 2^{nR^{(1)}} \rfloor \times \lfloor 2^{nR^{(2)}} \rfloor$ length-n codewords by randomly and independently drawing each entry according to a Bernoulli-$1/2$ distribution. Arrange the codewords in an array with $\lfloor 2^{nR^{(1)}} \rfloor$ rows and $\lfloor 2^{nR^{(2)}} \rfloor$ columns. The transmitter sends the codeword that lies in the row corresponding to Message $W_{d_2}^{(1)}$ and the column corresponding to Message $W_{d_2}^{(2)}$.

Receiver 2 decodes both messages $W_{d_1}^{(2)}$ and $W_{d_2}^{(2)}$. Receiver 1 retrieves Message $W_{d_2}^{(1)}$ from its cache memory and decodes $W_{d_2}^{(2)}$ using an optimal decoding rule for the row-codebook corresponding to $W_{d_2}^{(1)}$. If $d_1 \neq d_2$, it XORs the decoded $W_{d_2}^{(2)}$ with the XOR $W_1^{(2)} \oplus W_2^{(2)}$ stored in its cache memory.

Analysis: Due to the choice of rates $R^{(1)}$ and $R^{(2)}$ in (135), the probability of decoding error tends to 0 as the blocklength n tends to infinity. Since $\epsilon > 0$ can be chosen arbitrarily close to 0, we have proved achievability of the rate-memory pair in (134).

Appendix G
Proof of Upper Bound in Theorem 5

The first and last terms in (87) are special cases of Theorem 4 for $k_{\text{w}} = 1$ and $k_{\text{w}} = 0$, respectively. Here, we prove the second term by showing that for every achievable rate-memory pair $(R,\ M)$,

$$3R \leq M + (1 - \delta_{\text{w}}) + (1 - \delta_{\text{s}}). \tag{137}$$

Since the capacity-memory tradeoff only depends on the conditional marginal distributions of the channel law (4), we will assume that the erasure BC is physically degraded. So, for each $t \in \{1, \ldots, n\}$,

$$X_t \to Y_{2,t} \to Y_{1,t}. \tag{138}$$

For all sufficiently large blocklengths n, choose placement functions $\{g_i^{(n)}\}$ as in (10), encoding functions $f_{\mathbf{d}}^{(n)}$ as in (11), and decoding functions $\{\varphi_{i,\mathbf{d}}^{(n)}\}$ as in (13) so that the probability of worst-case error $P_{\text{e}}^{\text{worst}}$ tends to 0 as the blocklength $n \to \infty$. Consider now a fixed blocklength n that is sufficiently large for the purposes that we describe in the following. Let

$$V_1^{(n)} = g_1^{(n)}(W_1, W_2), \tag{139}$$

$$X_{\mathbf{d}}^n = f_{\mathbf{d}}^{(n)}(W_1, W_2), \tag{140}$$

denote cache contents and the input of the erasure BC for a given demand vector $\mathbf{d} \in \mathcal{D}^2$ and for above chosen placement and encoding functions. Also, let $Y_{1,\mathbf{d}}^n$ and $Y_{2,\mathbf{d}}^n$ denote the corresponding channel outputs at the weak and strong receivers.

We focus on the two demand vectors

$$\mathbf{d}_1 := (1,2) \qquad \text{and} \qquad \mathbf{d}_2 := (2,1).$$

So, W_1 should be decodable from $(Y_{1,\mathbf{d}_1}^n, V_1^{(n)})$ and from $Y_{2,\mathbf{d}_2}^n$, and W_2 should be decodable from $(Y_{1,\mathbf{d}_2}^n, V_1^{(n)})$. Thus, by Fano's inequality, for all $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ and sufficiently large blocklength n, we have

$$nR \leq I(W_1; V_1^{(n)}, Y_{1,\mathbf{d}_1}^n) + n\epsilon_1, \tag{141a}$$

$$nR \leq I(W_1; Y_{2,\mathbf{d}_2}^n) + n\epsilon_2, \tag{141b}$$

$$nR \leq I(W_2; V_1^{(n)}, Y_{1,\mathbf{d}_1}^n, Y_{1,\mathbf{d}_2}^n | W_1) + n\epsilon_3, \tag{141c}$$

where for the last inequality we also used the independence of messages W_1 and W_2.

We first develop the second constraint using the chain rule of mutual information:

$$\begin{aligned} nR &\leq \sum_{t=1}^{n} I(W_1; Y_{\mathbf{d}_2,t} | Y_{2,\mathbf{d}_2}^{t-1}) + n\epsilon_2 \\ &\leq (1-\delta_{\text{s}}) \sum_{t=1}^{n} I(W_1; X_{\mathbf{d}_2,t} | Y_{2,\mathbf{d}_2}^{t-1}) + n\epsilon_2. \end{aligned} \tag{142}$$

We then jointly develop the first and the third constraints, where we also define $\epsilon' := \epsilon_1 + \epsilon_3$:

$$\begin{aligned} &2nR \\ &\leq I(W_1, W_2; V_1^{(n)}, Y_{1,\mathbf{d}_1}^n) + I(W_2; Y_{1,\mathbf{d}_2}^n | W_1, V_1^{(n)}, Y_{1,\mathbf{d}_1}^n) \\ &\quad + n\epsilon' \\ &\overset{(a)}{\leq} I(W_1, W_2; V_1^{(n)}) + I(W_1, W_2; Y_{1,\mathbf{d}_1}^n | V_1^{(n)}) \\ &\quad + I(W_2; Y_{2,\mathbf{d}_2}^n | W_1, V_1^{(n)}, Y_{1,\mathbf{d}_1}^n) + n\epsilon' \\ &= I(W_1, W_2; V_1^{(n)}) + \sum_{t=1}^{n} I(W_1, W_2; Y_{1,\mathbf{d}_1,t} | V_1^{(n)}, Y_{1,\mathbf{d}_1}^{t-1}) \\ &\quad + \sum_{t=1}^{n} I(W_2; Y_{2,\mathbf{d}_2,t} | W_1, V_1^{(n)}, Y_{1,\mathbf{d}_1}^n, Y_{2,\mathbf{d}_2}^{t-1}) + n\epsilon' \\ &= I(W_1, W_2; V_1^{(n)}) \\ &\quad + (1-\delta_{\text{w}}) \sum_{t=1}^{n} I(W_1, W_2; X_{\mathbf{d}_1,t} | V_1^{(n)}, Y_{1,\mathbf{d}_1}^{t-1}) \\ &\quad + (1-\delta_{\text{s}}) \sum_{t=1}^{n} I(W_2; X_{\mathbf{d}_2,t} | W_1, V_1^{(n)}, Y_{1,\mathbf{d}_1}^n, Y_{2,\mathbf{d}_2}^{t-1}) + n\epsilon' \\ &\leq I(W_1, W_2; V_1^{(n)}) \\ &\quad + (1-\delta_{\text{w}}) \sum_{t=1}^{n} I(W_1, W_2; X_{\mathbf{d}_1,i} | V_1^{(n)}, Y_{1,\mathbf{d}_1}^{t-1}) \\ &\quad + (1-\delta_{\text{s}}) \sum_{t=1}^{n} I(W_2, V_1^{(n)}, Y_{1,\mathbf{d}_1}^n; X_{\mathbf{d}_2,t} | W_1, Y_{2,\mathbf{d}_2}^{t-1}) + n\epsilon' \\ &\leq nM + n(1-\delta_{\text{w}}) \\ &\quad + (1-\delta_{\text{s}}) \sum_{t=1}^{n} I(W_2, V_1^{(n)}, Y_{1,\mathbf{d}_1}^n; X_{\mathbf{d}_2,t} | W_1, Y_{2,\mathbf{d}_2}^{t-1}) + n\epsilon'. \end{aligned} \tag{143}$$

In (a), we used that the physically degradedness of the channel in (138) implies the Markov chain

$$(W_1, W_2, V_1^{(n)}, Y_{1,\mathbf{d}_1}^n) \to Y_{2,\mathbf{d}_2}^n \to Y_{1,\mathbf{d}_2}^n.$$

Adding up (142) and (143) and letting $\epsilon_1, \epsilon_2, \epsilon_3$ tend to 0, we obtain the missing converse bound in (137), because

$$\begin{aligned} &I(W_2, V_1^{(n)}, Y_{1,\mathbf{d}_1}^n; X_{\mathbf{d}_2,t} | W_1, Y_{2,\mathbf{d}_2}^{t-1}) + I(W_1; X_{\mathbf{d}_2,t} | Y_{2,\mathbf{d}_2}^{t-1}) \\ &= I(W_1, W_2, V_1^{(n)}, Y_{1,\mathbf{d}_1}^n; X_{\mathbf{d}_2,t} | Y_{2,\mathbf{d}_2}^{t-1}) \\ &\leq H(X_{\mathbf{d}_2,t}) \\ &\leq 1. \end{aligned} \tag{144}$$

References

[1] E. Baştuğ, M. Bennis, E. Zeydan, M. A. Kader, A. Karatepe, A. S. Er and M. Debbah, "Big data meets telcos: A proactive caching perspective," *J. commun. and networks* vol. 17, no. 6, pp. 549–557, Dec. 2015.

[2] M. A. Maddah-Ali, U. Niesen, “Fundamental limits of caching,” in *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[3] Z. Chen, P. Fan, and K. B. Letaief, “Fundamental limits of caching: Improved bounds for small buffer users,” *IET Commun.*, vol. 10, no. 17, nov. 2016, pp. 2315–2318.

[4] C. Tian, “A note on the fundamental limits of coded caching,” in *arXiv, 1503.00010*, Feb. 2015.

[5] K. Wan, D. Tuninetti, and P. Piantanida, “On caching with more users than files,” in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016.

[6] S. Sahraei and M. Gastpar, “K users caching two files: An improved achievable rate,” in *Proc. Annu. Conf. Inf. Sciences and Systems*, Princeton, NJ, Mar. 2016.

[7] A. Sengupta, R. Tandon, “Improved approximation of storage-rate tradeoff for caching with multiple demands,” *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 1940–1955, May 2017.

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

[9] H. Ghasemi and A. Ramamoorthy, “Further results on lower bounds for coded caching,” in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016.

[10] C. Tian, “Symmetry, demand types and outer bounds in caching systems,” in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016.

[11] A. Amiri, Q. Yang, and D. Gündüz, “Coded caching for a large number of users,” in *Proc. IEEE Inf. Theory Workshop,* Cambridge, United Kingdom, Sep. 2016.

[12] J. Zhang, X. Lin, and X. Wang, “Coded caching under arbitrary popularity distributions,” in *Proc. Inf. Theory and Appl. Workshop,* San Diego, CA, Feb. 2015.

[13] C.-Y. Wang, S. H. Lim, and M. Gastpar, “A new converse bound for coded caching,” in *Proc. Inf. Theory and Appl. Workshop*, La Jolla, CA, Feb. 2016.

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

[15] J. Gómez-Vilardebó, “Fundamental limits of caching: improved bounds with coded prefetching,” in *arXiv, 1612.09071*, Jan. 2017.

[16] R. Timo, S. Saeedi Bidokhti, M. Wigger and B. Gieger, “A rate-distortion approach to caching,” in *Proc. Int. Zurich Seminar on Commun.*, Zurich, Switzerland, March, 2016.

[17] C.-Y. Wang, S. H. Lim, and M. Gastpar, “Information-theoretic caching: Sequential coding for computing,” *IEEE Trans. Inf. Theory,* vol. 63, no. 7, pp. 4388–4413, Jul. 2017.

[18] S. H. Lim, C.-Y. Wang, and M. Gastpar, “Information theoretic caching: The multi-user case,” *IEEE Trans. Inf. Theory,* vol. 63, no. 7, pp. 4388–4413, Jul. 2017.

[19] Q. Yang and D. Gündüz, “Centralized coded caching for heterogeneous lossy requests,” in *Proc. IEEE Int. Symp. Inf. Theory*, Barcelona, Spain, Jul. 2016, pp. 405–409.

[20] R. Timo and M. Wigger, “Joint cache-channel coding over erasure broadcast channels,” in *Proc. IEEE Int. Symp. Wireless Commun. Systems*, Bruxelles, Belgium, Aug. 2015.

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

[22] S. Wang, X. Tian and H. Liu, “Exploiting the unexploited of coded caching for wireless content distribution,” in *Proc. IEEE Int. Conf. Computing, Networking and Commun.*, Feb. 2015.

[23] P. Hassanzadeh, E. Erkip, J. Llorca,and A. Tulino, “Distortion-memory tradeoffs in Cache-aided wireless video delivery,” in *Proc. Allerton Conf. Commun., Control and Computing* Monticello (IL), USA, Oct. 2015.

[24] A. S. Cacciapuoti, M. Caleffi, M. Ji, J. Llorca, A. M. Tulino,“Speeding up future video distribution via channel-aware caching-aided coded multicast,” *IEEE J. Selected Areas in Commun.*, vol. 34, no. 8, Aug., 2016.

[25] J. Hachem, N. Karamchandani, and S. Diggavi, “Content caching and delivery over heterogeneous wireless networks,” *arXiv*, 1404.6560, March 2015.

[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] A. Ghorbel, K.H. Ngo, R. Combes, M. Kobayashi, S. Yang, “Opportunistic content delivery in fading broadcast channels,” *arXiv*, 1702.02179, Feb. 2017.

[28] K. H. Ngo, S. Yang, M. Kobayashi, “Scalable content delivery with coded caching in multi-antenna fading channels,” *arXiv*, 1703.06538, March 2017.

[29] 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.

[30] J. Zhang, P. Elia, “Wireless coded caching: A topological perspective,” in *Proc. IEEE Int. Symp. Inf. Theory*, June 2017, Aachen, Germany, pp. 401–405.

[31] J. Zhang and P. Elia, “Feedback-aided coded caching for the MISO BC with small caches,” in *Proc. Int. Conf. Commun.*, May 2017, Paris, France.

[32] S. P. Shariatpanahi, G. Caire, and B. H. Khalaj, “Multi-antenna coded caching,” *arXiv*, 1701.02979, Jan. 2017.

[33] X. Yi and G. Caire, “Topological coded caching,” in *Proc. IEEE Int. Symp. Inf. Theory*, July 2016, Barcelona, Spain, pp. 2039–2043.

[34] E. Piovano, H. Joudeh, B. Clerckx, “On coded caching in the overloaded MISO broadcast channel,” *arXiv* 1702.01672.

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

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

[37] G. Kramer and S. Shamai, “Capacity for classes of broadcast channels with receiver side information,” in *Proc. IEEE Inf. Theory Workshop*, Lake Tahoe, California, September, 2007.

[38] T. J. Oechtering, C. Schnurr, I. Bjelakovic, and H. Boche, “Broadcast capacity region of two-phase bidirectional relaying,” *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 454–458, Jan. 2008.

[39] R. F. Wyrembelski, A. Sezgin and H. Boche, ”Secrecy in broadcast channels with receiver side information,” *Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, 2011, pp. 290-294.

[40] Y. Gao and E. Tuncel, “Wyner-Ziv coding over broadcast channels: hybrid digital/analog schemes,” *IEEE Trans. Inf. Theory*, vol. 57, no. 9, pp. 5660–5672, September, 2009.

[41] J. Nayak, E. Tuncel, and D. Gunduz “Wyner-Ziv coding over broadcast channels: digital schemes,” *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1782–1799, Apr. 2010.

[42] R. Timo, A. Grant, and G. Kramer, “Lossy broadcasting with complementary side information,” *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 104–131, Jan. 2013.

[43] R. Timo and M. Wigger, ”Slepian-Wolf coding for broadcasting with cooperative base-stations,” *IEEE Trans. on Commun.*, vol. 63, no. 5, pp. 1850–1866, May 2015.

[44] A. Bracher and M. Wigger, “Feedback and partial message side-information on the semideterministic broadcast channel,” *IEEE Trans. Inf. Theory,* vol. 63, no. 7, pp. 4388–4413, Jul. 2017.

[45] B. Asadi, L. Ong, and S. J. Johnson, “A unified inner bound for the two-receiver memoryless broadcast channel with channel state and message side information,” *arXiv*, 1601.03121, Jan. 2016.

[46] T. Laich and M. Wigger, ”Utility of encoder side information for the lossless Kaspi/Heegard-Berger problem, ” in *Proc. IEEE Int. Symp. Inf. Theory*, July 2013, pp. 3065–3069.

[47] S. Diggavi, E. Perron, and E. Telatar, ”Lossy source coding with Gaussian or erased side-information,” in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, Korea, June 28–July 3 2009, pp. 1035–1039.

[48] Y. Wu and M. Wigger, “Coding schemes with rate-limited feedback that improve over the nofeedback capacity for a large class of broadcast channels,” *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 2009–2033, Apr. 2016.

[49] O. Shayevitz and M. Wigger, “On the capacity of the discrete memoryless broadcast channel with feedback,” *IEEE Trans. Inf. Theory*, vol. 59, no. 3, pp. 1329–1345, Mar. 2013.

[50] R. Venkataramanan and S. S. Pradhan, “An achievable rate region for the broadcast channel with feedback,” *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6175–6191, Oct. 2013.

[51] C. C. Wang, “Capacity of 1-to-k broadcast erasure channel with output feedback,” in *Proc. Allerton Conf. Commun., Control and Computing*, Monticello, IL, Sep. 2010.

[52] L. Georgiadis and L. Tassiulas, “Broadcast erasure channel with feedback capacity and algorithms,” *in Proc. Workshop on Network Coding, Theory, & Applications,* Lausanne, Switzerland, June 2009.

[53] M. A. Maddah-Ali and D. Tse, “Completely Stale Transmitter Channel State Information is Still Very Useful,” *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, July 2012.

[54] H. Kim, Y.-K. Chia, and A. El Gamal, “A note on broadcast channels with stale state information at the transmitter,” *IEEE Trans. Inf. Theory*, vol. 61, no. 7, pp. 3622–3631, July 2015.

[55] Y. Wu, "Coding schemes for discrete memoryless multicast networks with and without Feedback," in *Proc. Allerton Conf. Commun., Control and Computing*, Monticello, IL, Sept. 2015.

[56] J. Chen and P. Elia, "Toward the performance versus feedback tradeoff for the two-user MISO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8336–8356, Dec. 2013.

[57] S. Yang, M, Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.

[58] 3GPP TR 38.913 "Study on scenarios and requirements for next generation access technologies," V14.3.0, 2017.

[59] R. Urbanke and A. Wyner, "Packetizing for the erasure broadcast channel with an internet application," technical report, 1999.

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

[61] A. El Gamal and Y. H. Kim, "Network information theory," Cambridge University Press, 2012.

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

[63] N. Naderializadeh, M. Maddah-Ali, and A. Avestimehr, "On the Optimality of Separation between Caching and Delivery in General Cache Networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Aachen, Germany, Jun. 2017.

[64] M. M. Amiri and D. Gündüz, "Cache-aided content delivery over erasure broadcast channels." *IEEE Trans. on Comm.*, vol. 66, no. 1, pp. 370–381, Jan. 2018.

[65] S. Saeedi Bidokhti, M. Wigger and A. Yener, "Benefits of cache assignment on degraded broadcast channels." *ArXiv:1702.08044*, Feb. 2017.