Noisy Broadcast Networks with Receiver Caching

Shirin Saeedi Bidokhti, Michèle Wigger and Roy Timo

Abstract

The paper studies an erasure broadcast network 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 bounds match when there is a single weak receiver (and any number of strong receivers) and the cache size does not exceed a given threshold. Improved bounds are presented for the special case with a single weak receiver, a single strong receiver, and two files; these new bounds match over a larger regime of cache sizes than the previous bounds. The lower bound is achieved by a new *joint cache-channel coding scheme* and significantly improves over traditional schemes based on separate cache-channel coding. The upper bound holds for all stochastically degraded broadcast channels.

The derived upper and lower bounds on the capacity-memory tradeoff are also converted to lower and upper bounds on the delivery rate-memory tradeoff.

I. Introduction

We address a one-to-many broadcast communications 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 at the network edge during off-peak times when the communications 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 behaviour and identify files that are popular in peak-traffic [1].

The above caching problem is particularly relevant to video-streaming services in mobile networks. Here network operators pre-place information in clients' caches (or, on servers near the clients) to improve latency and throughput during peak-traffic times. The network operator does not know in advance which movies the clients will request, and thus the cached information cannot depend on the clients' specific demands. It is now widely expected that there will be a nine-fold increase in mobile data traffic by 2020, and around 60 percent of this traffic will be mobile video [2]. Smart data caching strategies, new bandwidth allocations, reduced cell sizes and new radio-access technologies will all be needed to meet these growing demands [3].

The information-theoretic aspects of cache-aided communications have received significant attention in recent years [4]–[46]. Maddah-Ali and Niesen [4] 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 perfect, noise-free broadcast channel (BC) where each receiver directly observes the inputs. They showed that a smart design of the cache contents enables the server to send coded (XOR-ed) data during the delivery phase that can 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 was thus termed [4] *global caching gain*. Several recent works [4]–[18] have presented upper and lower bounds on the minimum delivery rate as a function of the cache sizes. The works in [19]–[22] considered related scenarios but where the various files can be correlated.

In this paper we assume that the delivery phase takes place over a noisy BC, and we will see 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 adopteda *separate cache-channel coding architectures* with encoder/decoders consisting of a cache encoder/decoder and a channel encoder/decoder that only depend on the cache contents or only on 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 [50]–[60]. Tuncel [50], 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 [59], [60], a scenario that also relates to the BC with *feedback* because the fed back channel

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

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

Parts of the material in this paper have been presented at the *IEEE International Symposium on Wireless Communications Systems (ISWCS)*, Bruxelles, Belgium, August 2015 [23], and will be presented at the *2016 IEEE International Symposium on Information Theory*, Barcelona, Spain and at the *International Symposium on Turbo Codes & Iterative Information Processing*, Brest, France, September 2016.

outputs can be viewed as such a side-information. The schemes in [57], [61]–[70] exactly exploit this side-information to improve over the nofeedback capacity region of the considered BCs. Joint source-channel coding is however only used in [57], [61].

The importance of including a noisy channel model for the delivery phase was also observed in [23]–[30], [32]–[35]. For example, [29]–[32] illustrate interesting interplays between feedback, channel state information, and massive MIMO with caching; [36]–[38] show that in Gaussian interference networks caches at the transmitter-side and receiver-side allow for load-balancing and interference mitigation in noisy interference networks; and [38]–[46] focus on cellular networks where caching allows to cancel inter-cell interference [39].

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 K 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 M.
- 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.

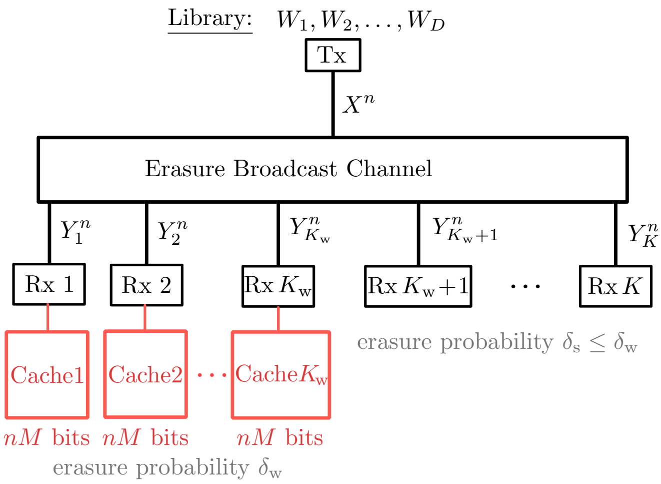

Fig. 1: K user packet-erasure BC with K_{w} weak and K_{s} strong receivers and where the weak receivers have cache memories.

This scenario is motivated by previous studies [23], [24] 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.

A lower bound on the capacity-memory tradeoff $C(M)$ is presented for the network in Figure 1, as well as an upper bound for general degraded BCs with arbitrary receiver cache memories. The 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

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

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 (1) or large:

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

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

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

The proposed lower bounds are based on joint cache-channel coding building on the *piggyback coding* idea in [23], [61]. 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. Notice that (after recasting the cached contents as message side-information) Tuncel-coding [50] could be used instead of piggyback coding. The former is however more complicated and includes binning etc. The new lower bounds substantially improve over the obvious separate cache-channel coding scheme that combines coded caching [4] with a capacity-achieving scheme for the erasure BC. For example, when M is smaller than a given threshold that depends on the problem parameters, the joint cache-channel coding scheme achieves the following lower bound on the capacity-memory tradeoff:

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

Here, D denotes the number of files in the system; 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

$$\gamma_\mathrm{joint} := 1 + \frac{2K_\mathrm{w}}{1+K_\mathrm{w}} \cdot \frac{K_\mathrm{s}(1-\delta_\mathrm{w})}{K_\mathrm{w}(1-\delta_\mathrm{s})} \geq 1 \tag{3}$$

describes this scheme's gain over separate cache-channel coding. That means, separate cache-channel coding achieves the lower bound in (2) but when γ_joint is replaced by 1. Inequalities (2) and (3) show that the improvement of our joint cache-channel coding schemes over the separate cache-channel coding scheme 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. On a more intuitive level, the benefit of joint cache-channel coding is that it can provide global caching gains also to the strong receivers. That means, the effective communication rate to the strong receivers can be decreased in function of the cache memories available at the weak receivers.

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 [4].

A. Paper Organization

In Section II we state the problem setup. Section III presents a separate-cache channel coding scheme based on coded caching and Section IV a new 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 identified by uppercase letters, e.g. A, their alphabets by matching calligraphic font, e.g. $\mathcal{A}$, and elements of an alphabet by lowercase letters, e.g. $a \in \mathcal{A}$. We also use uppercase letters for deterministic quantities like rate R, capacity C, number of users K, cache size M, and number of files in the library D. 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 $(A_1, \ldots, A_n)$.

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

Finally, the notation $W_1 \oplus W_2$ is used for the bitwise XOR over the binary strings corresponding to the messages W_1 and W_2, where these strings are assumed to be of equal length.

B. Message and Channel Models

Consider a broadcast channel (BC) with a single transmitter and K receivers as depicted in Figure 1. We have two sets of receivers: K_w weak receivers that statistically have a bad channel and $K_\mathrm{s} = K - K_\mathrm{w}$ strong receivers that statistically have a good channel. (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 sets

$$\mathcal{K}_\mathrm{w} := \{1, \ldots, K_\mathrm{w}\}$$

and

$$\mathcal{K}_\mathrm{s} := \{K_\mathrm{w}+1, \ldots, K\}.$$

We model the channel from the transmitter to the receivers by a memoryless *erasure BC*[1] with input alphabet

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

[1]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 this packet size.

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 $k \in \mathcal{K} := \{1, \ldots, K\}$ 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[2] of the memoryless BC are thus described by

$$\mathbb{P}[Y_k = y_k | X = x] = \begin{cases} 1 - \delta_k & \text{if } y_k = x \\ \delta_k & \text{if } y_k = \Delta \\ 0 & \text{otherwise} \end{cases} \qquad \forall\, k \in \mathcal{K}. \tag{4}$$

We will assume throughout that

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

for fixed erasure probabilities[3]

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

Since $\delta_{\text{s}} \leq \delta_w$, 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. We explain shortly how the cache memory at the weak receivers can be exploited.

C. Message Library and Receiver Demands

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

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

These messages are all independent of each other, each being uniformly distributed over the message set $\{1, \ldots, \lfloor 2^{nR} \rfloor\}$, where $R \geq 0$ is the rate of each message and n the blocklength of transmission.

Each receiver will demand (i.e., request and download) exactly one of these 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 demand vector

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

can take on any value in $\mathcal{D}^K$.

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 first communication phase the transmitter sends caching information V_i to each weak receiver $i \in \mathcal{K}_{\text{w}}$, which then stores this information in its cache memory. The strong receivers do not take part in the placement phase.

The 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 the users' specific demands $\mathbf{d}$. Instead, V_i is a function of the entire library only:

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

for some function

$$g_i \colon \left\{1, \ldots, \lfloor 2^{nR} \rfloor\right\}^D \to \mathcal{V}, \qquad i \in \mathcal{K}_{\text{w}}, \tag{9}$$

where

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

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 $i \in \mathcal{K}_{\text{w}}$ can thus store V_i in its cache memory.

[2]As will become clear in the following, for our problem setup only this marginal transition law is relevant, but not the joint transition law.

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

E. Delivery Phase

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 packet-erasure BC. The entire demand vector $\mathbf{d}$ is assumed to be known to the transmitter and all receivers[4].

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

$$f_{\mathbf{d}} \colon \{1, \ldots, \lfloor 2^{nR} \rfloor\}^D \to \mathcal{X}^n, \tag{10}$$

and it sends

$$X^n = f_{\mathbf{d}}(W_1, \ldots, W_D), \tag{11}$$

over the packet-erasure BC.

Each Receiver $k \in \mathcal{K}$ observes Y_k^n according to the memoryless transition law in (4). Each weak receiver attempts to reconstruct its desired message from it channel output, cache contents and demand vector $\mathbf{d}$. Similarly, each strong receiver attempts to reconstruct its desired message from its channel output and the demand vector $\mathbf{d}$. More formally,

$$\hat{W}_i := \begin{cases} \varphi_{i,\mathbf{d}}(Y_i^n, V_i) & \text{if } i \in \mathcal{K}_{\text{w}} \\ \varphi_{i,\mathbf{d}}(Y_i^n) & \text{if } i \in \mathcal{K}_{\text{s}}, \end{cases} \tag{12a}$$

where

$$\varphi_{i,\mathbf{d}} \colon \mathcal{Y}^n \times \mathcal{V} \to \{1, \ldots, \lfloor 2^{nR} \rfloor\}, \qquad i \in \mathcal{K}_{\text{w}}, \tag{12b}$$

and

$$\varphi_{i,\mathbf{d}} \colon \mathcal{Y}^n \to \{1, \ldots, \lfloor 2^{nR} \rfloor\}, \qquad i \in \mathcal{K}_{\text{s}}. \tag{12c}$$

F. Capacity-Memory Tradeoff

An error occurs whenever

$$\hat{W}_k \neq W_{d_k} \quad \text{for some} \quad k \in \mathcal{K}. \tag{13}$$

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

$$\mathsf{P}_{\text{e}}(\mathbf{d}) := \mathbb{P}\left[\bigcup_{k=1}^{K} \hat{W}_k \neq W_{d_k} \right]. \tag{14}$$

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

$$\mathsf{P}_{\text{e}}^{\text{worst}} := \max_{\mathbf{d} \in \mathcal{D}^K} \mathsf{P}_{\text{e}}(\mathbf{d}). \tag{15}$$

In Definitions (9)–(15), we sometimes add a superscript (n) to emphasise the dependency on the blocklength n.

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 (9), (10) and (12) such that $\mathsf{P}_{\text{e}}^{\text{worst}} < \epsilon$. The main problem of interest in this paper is to determine the following capacity versus cache-memory tradeoff.

Definition 1: Given 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 to all K receivers [71]:

$$C(M = 0) = R_0, \tag{16}$$

where

$$R_0 := \left(\frac{K_{\text{w}}}{1 - \delta_{\text{w}}} + \frac{K_{\text{s}}}{1 - \delta_{\text{s}}} \right)^{-1}. \tag{17}$$

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

$$C(M) \leq \frac{1 - \delta_{\text{s}}}{K_{\text{s}}}, \quad \forall\ M \geq 0. \tag{18}$$

[4]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.

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,

$$C(M) = \frac{1-\delta_{\text{s}}}{K_{\text{s}}}, \qquad \forall\, M \geq D \cdot \frac{1-\delta_{\text{s}}}{K_{\text{s}}}. \tag{19}$$

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 capacity-achieving codes for the erasure BC and Maddah-Ali and Niesen's coded caching scheme [4]. 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 $\mathcal{K}_{\text{w}}$ of size $\tilde{t}$, and split each message W_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 being of equal rate

$$R_{\text{sub}} := R \binom{K_{\text{w}}}{\tilde{t}}^{-1}. \tag{20}$$

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

`Method Ca`: Takes as input the entire library $W_1,\ldots,W_D$ and outputs the cache contents

$$V_i = \left\{ W_{d,\mathcal{G}_\ell}\colon\;\; d \in \{1,\ldots,D\} \text{ and } i \in \mathcal{G}_\ell \right\}, \qquad i \in \mathcal{K}_{\text{w}}. \tag{21}$$

In other words, when $\tilde{t} \neq 0$, then during the placement phase, the tuple

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

is stored in the cache memory of every Receiver 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

$$W_{d_{K_{\text{w}}+1}},\;\, W_{d_{K_{\text{w}}+2}},\;\, \ldots,\;\, W_{d_K} \tag{22}$$

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 library $W_1, \ldots, W_D$ and the reduced demand vector $\mathbf{d}_\text{w} := (d_1, \ldots, d_{K_\text{w}})$. It produces the outputs
>
> $$\left\{W_{\text{XOR},\mathcal{S}}: \quad \mathcal{S} \subseteq \mathcal{K}_\text{w} \text{ and } |\mathcal{S}| = \tilde{t}+1\right\}, \tag{23}$$
>
> where
>
> $$W_{\text{XOR},\mathcal{S}} := \bigoplus_{k\in\mathcal{S}} W_{d_k,\mathcal{S}\backslash\{k\}}. \tag{24}$$

The transmitter then uses a capacity-achieving scheme for erasure BCs to send the messages in (24) to all[5] weak receivers $\mathcal{K}_\text{w}$ and the messages in (22) to all strong receivers $\mathcal{K}_\text{s}$.

4) Delivery-Decoding: The strong receivers decode their intended messages in (22) 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 $\tilde{t} = K_\text{w}$, the weak receivers can directly retrieve their desired messages from their cache memories.

If $1 \leq \tilde{t} \leq K_\text{w} - 1$, the weak receivers first decode the XOR-messages in (24) using a capacity-achieving decoder for the erasure BC. Each Receiver $i \in \mathcal{K}_\text{w}$ then applies the following Method $\texttt{De}_i$:

> Method $\texttt{De}_i$: Takes as inputs the demand vector $\mathbf{d}_\text{w}$, the decoded messages $\{W_{\text{XOR},\mathcal{S}} : i \in \mathcal{S}\}$, and the cache content V_i. It outputs the reconstruction
>
> $$\hat{W}_i := \left(\hat{W}_{d_i,\mathcal{G}_1}, \ldots, \hat{W}_{d_i,\mathcal{G}_{\binom{K_\text{w}}{\tilde{t}}}}\right), \tag{25}$$
>
> where
>
> $$\hat{W}_{d_i,\mathcal{G}_\ell} = \begin{cases} W_{d_i,\mathcal{G}_\ell} & \text{if } i \in \mathcal{G}_\ell \\ \left(\bigoplus_{s\in\mathcal{G}_\ell} W_{d_s,\mathcal{G}_\ell\cup\{i\}\backslash\{s\}}\right) \oplus W_{\text{XOR},\mathcal{G}_\ell\cup\{i\}} & \text{if } i \notin \mathcal{G}_\ell. \end{cases} \tag{26}$$

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

$$R_{\tilde{t},\text{sep}} := \left(\frac{K_\text{w} - \tilde{t}}{(\tilde{t}+1)(1-\delta_\text{w})} + \frac{K_\text{s}}{(1-\delta_\text{s})}\right)^{-1}, \tag{27a}$$

and it requires the weak receivers to have cache memories of size

$$M_{\tilde{t},\text{sep}} := D\frac{\tilde{t}}{K_\text{w}} R_{\tilde{t},\text{sep}}. \tag{27b}$$

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

Proposition 1: The upper convex hull of the rate-memory pairs in (27) is achievable:

$$C(M) \geq \text{upp hull}\left(\left\{(R_{\tilde{t},\text{sep}}, M_{\tilde{t},\text{sep}}); \quad \tilde{t} \in \{0, \ldots, K_\text{w}\}\right\}\right). \tag{28}$$

IV. A Joint Cache-Channel Coding Scheme

We now describe a joint cache-channel coding scheme for the scenario in Figure 2 with $K_\text{w} = 3$ weak receivers and $K_\text{s} = 1$ strong receivers. The general scheme is parameterized by a positive integer $t \in \mathcal{K}_\text{w}$. The example here corresponds to $t = 2$.

A. A Simple Example

1) Scheme: Define

$$R^{(1)} := \frac{1-\delta_\text{w}}{1-\delta_\text{s}} R, \tag{29a}$$

$$R^{(2)} := \frac{\delta_\text{w}-\delta_\text{s}}{1-\delta_\text{s}} R, \tag{29b}$$

where we notice that

$$\frac{R^{(1)}}{R^{(2)}} = \frac{1-\delta_\text{w}}{\delta_\text{w}-\delta_\text{s}}. \tag{30}$$

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

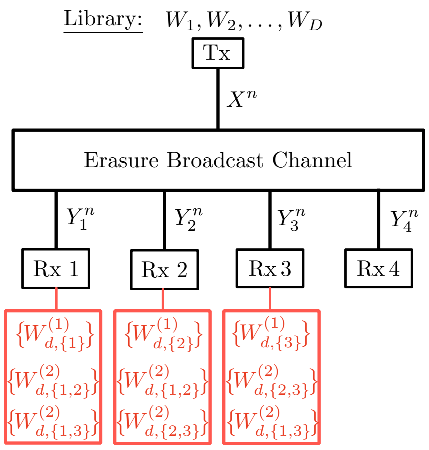

Fig. 2: An example network with 3 weak and a single strong receiver. The figure illustrates the contents cached in the proposed joint cache-channel coding scheme when $t = 2$.

Split each message W_d into 2 submessages

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

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

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

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

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

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

Placement Phase: Cache all messages $\{W^{(1)}_{d,\{i\}}\}_{d=1}^{D}$ at Receiver i, for $i \in \{1,2,3\}$, and all messages $\{W^{(2)}_{d,\{i,j\}}\}_{d=1}^{D}$ 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 `Ca` given in the previous Section III: First apply method `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. Table I shows the messages transmitted in the various subphases. Notice that given the cache contents in Figure 2, each receiver can recover its desired message without errors, if the submessages in Table I are correctly decoded by the appropriate receivers.

	Subphase 1	Subphase 2	Subphase 3
Messages for 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 for 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.

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

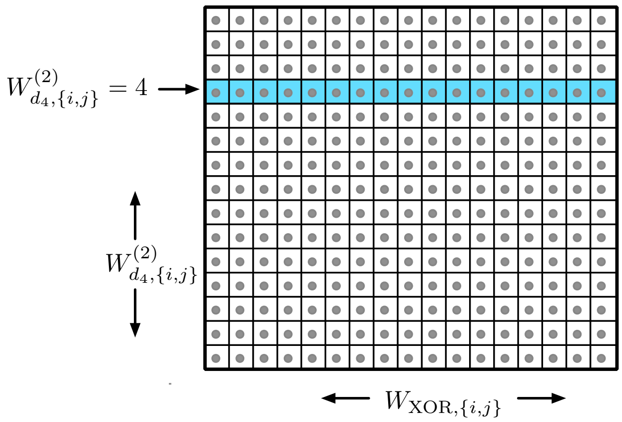

Fig. 3: Codebook used for piggyback coding in Period $\{i, j\}$ of Subphase 2.

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 the XOR-message $W_{\text{XOR},\{1,2,3\}}$ as $\hat{W}_{\text{XOR},\{1,2,3\}}$, then it retrieves messages $W^{(2)}_{d_3,\{1,2\}}$ and $W^{(2)}_{d_2,\{1,3\}}$ from its cache memory and produces:

$$\hat{W}^{(2)}_{d_1,\{2,3\}} := W^{(2)}_{d_3,\{1,2\}} \oplus W^{(2)}_{d_2,\{1,3\}} \oplus \hat{W}_{\text{XOR},\{1,2,3\}}. \tag{31}$$

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 the only one using 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

$$W_{\text{XOR},\{i,j\}} := W^{(1)}_{d_j,\{i\}} \oplus W^{(1)}_{d_i,\{j\}} \tag{32a}$$

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

$$W^{(2)}_{d_4,\{i,j\}} \tag{32b}$$

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 (32), 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.

The strong receiver 4 decodes both messages $W_{\text{XOR},\{i,j\}}$ and $W^{(2)}_{d_4,\{i,j\}}$ using a standard decoder. It will further use only the guess $\hat{W}^{(2)}_{d_4,\{i,j\}}$.

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

$$\mathcal{C}_{i,j,\text{row}}\left(W^{(2)}_{d_4,\{i,j\}}\right) := \left\{x^{n_{2,\text{per}}}_{i,j}\left(W^{(2)}_{d_4,\{i,j\}},\, w\right)\right\}_{w=1}^{\left\lfloor 2^{nR^{(1)}/3}\right\rfloor}. \tag{33}$$

(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 guess $\hat{W}_{\text{XOR},\{i,j\}}$ 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 to this receiver. At the end of Subphase 3, Receiver 4 applies a standard decoder and produces the guess $\hat{W}^{(1)}_{d_4}$.

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

$$\hat{W}_1 := \left(W^{(1)}_{d_1,\{1\}}, \hat{W}^{(1)}_{d_1,\{2\}}, \hat{W}^{(1)}_{d_1,\{3\}}, W^{(2)}_{d_1,\{1,2\}}, W^{(2)}_{d_1,\{1,3\}}, \hat{W}^{(2)}_{d_1,\{2,3\}}\right); \tag{34}$$

$$\hat{W}_2 := \left(\hat{W}^{(1)}_{d_2,\{1\}}, W^{(1)}_{d_2,\{2\}}, \hat{W}^{(1)}_{d_2,\{3\}}, W^{(2)}_{d_2,\{1,2\}}, \hat{W}^{(2)}_{d_2,\{1,3\}}, W^{(2)}_{d_2,\{2,3\}}\right); \tag{35}$$

$$\hat{W}_3 := \left(\hat{W}^{(1)}_{d_3,\{1\}}, \hat{W}^{(1)}_{d_3,\{2\}}, W^{(1)}_{d_3,\{3\}}, \hat{W}^{(2)}_{d_3,\{1,2\}}, W^{(2)}_{d_3,\{1,3\}}, W^{(2)}_{d_3,\{2,3\}}\right); \tag{36}$$

$$\hat{W}_4 := \left(\hat{W}^{(1)}_{d_4}, \hat{W}^{(2)}_{d_4,\{1,2\}}, \hat{W}^{(2)}_{d_4,\{1,3\}}, \hat{W}^{(2)}_{d_4,\{2,3\}}\right). \tag{37}$$

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

$$M = D\left(\frac{1}{3}R^{(1)} + \frac{2}{3}R^{(2)}\right) = D \cdot \frac{1-\delta_{\text{s}}+\delta_{\text{w}}-\delta_{\text{s}}}{3(1-\delta_{\text{s}})}R \tag{38}$$

The probability of error in Subphase 1 tends to 0 as $n \to \infty$, if

$$\overline{\lim_{n\to\infty}}\, \frac{n}{n_1} \cdot \frac{1}{3}R^{(2)} < 1 - \delta_{\text{w}}. \tag{39}$$

The probability of error of each period in Subphase 2 tends to 0 as $n \to \infty$, if

$$\overline{\lim_{n\to\infty}}\, \frac{n}{n_2/3} \cdot \frac{1}{3}R^{(1)} < 1 - \delta_{\text{w}}, \tag{40}$$

and if

$$\overline{\lim_{n\to\infty}}\, \frac{n}{n_2/3} \cdot \frac{1}{3}\big(R^{(1)} + R^{(2)}\big) < 1 - \delta_{\text{s}}. \tag{41}$$

Here, Condition (40) ensures that the probability of decoding error at the weak receivers vanishes, because they decode their desired messages based on a row-codebook containing only $\lfloor 2^{nR^{(1)}/3}\rfloor$ codewords. Condition (41) ensures that the probability of decoding error at the strong receiver vanishes. By (30), Conditions (40) and (41) are equivalent, and we drop (41) in the following.

The probability of error in Subphase 3 tends to 0 as $n \to \infty$, if

$$\overline{\lim_{n\to\infty}}\, \frac{n}{n_3}R^{(1)} < 1 - \delta_{\text{s}}. \tag{42}$$

In summary, since $n_1 + n_2 + n_3 = n$, whenever

$$\frac{R^{(2)}}{3(1-\delta_{\text{w}})} + \frac{R^{(1)}}{1-\delta_{\text{w}}} + \frac{R^{(1)}}{1-\delta_s} \leq 1, \tag{43}$$

there exist appropriate choices of the lengths n_1, n_2, n_3 (as a function of the total blocklength n) so that the probability of error tends to 0 as $n \to \infty$.

Notice finally that by (29), Condition (43) is equivalent to

$$R \leq (1-\delta_{\text{s}})\left(\frac{\delta_{\text{w}}-\delta_{\text{s}}}{3(1-\delta_{\text{w}})} + 1 + \frac{1-\delta_{\text{w}}}{1-\delta_{\text{s}}}\right)^{-1}, \tag{44}$$

and the described scheme achieves any rate $R > 0$ satisfying (44).

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

$$W^{(2)}_{d_4,\{1,2\}}, W^{(2)}_{d_4,\{1,3\}}, W^{(2)}_{d_4,\{2,3\}} \tag{45}$$

can be piggybacked on the communications of the XOR messages

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

without harming the performance of this latter. In fact, by our choice (30), the probability of error in Subphase 2 vanishes whenever Condition (40) holds, which coincides with the required condition when solely Messages (46) are transmitted but not Messages (45).

If in Subphase 2 the weak receivers applied 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 (43), or equivalently, the additional summand

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

would appear on the RHS of (44). In other words, joint cache-channel coding allows reducing the communicated message rate to the strong receiver 4 from $\frac{R}{1-\delta_\text{s}}$ to $\frac{R^{(2)}}{1-\delta_\text{s}}$. In this sense, joint cache-channel coding can provide global caching gains also to receivers without cache memories.

B. General Scheme

The general scheme is parameterised by a positive integer

$$t \in \mathcal{K}_\text{w}. \tag{47}$$

We show in Appendix A that, for a given parameter t, this scheme achieves the rate-memory pair

$$R_t := \frac{(1-\delta_\text{w})\left(1+\dfrac{K_\text{w}-t+1}{tK_\text{s}}\ \dfrac{\delta_\text{w}-\delta_\text{s}}{1-\delta_\text{w}}\right)}{\dfrac{K_\text{w}-t+1}{t}\left(1+\dfrac{K_\text{w}-t}{(t+1)K_\text{s}}\ \dfrac{\delta_\text{w}-\delta_\text{s}}{1-\delta_\text{w}}\right)+K_\text{s}\dfrac{1-\delta_w}{1-\delta_\text{s}}}, \tag{48a}$$

$$M_t := R_t\ \frac{D}{K_\text{w}}\left(t-\left(1+\frac{K_\text{w}-t+1}{tK_\text{s}}\ \frac{\delta_\text{w}-\delta_\text{s}}{1-\delta_\text{w}}\right)^{-1}\right). \tag{48b}$$

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

$$W_d = \left(W_d^{(t-1)},\ W_d^{(t)}\right) \tag{49}$$

of rates

$$R^{(t-1)} = R\cdot\frac{tK_\text{s}(1-\delta_\text{w})}{(K_\text{w}-t+1)(\delta_\text{w}-\delta_\text{s})+tK_\text{s}(1-\delta_\text{w})}, \tag{50a}$$

$$R^{(t)} = R\cdot\frac{(K_\text{w}-t+1)(\delta_\text{w}-\delta_\text{s})}{(K_\text{w}-t+1)(\delta_\text{w}-\delta_\text{s})+tK_\text{s}(1-\delta_\text{w})}. \tag{50b}$$

Notice that $R^{(t-1)}+R^{(t)}=R$.

2) Placement phase: First, apply the coded-caching Method $\texttt{Ca}$ (from the previous Section III-2) with parameter $\tilde{t}=t$ to messages $W_1^{(t)},\ldots,W_D^{(t)}$ to produce the cache contents

$$V_i^{(t)} = \left\{W^{(t)}_{d,\mathcal{G}^{(t)}_\ell}:\ d\in\mathcal{D} \text{ and } i\in\mathcal{G}^{(t)}_\ell\right\},\quad i\in\mathcal{K}_\text{w}.$$

Here, $\mathcal{G}_1^{(t)},\ldots,\mathcal{G}^{(t)}_{\binom{K_\text{w}}{t}}$ denote the $\binom{K}{t}$ subsets of $\mathcal{K}_\text{w}$ of size t, and each message $W^{(t)}_{d,\mathcal{G}^{(t)}_\ell}$ is of equal rate

$$R^{(t)}_\text{sub} = R^{(t)}\cdot\binom{K_\text{w}}{t}^{-1}. \tag{51a}$$

Then, apply Method Ca with parameter $\tilde{t} = t-1$ to messages $W_1^{(t-1)}, \ldots, W_D^{(t-1)}$ to produce the cache contents

$$V_i^{(t)} = \left\{ W^{(t-1)}_{d,\mathcal{G}_\ell^{(t-1)}} : \ d \in \mathcal{D} \text{ and } i \in \mathcal{G}_\ell^{(t-1)} \right\}, \quad i \in \mathcal{K}_\text{w}.$$

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

$$R_\text{sub}^{(t-1)} = R^{(t-1)} \cdot \binom{K_\text{w}}{t-1}^{-1}. \tag{51b}$$

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

$$V_i = V_i^{(t)} \cup V_i^{(t-1)} \tag{52}$$

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"

$$W_{d_1}^{(t)}, \ W_{d_2}^{(t)}, \ \ldots, W_{d_{K_\text{w}}}^{(t)}, \tag{53}$$

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

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

$$W_{d_{K_w+1}}^{(t)}, \ W_{d_{K_\text{w}+2}}^{(t)}, \ \ldots, W_{d_K}^{(t)}, \tag{54}$$

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

$$W_{d_1}^{(t-1)}, \ W_{d_2}^{(t-1)}, \ \ldots, W_{d_{K_\text{w}}}^{(t-1)}, \tag{55}$$

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 restricted demand vector $\mathbf{d}_\text{w} = (d_1, \ldots, d_\text{w})$ and to the messages

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

This produces the XOR-messages

$$\left\{ W^{(t-1)}_{\text{XOR},\mathcal{G}_\ell^{(t)}} : \ \ell = 1, \ldots, \binom{K_\text{w}}{t} \right\}, \tag{56}$$

which are of rate $R_\text{sub}^{(t-1)}$ given in (51b).

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

$$W^{(t-1)}_{\text{XOR},\mathcal{G}_\ell^{(t)}} \tag{57a}$$

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

$$\mathbf{W}^{(t)}_{\ell,\text{strong}} := \left(W^{(t)}_{d_{K_\text{w}+1},\mathcal{G}_\ell^{(t)}}, \ \ldots, \ W^{(t)}_{d_K,\mathcal{G}_\ell^{(t)}} \right) \tag{57b}$$

(which consists of K_s messages of rate $R_\text{sub}^{(t)}$) is conveyed to all strong receivers $\mathcal{K}_\text{s}$. For this purpose, we generate a codebook $\mathcal{C}_\ell$ with $\lfloor 2^{nR_\text{sub}^{(t-1)}} \rfloor \times \lfloor 2^{nK_\text{s}R_\text{sub}^{(t)}} \rfloor$ 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

$$x_\ell^{n_{2,\text{per}}}(w_\text{row}, \ w_\text{column}). \tag{58}$$

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

The transmitter sends the codeword

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

over the channel.

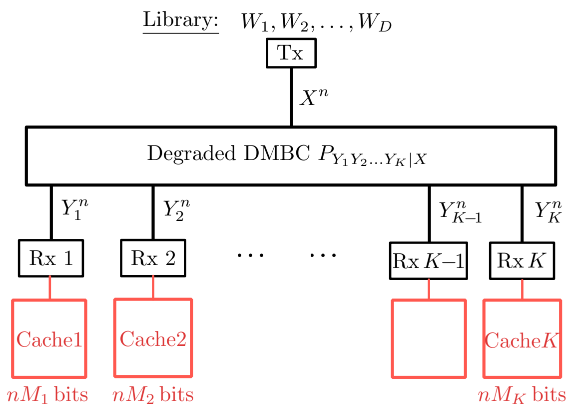

Fig. 4: Degraded K-user BC $P_{Y_1Y_2\cdots Y_K|X}$ where each Receiver $k \in \mathcal{K}$ has cache memory of size nM_k bits.

Each strong receiver $j \in \mathcal{K}_\text{s}$ 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 use only the guess $\hat{W}^{(t)}_{d_j,\mathcal{G}^{(t)}_\ell}$, i.e., the $j-K_\text{w}$-th component of its guess $\hat{\mathbf{W}}^{(t)}_{\ell,\text{strong}}$. At the end of the last period $\binom{K_\text{w}}{t}$, the strong receiver $j \in \mathcal{K}_\text{s}$ produces

$$\hat{W}^{(t)}_j := \left(\hat{W}^{(t)}_{d_j,\mathcal{G}^{(t)}_1}, \ldots, \hat{W}^{(t)}_{d_j,\mathcal{G}^{(t)}_{\binom{K_\text{w}}{t}}}\right). \tag{59}$$

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 $\mathcal{C}_{\ell,\text{row}}\big(\mathbf{W}^{(t)}_{\ell,\text{strong}}\big)$ from $\mathcal{C}_\ell$:

$$\mathcal{C}_{\ell,\text{row}}\left(\mathbf{W}^{(t)}_{\ell,\text{strong}}\right) := \left\{x_\ell^{n_{2,\text{per}}}\left(\mathbf{W}^{(t)}_{\ell,\text{strong}}, W^{(t-1)}_{\text{XOR},\mathcal{G}^{(t)}_\ell}\right)\right\}_{w=1}^{\left\lfloor 2^{nR^{(t-1)}_\text{sub}}\right\rfloor}. \tag{60}$$

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 $\mathcal{C}_{\ell,\text{row}}\big(\mathbf{W}^{(t)}_{\ell,\text{strong}}\big)$.

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

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

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

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

$$W^{(t-1)}_{d_{K_w+1}},\ W^{(t-1)}_{d_{K_\text{w}+2}},\ \ldots, W^{(t-1)}_{d_K}, \tag{61}$$

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

$$\hat{W}^{(t-1)}_j,\ j \in \mathcal{K}_\text{s}. \tag{62}$$

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

$$\hat{W}_k = \left(\hat{W}^{(t-1)}_k, \hat{W}^{(t)}_k\right). \tag{63}$$

V. A Converse for General Degraded BCs

In this section, we present an upper bound on the capacity-memory tradeoff of a more general broadcast network where each Receiver i has a cache of size M_i and the BC is discrete and memoryless with input alphabet $\mathcal{X}$, output alphabets $\mathcal{Y}_1, \ldots, \mathcal{Y}_K$, and channel transition law $P_{Y_1Y_2\cdots Y_K|X}(y_1,\ldots,y_K|x)$. Let us first assume that the BC is *physically degraded*, i.e., the transition law satisfies the Markov chain

$$X \to Y_K \to Y_{K-1} \to \cdots \to Y_1. \tag{64}$$

(The extension to general degraded BCs will trivially follow later.)

The library and the probability of worst-case error $\mathsf{P}_\text{e}^\text{worst}$ 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 (9)–(12) such that $\mathsf{P}_e^{\text{worst}} < \epsilon$. 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

$$j_1 \leq j_2 \leq \ldots \leq j_{|\mathcal{S}|}, \tag{65}$$

define

$$R_{\text{sym},\mathcal{S}} := \max \min \left\{ I(U_1; Y_{j_1}),\ I(U_2; Y_{j_2}|U_1),\ \ldots,\ I\big(U_{|\mathcal{S}|}; Y_{j_{|\mathcal{S}|-1}} \big| U_{|\mathcal{S}|-2}\big),\ I\big(X; Y_{j_{|\mathcal{S}|}} \big| U_{|\mathcal{S}|-1}\big) \right\}, \tag{66}$$

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

$$U_1 \to U_2 \to \cdots \to U_{|\mathcal{S}|-1} \to X \to \big(Y_{j_1}, \ldots, Y_{j_{|\mathcal{S}|}}\big). \tag{67}$$

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 [72].

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

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

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

$$M_{\mathcal{S}} = \sum_{k \in \mathcal{S}} M_k. \tag{68}$$

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

$$R_0 = \left(\frac{K_{\text{w}}}{1 - \delta_{\text{w}}} + \frac{K_{\text{s}}}{1 - \delta_{\text{s}}} \right)^{-1}, \qquad M_0 := 0; \tag{69}$$

and

$$R_{K_{\text{w}}+1} := \frac{1 - \delta_{\text{s}}}{K_{\text{s}}}, \qquad M_{K_{\text{w}}+1} := D\ \frac{1 - \delta_{\text{s}}}{K_{\text{s}}}; \tag{70}$$

and recall the rate-memory pairs $(R_1, M_1), \ldots, (R_{K_{\text{w}}}, M_{K_{\text{w}}})$ in (48):

$$R_t = \frac{(1 - \delta_{\text{w}}) \left(1 + \frac{K_{\text{w}} - t + 1}{t K_{\text{s}}}\ \frac{\delta_{\text{w}} - \delta_{\text{s}}}{1 - \delta_{\text{w}}} \right)}{\frac{K_{\text{w}} - t + 1}{t} \left(1 + \frac{K_{\text{w}} - t}{(t+1) K_{\text{s}}}\ \frac{\delta_{\text{w}} - \delta_{\text{s}}}{1 - \delta_{\text{w}}} \right) + K_{\text{s}} \frac{1 - \delta_w}{1 - \delta_{\text{s}}}}, \quad t \in \mathcal{K}_{\text{w}}, \tag{71a}$$

$$M_t = R_t\ \frac{D}{K_{\text{w}}} \left(t - \left(1 + \frac{K_{\text{w}} - t + 1}{t K_{\text{s}}}\ \frac{\delta_{\text{w}} - \delta_{\text{s}}}{1 - \delta_{\text{w}}} \right)^{-1} \right), \quad t \in \mathcal{K}_{\text{w}}. \tag{71b}$$

Theorem 3: The upper convex hull of the $K_{\text{w}}+2$ rate-memory pairs in (69)–(70) forms a lower bound on the capacity-memory tradeoff:

$$C(M) \geq \text{upper hull}\big\{ (R_t, M_t) \colon\ t = 0, \ldots, K_{\text{w}} + 1 \big\}. \tag{72}$$

Proof outline: The pair $(R_0,\ M_0 = 0)$ corresponds to the case without caches, and achievability follows from (16). Achievability of the pair $(R_{K_{\text{w}}+1},\ M_{K_{\text{w}}+1})$ follows from (19). The pairs $(R_1, M_1), \ldots, (R_{K_{\text{w}}}, M_{K_{\text{w}}})$ are achieved by the joint cache-channel coding scheme in Section IV.

The upper convex hull of $\{(R_t, M_t); t = 0, 1, \ldots, K_{\text{w}} + 1\}$, finally, 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 (71b) and equals

$$M_1 = D \cdot \frac{(\delta_{\text{w}} - \delta_{\text{s}}) K_{\text{s}}^{-1}}{K_{\text{w}} + \frac{K_{\text{w}} - 1}{2} \cdot \frac{K_{\text{w}}(\delta_{\text{w}} - \delta_{\text{s}})}{K_{\text{s}}(1 - \delta_{\text{w}})} + K_{\text{s}} \frac{1 - \delta_{\text{w}}}{1 - \delta_{\text{s}}}}. \tag{73}$$

In this regime, Theorem 3 specializes as:

$$C(M) \geq R_0 + \frac{M}{D} \cdot \frac{K_{\text{w}}(1-\delta_{\text{s}})}{K_{\text{w}}(1-\delta_{\text{s}}) + K_{\text{s}}(1-\delta_{\text{w}})} \cdot \frac{1+K_{\text{w}}}{2} \cdot \gamma_{\text{joint}}, \qquad M \leq M_1, \tag{74}$$

where

$$\gamma_{\text{joint}} := 1 + \frac{2K_{\text{w}}}{1+K_{\text{w}}} \cdot \frac{K_{\text{s}}(1-\delta_{\text{w}})}{K_{\text{w}}(1-\delta_{\text{s}})}. \tag{75}$$

Replacing in lower bound (74) the factor γ_{joint} by 1 recovers 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. Notice that γ_{joint} is unbounded when one increases the number of strong receivers K_{s} while keeping K_{w} and the erasure probabilities δ_{s} and δ_{w} constant. More generally, γ_{joint} is increasing in the ratio $\frac{K_{\text{s}}(1-\delta_{\text{w}})}{K_{\text{w}}(1-\delta_{\text{s}})}$ when $K_{\text{w}} \geq 1$.

B. *General Upper Bound on $C(M)$*

Define for each $k_{\text{w}} \in \{0, \dots, K_{\text{w}}\}$

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

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

$$C(M) \leq \min_{k_{\text{w}} \in \{0,\dots,K_{\text{w}}\}} R_{k_{\text{w}}}(M). \tag{76}$$

Proof: Specialize the upper bound in Theorem 2 to the erasure BC. Since strong receivers do not have cache memories and since the symmetric capacity $R_{\text{sym},\mathcal{S}}$ decreases as the receiver set $\mathcal{S}$ increases, in Theorem 2 it suffices to consider the bounds that correspond to subsets $\mathcal{S} \subseteq \mathcal{K}$ containing all indices $K_{\text{w}}+1, \dots, K$, i.e., all strong receivers. ■

The choice of k_{w} in (76) that leads to the tightest upper bound depends on the cache size M. For small values of M, choosing $k_{\text{w}} = K_{\text{w}}$ leads to the tightest bound, and for increasing 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 Figures 5 and 6.

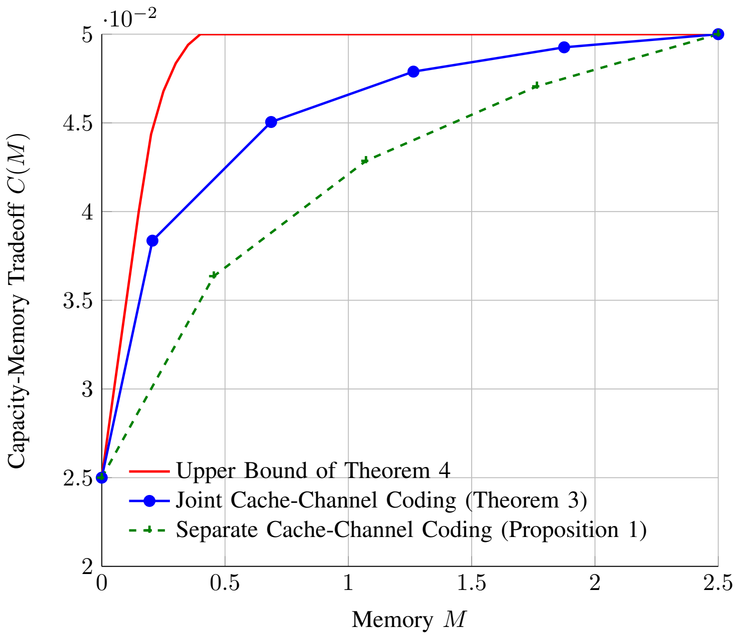

Fig. 5: Bounds on the capacity-memory tradeoff $C(M)$ for $K_{\text{w}} = 4$, $K_{\text{s}} = 16$, $D = 50$, $\delta_{\text{w}} = 0.8$, $\delta_{\text{s}} = 0.2$.

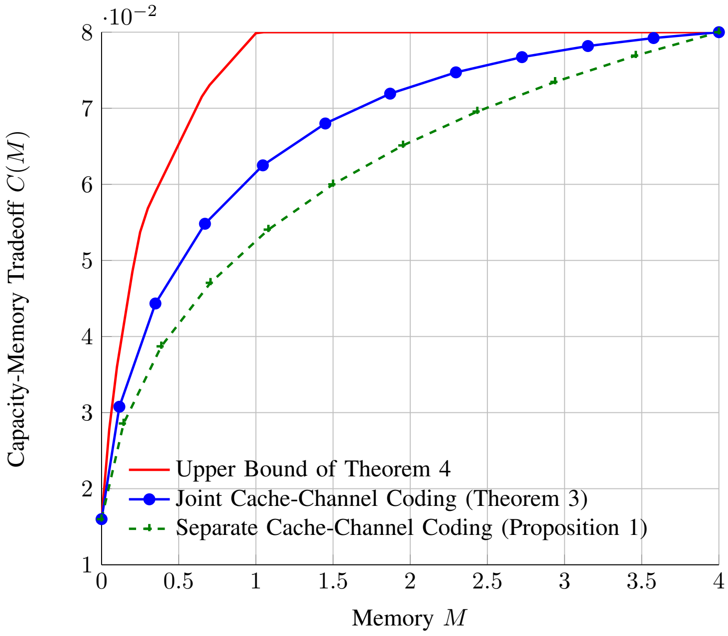

Fig. 6: Bounds on capacity-memory tradeoff $C(M)$ for $K_{\text{w}} = K_{\text{s}} = 10$, $D = 50$, $\delta_{\text{w}} = 0.8$, $\delta_{\text{s}} = 0.2$.

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

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

$$\Gamma_1 := \frac{(1-\delta_{\text{s}})}{K_{\text{s}}} \cdot \frac{(\delta_w - \delta_{\text{s}})}{(K_{\text{s}}(1-\delta_w) + (1-\delta_{\text{s}}))}, \tag{77}$$

$$\Gamma_2 := \frac{(1-\delta_{\text{s}})}{K_{\text{s}}} \cdot \frac{(1-\delta_{\text{s}})}{(K_{\text{s}}(1-\delta_w) + (1-\delta_{\text{s}}))}, \tag{78}$$

$$\Gamma_3 := \frac{(1-\delta_{\text{s}})}{K_{\text{s}}}. \tag{79}$$

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

Corollary 4.1: If $K_{\text{w}} = 1$ the capacity-memory tradeoff is lower bounded by:

$$C(M) \geq \begin{cases} \frac{(1-\delta_{\text{w}})(1-\delta_{\text{s}})}{K_{\text{s}}(1-\delta_{\text{w}})+(1-\delta_{\text{s}})} + \frac{M}{D}, & \text{if } \frac{M}{D} \in [0, \Gamma_1] \\ \frac{(1-\delta_{\text{s}})}{1+K_{\text{s}}} + \frac{M}{(1+K_{\text{s}})D}, & \text{if } \frac{M}{D} \in (\Gamma_1, \Gamma_3], \end{cases} \tag{80}$$

and upper bounded by:

$$C(M) \leq \begin{cases} \frac{(1-\delta_{\text{w}})(1-\delta_{\text{s}})}{K_{\text{s}}(1-\delta_{\text{w}})+(1-\delta_{\text{s}})} + \frac{M}{D}, & \text{if } \frac{M}{D} \in [0, \Gamma_2] \\ \frac{(1-\delta_{\text{s}})}{K_{\text{s}}} & \text{if } \frac{M}{D} \in (\Gamma_2, \Gamma_3]. \end{cases} \tag{81}$$

Figure 7 shows these two bounds and the bound in Proposition 1 for $K_{\text{w}} = 1$, $K_{\text{s}} = 10$, $D = 22$, $D = 10$, $\delta_{\text{w}} = 0.8$, $\delta_{\text{s}} = 0.2$.

We identify two regimes. In the first regime $0 \leq \frac{M}{D} \leq \Gamma_1$, the joint cache-channel coding scheme allows to increase the rate R by $\frac{M}{D}$. This is the same performance as if all the K_{s} strong receivers could directly access Receiver 1's cache contents. This is the highest possible caching gain, and here the upper and lower bounds match.

In the second regime $\Gamma_1 < \frac{M}{D} \leq \Gamma_3$, the joint cache-channel coding scheme still profits from an increasing cache size, but the gain is less significant: the rate only increases as $\frac{1-\delta_{\text{s}}}{K_{\text{s}}} \cdot \frac{M}{D}$.

D. Special Case $K_{\text{w}} = K_{\text{s}} = 1$ and $D = 2$

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

$$\tilde{\Gamma}_1 := \frac{(1-\delta_{\text{s}})^2 - (1-\delta_{\text{w}})(\delta_{\text{w}} - \delta_{\text{s}})}{(1-\delta_{\text{w}}) + (1-\delta_{\text{s}})}, \tag{82}$$

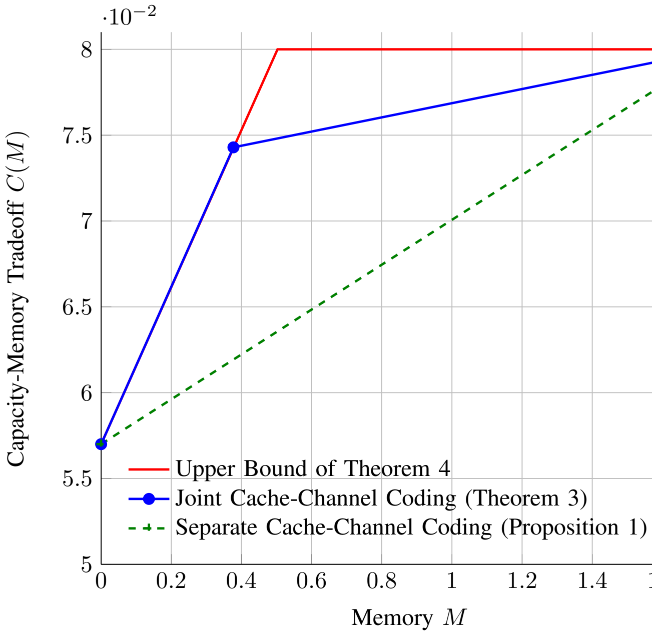

Fig. 7: Bounds on the capacity-memory tradeoff for $K_{\text{w}} = 1$, $K_{\text{s}} = 10$, $D = 22$, $\delta_{\text{w}} = 0.8$, $\delta_{\text{s}} = 0.2$.

$$\tilde{\Gamma}_2 := \frac{1}{2}\left((1-\delta_{\text{s}}) + (\delta_{\text{w}} - \delta_{\text{s}})\right). \tag{83}$$

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

Theorem 5: If $K_{\text{w}} = K_{\text{s}} = 1$ and $D = 2$, the capacity-memory tradeoff is upper bounded as:

$$C(M) \leq \begin{cases} \frac{(1-\delta_{\text{w}})(1-\delta_{\text{s}})}{(1-\delta_{\text{w}})+(1-\delta_{\text{s}})} + \frac{M}{2}, & \text{if } \frac{M}{2} \in [0, \tilde{\Gamma}_1] \\ \frac{1}{3}(2-\delta_{\text{s}}-\delta_{\text{w}}) + \frac{M}{3}, & \text{if } \frac{M}{2} \in (\tilde{\Gamma}_1, \tilde{\Gamma}_2] \\ 1-\delta_{\text{s}} & \text{if } \frac{M}{2} \in (\tilde{\Gamma}_2, \Gamma_3]. \end{cases} \tag{84}$$

and lower bounded as:

$$C(M) \geq \begin{cases} \dfrac{(1-\delta_{\text{w}})(1-\delta_{\text{s}})}{(1-\delta_{\text{w}})+(1-\delta_{\text{s}})} + \dfrac{M}{2}, & \dfrac{M}{2} \in [0, \Gamma_1] \\ \dfrac{(1-\delta_s)}{3(1-\delta_{\text{s}})-(1-\delta_{\text{w}})}\left((1-\delta_{\text{s}})+M\right), & \dfrac{M}{2} \in (\Gamma_1, \tilde{\Gamma}_2] \\ 1-\delta_{\text{s}} & \dfrac{M}{2} \in (\tilde{\Gamma}_2, \Gamma_3]. \end{cases} \tag{85}$$

Proof: Lower bound (85) coincides with the upper convex hull of the three rate-memory pairs: $(R_0,\ M_0)$ in (69); $(R_1,\ M_1)$ in (48); and $((1-\delta_{\text{s}}),\ 2\tilde{\Gamma}_3)$. Achievability of the former two pairs follows from Theorem 3. Achievability of the last pair follows from the joint cache-channel coding scheme in Appendix F. The upper bound is proved in Appendix G. ■

Figure 8 shows the bounds of Theorem 5 for $\delta_{\text{w}} = 0.8$ and $\delta_{\text{s}} = 0.2$. The upper and lower bounds of Theorem 5 coincide in general for $0 \leq M \leq \Gamma_1$ and for $M \geq \tilde{\Gamma}_2$. The theorem allows to conclude that the minimum cache size M for which communication is possible at the maximum rate $(1-\delta_{\text{s}})$ is $M = 2\tilde{\Gamma}_2$. When $\delta_{\text{w}} = \delta_{\text{s}}$, then the upper and lower bounds in Theorem 5 coincide for all values of M.

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., [4]). 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 $F \to \infty$. The minimum delivery rate ρ for given cache size m for which (ρ, m) is achievable, is called the *delivery rate-memory tradeoff* and is called $\rho^\star(m)$.

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:

$(R,\ M)$ achievable in original setup

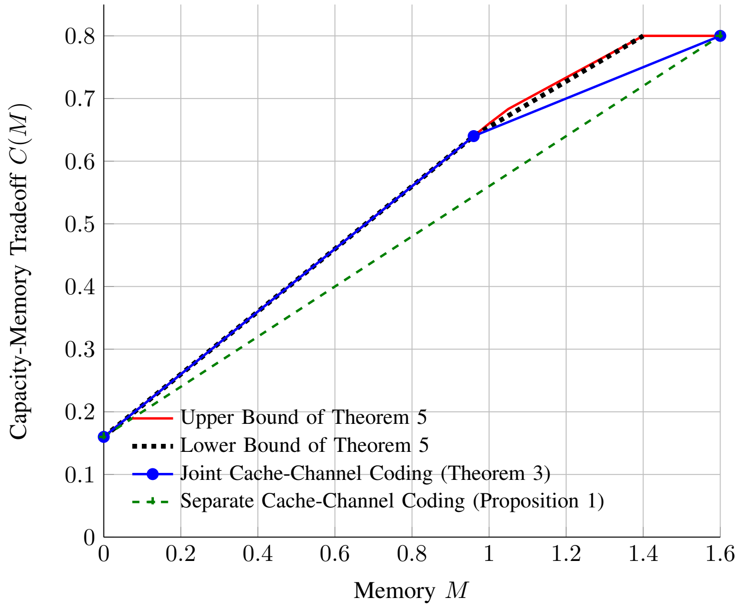

Fig. 8: Bounds on the capacity-memory tradeoff for $K_\mathrm{w} = 1$, $K_\mathrm{s} = 1$, $D = 2$, $\delta_\mathrm{w} = 0.8$, $\delta_\mathrm{s} = 0.2$.

$$\iff$$

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

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 normalized cache memory m. The separate-cache channel coding scheme in Section III achieves for each $\tilde{t} \in \{0, 1, \ldots, K_\mathrm{w}\}$ the delivery-rate memory pair

$$\rho_{\tilde{t},\mathrm{sep}} := \frac{K_\mathrm{w} - \tilde{t}}{(\tilde{t}+1)(1-\delta_\mathrm{w})} + \frac{K_\mathrm{s}}{(1-\delta_\mathrm{s})}, \tag{86a}$$

$$m_{\tilde{t},\mathrm{sep}} := D\frac{\tilde{t}}{K_\mathrm{w}}. \tag{86b}$$

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

$$\rho_t := \nu_t \frac{K_\mathrm{w} - t}{(t+1)(1-\delta_\mathrm{w})} + (1-\nu_t)\frac{K_\mathrm{w} - t + 1}{t(1-\delta_\mathrm{w})} + (1-\nu_t)\frac{K_\mathrm{s}}{1-\delta_\mathrm{s}}, \tag{87a}$$

$$m_t := \nu_t D \frac{t}{K_\mathrm{w}} + (1-\nu_t) D \frac{t-1}{K_\mathrm{w}}, \tag{87b}$$

where

$$\nu_t := \frac{(K_\mathrm{w} - t + 1)(\delta_\mathrm{w} - \delta_\mathrm{s}))}{(K_\mathrm{w} - t + 1)(\delta_\mathrm{w} - \delta_\mathrm{s}) + tK_\mathrm{s}(1-\delta_\mathrm{w})}. \tag{88}$$

The lower convex hull of the delivery rate-memory pairs in (86) and (87) 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 $\rho^\star(m)$:

$$\rho^\star(m) \geq \max_{k_\mathrm{w} \in \{0,1,\ldots,K_\mathrm{w}\}} \left[\left(\frac{k_\mathrm{w}}{1-\delta_\mathrm{w}} + \frac{K_\mathrm{s}}{1-\delta_\mathrm{s}} \right) \left(1 - \frac{k_\mathrm{w} m}{D} \right) \right]. \tag{89}$$

Figures 9 and 10 present upper and lower bounds on $\rho^\star(m)$ when $K_\mathrm{w} = K_\mathrm{s} = 10$, $D = 50$, $\delta_\mathrm{w} = 0.8$, $\delta_\mathrm{s} = 0.2$ and when $K_\mathrm{w} = 10$, $K_\mathrm{s} = 1000$, $D = 5000$, $\delta_\mathrm{w} = 0.8$, $\delta_\mathrm{s} = 0.2$. The lower bound is given by (89) and the upper bounds depict the convex hulls of rate-memory pairs $\{(\rho_{\tilde{t},\mathrm{sep}}, m_{\tilde{t},\mathrm{sep}})\}_{\tilde{t}=0}^{K_\mathrm{w}}$ and rate-memory pairs $\big\{(\rho_{0,\mathrm{sep}}, m_{0,\mathrm{sep}}),\ \{(\rho_t, m_t)\}_{t=1}^{K_\mathrm{w}},\ (\rho_{K_\mathrm{w},\mathrm{sep}}, m_{K_\mathrm{w},\mathrm{sep}})\big\}$.

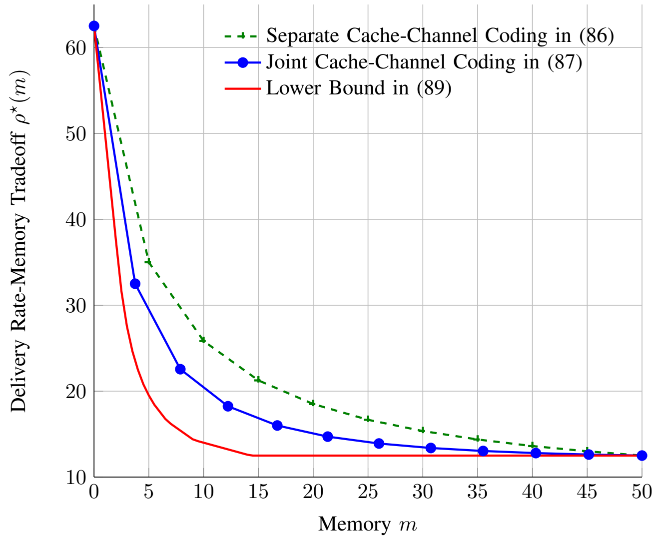

Fig. 9: Bounds on $\rho^\star(m)$ for $K_\mathrm{w} = K_\mathrm{s} = 10$, $D = 50$, $\delta_\mathrm{w} = 0.8$, $\delta_\mathrm{s} = 0.2$.

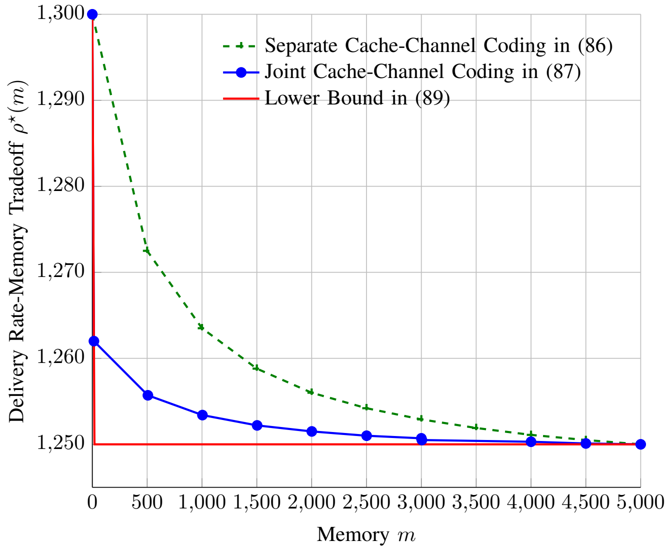

Fig. 10: Bounds on $\rho^\star(m)$ for $K_\mathrm{w} = 10$ and $K_\mathrm{s} = 1000$, $D = 5000$, $\delta_\mathrm{w} = 0.8$, $\delta_\mathrm{s} = 0.2$.

VIII. Summary and Concluding Remarks

In this paper we consider an erasure broadcast network with a set of weak receivers with equal cache size M and a set of strong receivers with no cache memories. Upper and lower bounds on the capacity-memory tradeoff are derived. These bounds are generally close, and they match when there is only a single weak receiver and its cache size is small. In practice 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. 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. In this regime the

To facilitate comparison with previous works that mostly focused on the delivery-rate memory tradeoff, we express our main results also in terms of this related quantity. 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 [4].

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

The considered scenario with two sets of receivers and cache memories only at weak receivers is motivated by previous works showing that in networks where some receivers have stronger channel conditions than others—e.g., because they are closer to a fiber optical access point or a helper basestation—this asymmetric cache configuration allows for significantly improved performance compared to a more traditional uniform cache configuration. The present work illustrates that by applying joint cache-channel coding in such an asymmetric cache-assignment allows obtaining further caching gains, which can even be unbounded in the network parameters. The benefit of joint cache-channel coding is that it can provide global caching gains also to the strong receivers which have no cache memories. In the considered separate cache-channel coding scheme this is not the case.

The described asymmetric cache configuration arises also as part of a more complex system model in which every receiver is equipped with a cache. Suppose, for example, that the stronger receivers want to decode additional data that will never be demanded by the weak receivers. This additional data might represent, for example, a higher resolution of a video. A practical solution in this case is to separate transmission of files from the two libraries [47]–[49]: A first transmission sends the files that are of interest to all receivers, and a second transmission sends only files from the additional library to the strong receivers. The question is now how to divide the cache memory between the two transmissions. Based on the results we obtain in this paper, we propose to assign all the cache memory at the strong receivers to the second transmission, because through a careful design of the first transmission scheme, the strong receivers can already benefit from the weak receivers' caches without accessing their own cache memories.

ACKNOWLEDGMENTS

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

APPENDIX A
ANALYSIS OF JOINT CACHE-CHANNEL CODING SCHEME IN SECTION IV

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

$$\begin{aligned} M &= R^{(t)} \cdot D\frac{t}{K_\mathrm{w}} + R^{(t-1)} \cdot D\frac{t-1}{K_\mathrm{w}} \\ &= R \cdot \frac{D}{K_\mathrm{w}} \left(t - \left(1 + \frac{K_\mathrm{w} - t + 1}{tK_\mathrm{s}} \cdot \frac{\delta_\mathrm{w} - \delta_\mathrm{s}}{1 - \delta_\mathrm{w}} \right)^{-1} \right). \end{aligned} \tag{90}$$

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 vanishes as $n \to \infty$, whenever

$$\frac{R^{(t)} \cdot \frac{K_\mathrm{w} - t}{t+1}}{1 - \delta_\mathrm{w}} < \beta_1. \tag{91}$$

3) Delivery Subphase 2: Consider Period ℓ with the transmission of messages in (57). The probability that the strong receivers make a decoding error vanishes as $n \to \infty$, whenever

$$\frac{R_\text{sub}^{(t-1)} + K_\mathrm{s} R_\text{sub}^{(t)}}{1 - \delta_\mathrm{s}} < \frac{\beta_2}{\binom{K_\mathrm{w}}{t}}. \tag{92}$$

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 thus vanishes as $n \to \infty$, whenever

$$\frac{R_\text{sub}^{(t-1)}}{(1 - \delta_\mathrm{w})} < \frac{\beta_2}{\binom{K_\mathrm{w}}{t}}. \tag{93}$$

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

By our choice of the rates $R^{(t-1)}$ and $R^{(t)}$ in (50) the two Constraints (92) and (93) coincide. We ignore (92) in the following.

4) Delivery Subphase 3: The probability that the strong receivers err in their decoding vanishes as $n \to \infty$, whenever

$$\frac{K_\mathrm{s} R^{(t-1)}}{(1 - \delta_\mathrm{s})} < \beta_3. \tag{94}$$

5) Overall Scheme: Combining (91), (93), and (94) and using the definitions of $R_{\text{sub}}^{(t-1)}$ and $R_{\text{sub}}^{(t-1)}$ in (51), we conclude that the probability of decoding error vanishes as $n \to \infty$, if

$$\frac{R^{(t)} \frac{K_\text{w}-t}{t+1}}{1-\delta_\text{w}} + \frac{R^{(t-1)} \cdot \frac{K_\text{w}-t+1}{t}}{1-\delta_\text{w}} + \frac{K_\text{s} R^{(t-1)}}{1-\delta_\text{s}} < 1. \tag{95}$$

Using the definitions of $R^{(t-1)}$ and $R^{(t)}$ in (50), one obtains that the probability of decoding error vanishes, 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 (90), this proves achievability of the rate-memory pair $(R_t,\ M_t)$ in (48).

APPENDIX B
PROOF OF THEOREM 2

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

$$C(M_1, \ldots, M_K) \leq R_{\text{sym},\mathcal{K}} + \frac{1}{D}\sum_{k=1}^{K} M_k. \tag{96}$$

The bounds corresponding to other subsets $\mathcal{S}$ can be proved in a similar way.

We start the proof of (96). 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 $\mathsf{P}_\text{e}{}^{(n)}(\mathbf{d})$ 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 contents for the chosen placement functions.

Lemma 6: For any $\epsilon > 0$, any demand vector $\mathbf{d} = (d_1, \ldots, d_K)$ with all different entries, 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

$$U_{1,\mathbf{d}} - U_{2,\mathbf{d}} - \cdots - U_{K-1,\mathbf{d}} - X_\mathbf{d} - Y_{K,\mathbf{d}} - Y_{K-1,\mathbf{d}} \cdots - Y_{1,\mathbf{d}} \tag{97a}$$

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

$$(Y_{1,\mathbf{d}}, Y_{2,\mathbf{d}}, \ldots, Y_{K,\mathbf{d}}) \sim P_{Y_1 \cdots Y_K|X}(\cdots|x), \tag{97b}$$

and such that the following K inequalities hold:

$$R - \epsilon \leq \frac{1}{n} I\big(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}\big) + I\big(U_{1,\mathbf{d}}; Y_{1,\mathbf{d}}\big), \tag{98a}$$

$$\begin{aligned} R - \epsilon &\leq \frac{1}{n} I\big(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{k-1}}\big) \\ &\quad + I\big(U_{k,\mathbf{d}}; Y_{k,\mathbf{d}} | U_{k-1,\mathbf{d}}\big), \qquad \forall k \in \{2, \ldots, K-1\}, \\ R - \epsilon &\leq \frac{1}{n} I\big(W_{d_K}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{K-1}}\big) \\ &\quad + I\big(X_\mathbf{d}; Y_{K,\mathbf{d}} | U_{K-1,\mathbf{d}}\big). \end{aligned} \tag{98b}$$

Proof: The proof is similar to the converse proof of the capacity of degraded BCs without caching [73, 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 (98) 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 $\mathcal{Q}$ and independent of all previously defined random variables. Define: $U_1 := (U_{1,Q}, Q)$; $U_k := U_{k,Q}$, for $k \in \{2, \ldots, K-1\}$; $X_k := X_Q$; and $Y_k := Y_{k,Q}$ for $k \in \mathcal{K}$. Notice that they form the Markov chain

$$U_1 \to U_2 \to \cdots \to U_{K-1} \to X \to (Y_1, \ldots, Y_K), \tag{99}$$

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

$$(Y_1, Y_2, \ldots, Y_K) \sim P_{Y_1 \cdots Y_K|X}(\cdots|x). \tag{100}$$

Averaging inequalities (98) over the demand vectors in $\mathcal{Q}$ and using standard arguments to take care of the time-sharing random variable Q, we obtain:

$$R - \epsilon \leq \alpha_1 + I\big(U_1; Y_1\big), \tag{101a}$$
$$R - \epsilon \leq \alpha_k + I\big(U_k; Y_k|U_{k-1}\big), \quad \forall k \in \{2, \ldots, K-1\}, \tag{101b}$$
$$R - \epsilon \leq \alpha_k + I\big(X; Y_K|U_{K-1}\big), \tag{101c}$$

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

$$\alpha_1 := \frac{1}{\binom{D}{K} K!} \sum_{\mathbf{d} \in \mathcal{Q}} \frac{1}{n} I(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}), \tag{102a}$$
$$\alpha_k := \frac{1}{\binom{D}{K} K!} \sum_{\mathbf{d} \in \mathcal{Q}} \frac{1}{n} I(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{k-1}}), \quad k \in \{2, \ldots, K\}. \tag{102b}$$

Lemma 7: Parameters $\alpha_1, \ldots, \alpha_K$, defined in (102), satisfy the following constraints:

$$\alpha_k \geq 0, \qquad k \in \mathcal{K}, \tag{103a}$$
$$\alpha_{k'} \leq \alpha_k, \qquad k, k' \in \mathcal{K},\ k' \leq k, \tag{103b}$$
$$\sum_{k \in \mathcal{K}} \alpha_k \leq \frac{K}{D} \sum_{k \in \mathcal{K}} M_k. \tag{103c}$$

Proof: See Appendix D. ■

Taking $\epsilon \to 0$, by (101) and (102) and by Lemma 7, we conclude that the capacity-memory tradeoff $C(M_1, \ldots, M_K)$ is upper bounded by the following K inequalities:

$$C(M_1, \ldots, M_K) \leq \alpha_1 + I\big(U_1; Y_1\big), \tag{104a}$$
$$C(M_1, \ldots, M_K) \leq \alpha_k + I\big(U_k; Y_k \big| U_{k-1}\big), \qquad \forall k \in \{2, \ldots, K-1\}, \tag{104b}$$
$$C(M_1, \ldots, M_K) \leq \alpha_K + I\big(X; Y_K \big| U_{K-1}\big), \tag{104c}$$

for some $\alpha_1, \ldots, \alpha_K$ satisfying (103) and some $U_1, \ldots, U_{K-1}, X, Y_1, \ldots, Y_K$ satisfying (99) and (100).

Lemma 8: Replacing each and every real number $\alpha_1, \ldots, \alpha_K$ in (104) 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,

$$C(M_1, \ldots, M_K) - \frac{\sum_{k \in \mathcal{K}} M_k}{D} \leq I\big(U_1; Y_1\big), \tag{105a}$$
$$C(M_1, \ldots, M_K) - \frac{\sum_{k \in \mathcal{K}} M_k}{D} \leq I\big(U_k; Y_k \big| U_{k-1}\big), \quad \forall k \in \{2, \ldots, K-1\}, \tag{105b}$$
$$C(M_1, \ldots, M_K) - \frac{\sum_{k \in \mathcal{K}} M_k}{D} \leq I\big(X; Y_K \big| U_{K-1}\big), \tag{105c}$$

for some $U_1, \ldots, U_K, X, Y_1, \ldots, Y_K$ satisfying (99) and (100).

All K constraints in (105) have the same LHS, and their RHSs coincide with the rate-constraints determining 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 (105) with the desired bound in (96), 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 be come clear in the following. Also, let

$$V_k^{(n)} = g_k^{(n)}(W_1, \ldots, W_D), \qquad k \in \mathcal{K}, \tag{106}$$
$$X_{\mathbf{d}}^n = f_{\mathbf{d}}^{(n)}(W_1, \ldots, W_D) \tag{107}$$

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

By Fano's inequality, by 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 that for all sufficiently large n the following K inequalities hold:

$$R - \epsilon \leq \frac{1}{n} I\big(W_{d_1}; Y^n_{1,\mathbf{d}}, V_1^{(n)}, \ldots, V_K^{(n)}\big)$$
$$= \frac{1}{n} I\big(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}\big) + \frac{1}{n} I\big(W_{d_1}; Y^n_{1,\mathbf{d}} \big| V_1^{(n)}, \ldots, V_K^{(n)}\big), \tag{108a}$$

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

$$R - \epsilon_n \leq \frac{1}{n} I\big(W_{d_k}; Y^n_{k,\mathbf{d}}, V_1^{(n)}, \ldots, V_K^{(n)} \big| W_{d_1}, \ldots, W_{d_{k-1}}\big)$$
$$= \frac{1}{n} I\big(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} \big| W_{d_1}, \ldots, W_{d_{k-1}}\big)$$
$$+ \frac{1}{n} I\big(W_{d_k}; Y^n_{k,\mathbf{d}} \big| V_1^{(n)}, \ldots, V_K^{(n)}, W_{d_1}, \ldots, W_{d_{k-1}}\big) \tag{108b}$$

We further develop the second summands in (108a) and (108b). For the second summand in (108a) we obtain

$$\frac{1}{n} I\big(W_{d_1}; Y^n_{1,\mathbf{d}} \big| V_1^{(n)}, \ldots, V_K^{(n)}\big)$$
$$= \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_1}; Y_{1,\mathbf{d},t} \big| V_1^{(n)}, \ldots, V_K^{(n)}, Y^{t-1}_{1,\mathbf{d}}\big)$$
$$\leq \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_1}, V_1^{(n)}, \ldots, V_K^{(n)}, Y^{t-1}_{1,\mathbf{d}}; Y_{1,\mathbf{d},t}\big)$$
$$= I\big(U_{1,\mathbf{d},T}; Y_{1,\mathbf{d},T} \big| T\big)$$
$$\leq I(U_{1,\mathbf{d}}; Y_{1,\mathbf{d}}), \tag{109}$$

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

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

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

$$U_{k,\mathbf{d},T} := (V_1^{(n)} \ldots, V_K^{(n)}, W_{d_1}, W_{d_2}, \ldots, W_{d_k}, Y^{T-1}_{1,\mathbf{d}}, \ldots, Y^{T-1}_{k,\mathbf{d}}),$$
$$U_{k,\mathbf{d}} := (U_{k,\mathbf{d},T}, T),$$
$$Y_{k,\mathbf{d}} := Y_{k,\mathbf{d},T},$$

in order to expand the second summand in (108b) as follows:

$$\frac{1}{n} I\big(W_{d_k}; Y^n_{k,\mathbf{d}} \big| V_1^{(n)}, \ldots, V_K^{(n)}, W_{d_1}, \ldots, W_{d_{k-1}}\big)$$
$$= \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_k}; Y_{k,\mathbf{d},t} \big| V_1^{(n)}, \ldots, V_K^{(n)}, W_{d_1}, \ldots, W_{d_{k-1}}, Y^{t-1}_{k,\mathbf{d}}\big)$$
$$= \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_k}; Y_{k,\mathbf{d},t} \big| V_1^{(n)}, \ldots, V_K^{(n)}, W_{d_1}, \ldots, W_{d_{k-1}}, Y^{t-1}_{1,\mathbf{d}}, \ldots, Y^{t-1}_{k-1,\mathbf{d}}, Y^{t-1}_{k,\mathbf{d}}\big)$$
$$\leq \frac{1}{n} \sum_{t=1}^{n} I\big(W_{d_k}, Y^{t-1}_{k,\mathbf{d}} \; ; \; Y_{k,\mathbf{d},t} \big| V_1^{(n)}, \ldots, V_K^{(n)}, W_{d_1}, \ldots, W_{d_{k-1}}, Y^{t-1}_{1,\mathbf{d}}, \ldots, Y^{t-1}_{k-1,\mathbf{d}}\big)$$
$$= I\big(U_{k,\mathbf{d},T}; Y_{k,\mathbf{d},T} \big| U_{k-1,\mathbf{d},T}, T\big)$$
$$= I\big(U_{k,\mathbf{d}}; Y_{k,\mathbf{d}} \big| U_{k-1,\mathbf{d}}\big) \tag{110}$$

where the second equality follows from the degradedness of the outputs, see (64).

Similarly, for $k=K$:

$$\frac{1}{n}I\big(W_{d_K};Y^n_{K,\mathbf{d}}|V_1^{(n)},\ldots,V_K^{(n)},W_{d_1},\ldots,W_{d_{K-1}}\big)$$
$$\leq I(X;Y_{k,\mathbf{d}}|U_{K-1,\mathbf{d}}), \tag{111}$$

where

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

Since the defined random variables satisfy (97), Inequalities (108)–(111) conclude the proof.

Appendix D
Proof of Lemma 7

Constraint (103a) follows by the nonnegativity of mutual information. To prove Constraint (103b), 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,

$$\overrightarrow{d}_k^{(\ell)}=d_{(k-\ell)\bmod K} \tag{112}$$

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

$$\xi \mod K=\xi-bK \quad \text{for some positive integer } b. \tag{113}$$

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

$$I(W_{d_1};V_1^{(n)},\ldots,V_K^{(n)})\overset{(a)}{=}I(W_{\overrightarrow{d}_{k'}^{(k'-1)}};V_1^{(n)},\ldots,V_K^{(n)})$$
$$\overset{(b)}{\leq}I(W_{\overrightarrow{d}_{k'}^{(k'-1)}};V_1^{(n)},\ldots,V_K^{(n)}|W_{\overrightarrow{d}_1^{(k'-1)}},\ldots,W_{\overrightarrow{d}_{k'-1}^{(k'-1)}})$$
$$\overset{(a)}{=}I(W_{\overrightarrow{d}_k^{(k-1)}};V_1^{(n)},\ldots,V_K^{(n)}|W_{\overrightarrow{d}_{k-k'+1}^{(k-1)}},\ldots,W_{\overrightarrow{d}_{k-1}^{(k-1)}})$$
$$\overset{(b)}{\leq}I(W_{\overrightarrow{d}_k^{(k-1)}};V_1^{(n)},\ldots,V_K^{(n)}|W_{\overrightarrow{d}_1^{(k-1)}},\ldots,W_{\overrightarrow{d}_{k-1}^{(k-1)}}), \tag{114}$$

where (a) follows by (112), and (b) follows by the independence of the messages and because $k-k'+1\geq 2$.

Fix a demand vector $\mathbf{d}\in\mathcal{Q}$ and sum up the above inequality (114) over all K cyclic shifts $\mathbf{d}^{(0)},\mathbf{d}^{(1)},\ldots,\mathbf{d}^{(K-1)}$ of $\mathbf{d}$ to obtain:

$$\sum_{\ell=0}^{K-1}I(W_{\overrightarrow{d}_1^{(\ell)}};V_1^{(n)},\ldots,V_K^{(n)})$$
$$\leq\sum_{\ell=0}^{K-1}I(W_{\overrightarrow{d}_{k'}^{(\ell)}};V_1^{(n)},\ldots,V_K^{(n)}|W_{\overrightarrow{d}_1^{(\ell)}},\ldots,W_{\overrightarrow{d}_{k'-1}^{(\ell)}})$$
$$\leq\sum_{\ell=0}^{K-1}I(W_{\overrightarrow{d}_k^{(\ell)}};V_1^{(n)},\ldots,V_K^{(n)}|W_{\overrightarrow{d}_1^{(\ell)}},\ldots,W_{\overrightarrow{d}_{k-1}^{(\ell)}}). \tag{115}$$

Since the set $\mathcal{Q}$ can be partitioned into subsets of demand vectors that are cyclic shifts of each others and all cyclic shifts of a demand vector in $\mathcal{Q}$ are also in $\mathcal{Q}$, we conclude from (115):

$$\sum_{\mathbf{d}\in\mathcal{Q}}I(W_{d_1};V_1^{(n)},\ldots,V_K^{(n)})$$
$$\leq\sum_{\mathbf{d}\in\mathcal{Q}}I(W_{d_{k'}};V_1^{(n)},\ldots,V_K^{(n)}|W_{d_1},\ldots,W_{d_{k'-1}})$$
$$\leq\sum_{\mathbf{d}\in\mathcal{Q}}I(W_{d_k};V_1^{(n)},\ldots,V_K^{(n)}|W_{d_1},\ldots,W_{d_{k-1}}). \tag{116}$$

This proves (103b).

We proceed to prove Constraint (103c). For each $\mathbf{d}\in\mathcal{Q}$:

$$I(W_{d_1};V_1^{(n)},\ldots,V_K^{(n)})$$

$$+\sum_{k=2}^{K} I(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, W_{d_2}, \ldots, W_{d_{k-1}})$$
$$= I(W_{d_1}, W_{d_2}, \ldots, W_{d_K}; V_1^{(n)}, \ldots, V_K^{(n)}). \tag{117}$$

Thus,

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

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 $\ell \in \{1, \ldots, L\}$:

$$\frac{1}{\binom{L}{\ell}} \sum_{\mathcal{S}\subseteq\{1,\ldots,L\}:|\mathcal{S}|=\ell} \frac{H(A_{\mathcal{S}})}{\ell} \geq \frac{1}{L} H(A_1, \ldots, A_L). \tag{118}$$

Proof: See [74, 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 (99) and real numbers $\alpha_1, \ldots, \alpha_K$ satisfying (103). 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

$$\bar{U}_1 \to \bar{U}_2 \to \ldots \to \bar{U}_{K-1} \to X \to (Y_1, \ldots, Y_K), \tag{119}$$

and real numbers $\bar{\alpha}_1, \ldots, \bar{\alpha}_K$ satisfying (103) so that the upper bound on $C(M_1, \ldots, M_K)$ in (104) is relaxed if we replace

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

by

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

This proves that the upper bound on $C(M_1, \ldots, M_K)$ in (104) remains unchanged if we replace all numbers $\alpha_1, \ldots, \alpha_K$ by the same number α. By (103c) this number $\alpha \leq \frac{1}{D}\sum_{k\in\mathcal{K}} M_k$, and by the monotonicity of the RHSs of (104) 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 (103b), the strict inequality

$$\alpha_{\tilde{k}} < \alpha_{\tilde{k}+1} \tag{120}$$

must hold. Choose

$$\bar{\alpha}_k = \alpha_k, \qquad k \in \mathcal{K},\ k \notin \{\tilde{k}, \tilde{k}+1\}, \tag{121}$$

$$\bar{\alpha}_{\tilde{k}} = \bar{\alpha}_{\tilde{k}+1} = \frac{1}{2}(\alpha_{\tilde{k}} + \alpha_{\tilde{k}+1}), \tag{122}$$

$$\bar{U}_k = U_k, \qquad k \in \{1,\ldots,K-1\},\ k \neq \tilde{k}. \tag{123}$$

For convenience, define

$$\bar{U}_K := U_K := X. \tag{124}$$

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

$$I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}) \leq I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}), \tag{125a}$$

or

$$I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}) > I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}), \tag{125b}$$

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

If (125a) holds, choose

$$\bar{U}_{\tilde{k}} = U_{\tilde{k}}. \tag{126}$$

If (125b) holds, let $E \in \{0,1\}$ be a Bernoulli-β random variable independent of everything else, where

$$\beta := \frac{1}{2} + \frac{1}{2}\frac{I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}})}{I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1})}, \tag{127}$$

and choose

$$\bar{U}_{\tilde{k}} = \begin{cases} (U_{\tilde{k}}, E), & \text{if } E = 1 \\ (U_{\tilde{k}-1}, E), & \text{if } E = 0. \end{cases} \tag{128}$$

The proposed choice satisfies the Markov chain (119).

Trivially, for $k \notin \{\tilde{k}, \tilde{k}+1\}$, Constraint (104) 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 (125a) holds, then the proposed replacement relaxes Constraint (104) 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 (125a) holds, the upper bound on $C(M_1, \ldots, M_K)$ in (104) 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 (125b) holds. Then, by (127) and (128), because $\bar{U}_{\tilde{k}-1} = U_{\tilde{k}-1}$, and because E is independent of the pair $(Y_{\tilde{k}}, \bar{U}_{\tilde{k}-1})$:

$$\begin{aligned} & I(\bar{U}_{\tilde{k}}; Y_{\tilde{k}} | \bar{U}_{\tilde{k}-1}) \\ & = I(\bar{U}_{\tilde{k}}; Y_{\tilde{k}} | \bar{U}_{\tilde{k}-1}, E) \\ & = \beta \cdot I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}, E=1) \\ & = \frac{1}{2}\left(I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}) + I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}) \right). \end{aligned} \tag{129}$$

By (122) and (129), 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 (122) and because

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

$$\stackrel{(e)}{=} \frac{1}{2} I(U_{\tilde{k}+1}; Y_{\tilde{k}+1} | U_{\tilde{k}}) + \frac{1}{2} I(U_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1})$$
$$\stackrel{(f)}{=} I(\bar{U}_{\tilde{k}}; Y_{\tilde{k}} | U_{\tilde{k}-1}), \tag{130}$$

where (a) follows by the definition of $\bar{U}_{\tilde{k}}$ and $\bar{U}_{\tilde{k}+1}$; (b) by the Markov chain (99); (c) by the chain rule of mutual information and the Markov chain (99); (d) by the degradedness of the channel (64); (e) by the definition of β in (127); and (f) by (129).

We can thus conclude that also when (125b) holds, the upper bound on $C(M_1, \ldots, M_K)$ in (104) 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 $(F(1-\delta_{\mathrm{s}}),\ 2\tilde{\Gamma}_2)$

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

$$R = F(1-\delta_{\mathrm{s}}) \quad \text{and} \quad M = 2\tilde{\Gamma}_2. \tag{131}$$

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

$$R^{(1)} := F(\delta_{\mathrm{w}} - \delta_{\mathrm{s}}), \tag{132a}$$
$$R^{(2)} := F(1 - \delta_{\mathrm{w}}) - \epsilon, \tag{132b}$$

for an arbitrarily small $\epsilon > 0$.

Placement Phase: Cache the triple

$$V_1 := \left(W_1^{(1)}, W_2^{(1)}, W_1^{(2)} \oplus W_2^{(2)}\right) \tag{133}$$

in the weak receiver's cache.

Delivery Phase: The strong receiver has to learn $W_{d_2}^{(1)}$ and $W_{d_2}^{(2)}$. The weak receiver only needs to learn $W_{d_2}^{(2)}$, because it has already stored $W_{d_1}^{(1)}$ in its cache memory. 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 $W_{d_2}^{(2)}$ to both receivers. For this purpose, construct a random codebook with $\lfloor 2^{nR^{(2)}} \rfloor \times \lfloor 2^{nR^{(2)}} \rfloor$ length-n codewords by randomly and independently drawing each entry according to a Bernoulli-$1/2$ distribution. Arrange the codewords in an array with $\lfloor 2^{nR^{(2)}} \rfloor$ rows and $\lfloor 2^{nR^{(2)}} \rfloor$ columns. The transmitter sends the codeword that lies in the column corresponding to Message $W_{d_2}^{(2)}$ and in the row corresponding to Message $W_{d_2}^{(1)}$.

The strong receiver decodes both messages. The weak receiver 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 $W_{d_2}^{(2)}$ 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 (132), 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 (131).

APPENDIX G
PROOF OF UPPER BOUND IN THEOREM 5

The first and last terms in (84) are special cases of Theorem 4 for $k_{\mathrm{w}} = 1$ and $k_{\mathrm{w}} = 0$, respectively. Here, we prove the second term by showing that for every achievable rate-memory pair $(R,\ M)$,

$$3R \leq M + (1 - \delta_{\mathrm{w}}) + (1 - \delta_{\mathrm{s}}). \tag{134}$$

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

$$X_t \to Y_{2,t} \to Y_{1,t}. \tag{135}$$

For all sufficiently large blocklengths n, choose placement functions $\{g_i^{(n)}\}$ as in (9), encoding functions $f_{\mathbf{d}}^{(n)}$ as in (10), and decoding functions $\{\varphi_{i,\mathbf{d}}^{(n)}\}$ as in (12) so that the probability of worst-case error $P_{\mathrm{e}}^{\mathrm{worst}}$ 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

$$V_1^{(n)} = g_1^{(n)}(W_1, W_2), \tag{136}$$
$$X_{\mathbf{d}}^n = f_{\mathbf{d}}^{(n)}(W_1, W_2), \tag{137}$$

denote cache contents and the input of the packet-erasure BC for a given demand vector $\mathbf{d} \in \mathcal{D}^2$ and for above chosen placement and encoding functions. Also, let $Y^n_{1,\mathbf{d}}$ and $Y^n_{2,\mathbf{d}}$ 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 $(Y^n_{1,\mathbf{d}_1}, V_1^{(n)})$ and from $Y^n_{2,\mathbf{d}_2}$, and W_2 should be decodable from $(Y^n_{1,\mathbf{d}_2}, V_1^{(n)})$. Thus, by Fano's inequality, for all $\epsilon_1, \epsilon_2, \epsilon_3 > 0$ and sufficiently large blocklength n, we have

$$nR \leq I(W_1; V_1^{(n)}, Y^n_{1,\mathbf{d}_1}) + n\epsilon_1, \tag{138a}$$

$$nR \leq I(W_1; Y^n_{2,\mathbf{d}_2}) + n\epsilon_2, \tag{138b}$$

$$nR = I(W_2; V_1^{(n)}, Y^n_{1,\mathbf{d}_1}, Y^n_{1,\mathbf{d}_2} | W_1) + n\epsilon_3, \tag{138c}$$

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:

$$\begin{aligned} nR &\leq \sum_{t=1}^{n} I(W_1; Y_{\mathbf{d}_2,t} | Y^{t-1}_{2,\mathbf{d}_2}) + n\epsilon_2 \\ &\leq (1-\delta_{\text{s}}) \sum_{t=1}^{n} I(W_1; X_{\mathbf{d}_2,t} | Y^{t-1}_{2,\mathbf{d}_2}) + n\epsilon_2. \end{aligned} \tag{139}$$

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

$$\begin{aligned} &2nR \\ &\leq I(W_1, W_2; V_1^{(n)}, Y^n_{1,\mathbf{d}_1}) + I(W_2; Y^n_{1,\mathbf{d}_2} | W_1, V_1^{(n)}, Y^n_{1,\mathbf{d}_1}) + n\epsilon' \\ &\overset{(a)}{\leq} I(W_1, W_2; V_1^{(n)}) + I(W_1, W_2; Y^n_{1,\mathbf{d}_1} | V_1^{(n)}) \\ &\quad + I(W_2; Y^n_{2,\mathbf{d}_2} | W_1, V_1^{(n)}, Y^n_{1,\mathbf{d}_1}) + n\epsilon' \\ &= I(W_1, W_2; V_1^{(n)}) + \sum_{t=1}^{n} I(W_1, W_2; Y_{1,\mathbf{d}_1,t} | V_1^{(n)}, Y^{t-1}_{1,\mathbf{d}_1}) \\ &\quad + \sum_{t=1}^{n} I(W_2; Y_{2,\mathbf{d}_2,t} | W_1, V_1^{(n)}, Y^n_{1,\mathbf{d}_1}, Y^{t-1}_{2,\mathbf{d}_2}) + n\epsilon' \\ &= I(W_1, W_2; V_1^{(n)}) + (1-\delta_{\text{w}}) \sum_{t=1}^{n} I(W_1, W_2; X_{\mathbf{d}_1,t} | V_1^{(n)}, Y^{t-1}_{1,\mathbf{d}_1}) \\ &\quad + (1-\delta_{\text{s}}) \sum_{t=1}^{n} I(W_2; X_{\mathbf{d}_2,t} | W_1, V_1^{(n)}, Y^n_{1,\mathbf{d}_1}, Y^{t-1}_{2,\mathbf{d}_2}) + n\epsilon' \\ &\leq I(W_1, W_2; V_1^{(n)}) + (1-\delta_{\text{w}}) \sum_{t=1}^{n} I(W_1, W_2; X_{\mathbf{d}_1,i} | V_1^{(n)}, Y^{t-1}_{1,\mathbf{d}_1}) \\ &\quad + (1-\delta_{\text{s}}) \sum_{t=1}^{n} I(W_2, V_1^{(n)}, Y^n_{1,\mathbf{d}_1}; X_{\mathbf{d}_2,t} | W_1, Y^{t-1}_{2,\mathbf{d}_2}) + n\epsilon' \\ &\leq nM + n(1-\delta_{\text{w}}) \\ &\quad + (1-\delta_{\text{s}}) \sum_{t=1}^{n} I(W_2, V_1^{(n)}, Y^n_{1,\mathbf{d}_1}; X_{\mathbf{d}_2,t} | W_1, Y^{t-1}_{2,\mathbf{d}_2}) + n\epsilon'. \end{aligned} \tag{140}$$

In (a), we used that the physically degradedness of the channel in (135) implies the Markov chain

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

Adding up (139) and (140) and letting $\epsilon_1, \epsilon_2, \epsilon_3$ tend to 0, we obtain the missing converse bound in (134), because

$$\begin{aligned} &I(W_2, V_1^{(n)}, Y^n_{1,\mathbf{d}_1}; X_{\mathbf{d}_2,t} | W_1, Y^{t-1}_{2,\mathbf{d}_2}) + I(W_1; X_{\mathbf{d}_2,t} | Y^{t-1}_{2,\mathbf{d}_2}) \\ &= I(W_1, W_2, V_1^{(n)}, Y^n_{1,\mathbf{d}_1}; X_{\mathbf{d}_2,t} | Y^{t-1}_{2,\mathbf{d}_2}) \\ &\leq H(X_{\mathbf{d}_2,t}) \leq 1. \end{aligned} \tag{141}$$

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," *arXiv*, 1602.06215, Feb. 2016.

[2] P. Cerwall, P. Jonsson, R. Möller, S. Bävertoft, S. Carson, I. Godor, P. Kersch, A. Kälvemark, G. Lemne, and P. Lindberg, Ericsson Mobility Report, Jun. 2015.

[3] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, Jun. 2014.

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

[5] Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: Improved bounds for small buffer users," in *arXiv, 1407.1935*, Nov. 2015.

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

[7] K. Wan, D. Tuninetti, and P. Piantanida, "On caching with more users than files," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Barcelona, Spain, Jul. 2016.

[8] S. Sahraei and M. Gastpar, "K users caching two files: An improved achievable rate," in *Proc. 50th Annu. Conf. on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2016.

[9] A. Sengupta, R. Tandon, and T. C. Clancy, "Improved approximation of storage-rate tradeoff for caching via new outer bounds," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Hong Kong, China, Jun. 2015.

[10] C. Tian and J. Chen, "Caching and delivery via interference elimination," in *arXiv, 1604.08600*, Apr. 2016.

[11] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," in *arXiv, 1501.06003*, Feb. 2016.

[12] H. Ghasemi and A. Ramamoorthy, "Further results on lower bounds for coded caching," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Barcelona, Spain, Jul. 2016.

[13] C. Tian, "Symmetry, demand types and outer bounds in caching systems," in *Proc. IEEE Int. Symp. Information Theory (ISIT)*, Barcelona, Spain, Jul. 2016.

[14] A. Amiri, Q. Yang, and D. Gündüz, "Coded caching for a large number of users," in *Proc. IEEE Information Theory Workshop (ITW),* Cambridge, United Kingdom, Sep. 2016.

[15] J. Zhang, X. Lin, and X. Wang, "Coded caching under arbitrary popularity distributions," in *Proc. Inf. Theory and Appl. Workshop (ITA),* San Diego, CA, Feb. 2015.

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

[17] 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.

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

[19] R. Timo, S. Saeedi Bidokhti, M. Wigger and B. Gieger, "A rate-distortion approach to caching," in Proc. *International Zurich Seminar on Communications (IZS)*, Zurich, Switzerland, March, 2016.

[20] C.-Y. Wang, S. H. Lim, and M. Gastpar, "Information-theoretic caching: Sequential coding for computing," *arXiv*, 1504.00553, April, 2015.

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

[22] Q. Yang and D. Gündüz, "Centralized coded caching for heterogeneous lossy requests," in Proc. *IEEE Int. Symp. Information Theory (ISIT)*, Barcelona, Span, Jul. 2016, pp. 405–409.

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

[24] 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.

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

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

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

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

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

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

[31] J. Zhang and P. Elia, "Fundamental limits of cache-aided wireless BC: interplay of coded-caching and CSIT feedback." Online: *arXiv*, 1511.03961, Nov. 2015.

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

[33] J. Zhang and P. Elia, "Feedback-aided coded caching for the MISO BC with small caches," *arXiv* 1606.05396, June 2016.

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

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

[36] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in *Proc. IEEE Intl. Symp. Inform. Theory,* Hong Kong China, Jun. 2015.

[37] N. Naderializadeh, M. A. Maddah-Ali, and A. S. Avestimehr, "Fundamental limits of cache-aided interference management," *arXiv*, 1602.04207, Feb. 2016.

[38] S. Pooya Shariatpanah, S. Abolfazl Motahari, B. Hossein Khalaj, "Multi-server coded caching." Online: *arXiv*, 1503.00265, Mar. 2015.

[39] M. Wigger, R. Timo, and S. Shamai, "Complete Interference Mitigation through Receiver-Caching in Wyner's Networks," in *Inf. Theory Workshop*, Cambridge, Sep., 2016.

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

[41] S. P. Shariatpanahi, S. A. Motahari, and B. H. Khalaj, "Multi-server coded caching," *arXiv*, 1503.00265, Mar. 2015.

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

[43] S.-H. Park, O. Simeone, and S. Shamai (Shitz), "Joint optimization of cloud and edge processing for fog radio access networks." Online: *arXiv*, 1601.02460, Jan. 2016.

[44] R. Tandon and O. Simeone, "Cloud-aided wireless networks with edge caching: Fundamental latency trade-offs in fog radio access networks." Jan. 2016.

[45] B. Azari, O. Simeone, U. Spagnolini and A. Tulino, "Hypergraph-based analysis of clustered cooperative beamforming with application to edge caching." to appear in *IEEE Wireless Comm. Letters*.

[46] M. Peng, S. Yan, K. Zhang and C. Wang, "Fog computing based radio access networks: issues and challenges." Online: *arXiv*, 1506.04233, June 2015.

[47] J. Hachem, N. Karamchandani, and S. Diggavi, "Effect of number of users in multi-level coded caching," in Proc. *IEEE ISIT*, June 2015.

[48] J. Hachem, N. Karamchandani and S. Diggavi, "Multi-level coded caching," in proceedings *IEEE ISIT*, Honolulu (HI), USA, July, 2014.

[49] S. Sahraei and M. Gastpar, "Multi-library coded caching," *arXiv*, 1601.06016, January, 2016.

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

[51] G. Kramer and S. Shamai, "Capacity for classes of broadcast channels with receiver side information," in Proc. *IEEE Information Theory Workshop (ITW)*, Lake Tahoe, California, September, 2007.

[52] 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.
[53] Y. Gao and E. Tuncel, “Wyner-Ziv coding over broadcast channels: hybrid digital/analog schemes,” *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 5660–5672, September, 2009.
[54] J. Nayak, E. Tuncel, and D. Gunduz “Wyner-Ziv coding over broadcast channels: digital schemes,” *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1782–1799, Apr. 2010.
[55] 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.
[56] R. Timo and M. Wigger, ”Slepian-Wolf coding for broadcasting with cooperative base-stations,” in *IEEE Trans. on Comm.*, vol. 63, no. 5, pp. 1850–1866, May 2015.
[57] A. Bracher and M. Wigger, “Feedback and partial message side-information on the semideterministic broadcast channel,” *arXiv*, 1508.01880, Aug. 2015.
[58] 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.
[59] T. Laich and M. Wigger, ”Utility of encoder side information for the lossless kaspi/heegard-b erger problem, ” in *Proc. ISIT 2013*, July 2013, pp. 3065–3069.
[60] S. Diggavi, E. Perron, and E. Telatar, ”Lossy source coding with gaussian or erased side-information,” in *Proc. ISIT 2009*, Seoul, Korea, June 28–July 3 2009, pp. 1035–1039.
[61] 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 Transactions on Information Theory*, vol. 62, no. 4, pp. 2009–2033, Apr. 2016.
[62] O. Shayevitz and M. Wigger, “On the capacity of the discrete memoryless broadcast channel with feedback,” *IEEE Trans. on Inf. Th.,* vol. 59, no. 3, pp. 1329–1345, Mar. 2013.
[63] R. Venkataramanan and S. S. Pradhan, “An achievable rate region for the broadcast channel with feedback,” *IEEE Trans. on Inf. Th.,* vol. 59, no. 10, pp. 6175–6191, Oct. 2013.
[64] C. C. Wang, “Capacity of 1-to-k broadcast erasure channel with output feedback,” in *Proc. 48th Ann. Allerton Conf.* , Monticello, IL, Sep. 2010.
[65] L. Georgiadis and L. Tassiulas, “Broadcast erasure channel with feedback capacity and algorithms,” *in Proc. 5th Workshop on Network Coding, Theory, & Applications,* Lausanne, Switzerland, June 2009.
[66] M. A. Maddah-Ali and D. Tse, “Completely Stale Transmitter Channel State Information is Still Very Useful,” *IEEE Trans. on Inf. Th.,* vol. 58, no. 7, pp. 4418–4431, July 2012.
[67] H. Kim, Y.-K. Chia, and A. El Gamal, “A note on broadcast channels with stale state information at the transmitter,” *IEEE Trans. on Inf. Th.,* vol. 61, no. 7, pp. 3622–3631, July 2015.
[68] Y. Wu, “Coding schemes for discrete memoryless multicast networks with and without Feedback,” in *Proc. 53th Ann. Allerton Conf.* , Monticello, IL, Sept. 2015.
[69] J. Chen and P. Elia, “Toward the performance versus feedback tradeoff for the two-user MISO broadcast channel,” *IEEE Trans. on Inf. Th.,* vol. 59, no. 12, pp. 8336–8356, Dec. 2013.
[70] S. Yang, M, Kobayashi, D. Gesbert, and X. Yi, “Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT,” *IEEE Trans. on Inf. Th.*, vol. 59, no. 1, pp. 315–328, Jan. 2013.
[71] R. Urbanke and A. Wyner, “Packetizing for the erasure broadcast channel with an internet application,” technical report, 1999.
[72] R. G. Gallager, “Capacity and coding for degraded broadcast channels,” *Problem. Inf. Transm.*, vol 10, no. 3, pp. 3–14, 1974.
[73] A. El Gamal and Y. H. Kim, “Network information theory,” Cambridge University Press, 2012.
[74] T. M. Cover and J. Y. Thomas. *Elements of Information Theory*, Wiley-Interscience, New York, 2006.