An Upper Bound on the Capacity-Memory Tradeoff of Degraded Broadcast Channels

Shirin Saeedi Bidokhti[1,3], Michèle Wigger[2] and Roy Timo[3]
[1]Stanford University, saeedi@stanford.edu
[2]LTCI CNRS, Telecom ParisTech, michele.wigger@telecom-paristech.fr
[3]Technical University of Munich, roy.timo@tum.de

***Abstract*—We provide a general upper bound on the capacity-memory tradeoff over degraded broadcast channels (BCs) with cache memories at the receivers. The bound holds for arbitrary demands and considers a worst-case scenario.**

I. Introduction

We address a one-to-many broadcast scenario 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 This pre-placement of information is called the *caching communications phase*, and it occurs 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 can try to cache information that is likely to be useful for many users during the *delivery communications phase* (the peak-traffic time when the users demand files from the server).

The information-theoretic aspects of cache-aided communications have received significant attention in recent years [1]–[13]. Maddah-Ali and Niesen [1] considered a cache-aided noise-less broadcast network and presented achievability and converse results on the total required delivery rate over the noiseless broadcast link in function of the cache memory sizes at the receivers. Tighter converse bounds were presented in [2]–[5]. The converse bound in [1] applies to *worst-case scenarios* where the delivery rate needs to suffice for all possible receiver demands. The converse bounds in [2]–[5] applied to *average-case scenarios* where the receivers' demands follow a given probability distribution and the delivery rate is averaged over this demand distribution.

In contrast to these previous works [1]–[5], we assume in this paper that the delivery phase takes place over a noisy broadcast channel (BC). (Noisy channel models for the delivery phase were also considered in [6]–[13].) For simplicity we focus on the class of (stochastically) degraded BCs. Our main result is a converse on the fundamental rate-memory tradeoff for cache-aided degraded BCs. That is, we provide an upper bound on the maximum equal rate at which messages can be reliably communicated to the receivers over a degraded BC in function of the receivers' cache sizes. We assume that the receivers' demands are arbitrary and our converse result holds for a worst-case scenario. Finally, we specialize our converse result to packet erasure BCs.

This work was supported by the Alexander von Humboldt Foundation and the Swiss National Science Foundation.

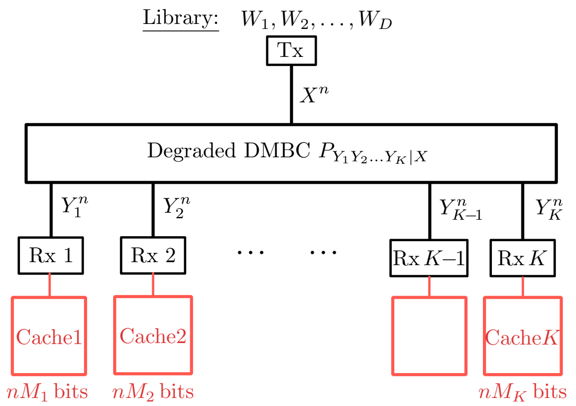

Fig. 1: Degraded K-user BC $P_{Y_1Y_2\cdots Y_K|X}$ where each Receiver $k \in \{1, \ldots, K\}$ has cache memory of size nM_k bits.

II. Problem Definition

Consider a degraded broadcast channel (BC) with a single transmitter and K receivers as shown in Fig. 1. Each user $k \in \mathcal{K} := \{1, \ldots, K\}$ has cache memory $V_K \in \{1, \ldots, \lfloor 2^{nM_k} \rfloor\}$.

We model the channel from the transmitter to the receivers by a memoryless *degraded BC* with input alphabet $\mathcal{X}$ and equal output alphabets $\mathcal{Y}$. The joint transitional law of the memoryless BC is given by $P_{Y_1Y_2\cdots Y_K|X}(y_1, \ldots, y_K|x)$. We assume that the BC is *degraded*, i.e., the transition law satisfies

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

For our problem setup only the marginal transition law is relevant. Therefore our main result in Theorem 1 holds also for *stochastically degraded BC*, i.e., for transition laws $P_{Y_1,\ldots,Y_K|X}$ for which there exist conditional probability distributions $\tilde{P}_{Y_2|Y_1}, \tilde{P}_{Y_3|Y_2}, \ldots, \tilde{P}_{Y_{K-1}|Y_K}$ such that for all $(x, y_1, y_2, \ldots, y_K) \in \mathcal{X} \times \mathcal{Y}_1 \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_K$ the BC's transition law satisfies

$$\begin{aligned} &P_{Y_1Y_2\cdots Y_K|X}(y_1, \ldots, y_K|x) \\ &= P_{Y_K|X}(y_k|x)\tilde{P}_{Y_{K-1}|Y_K}(y_{k-1}|y_k)\ldots\tilde{P}_{Y_1|Y_2}(y_1|y_2). \end{aligned} \tag{2}$$

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

$$W_1, \ldots, W_D. \tag{3}$$

These messages are all independent of each other and each of them is uniformly distributed over the message set

$\{1,\dots,\lfloor 2^{nR}\rfloor\}$, 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

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

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 that the demand vector

$$\mathbf{d} := (d_1,\dots,d_K) \tag{4}$$

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

Communication takes place in two phases: a first *caching phase* and a subsequent *delivery phase*.

During the caching phase, the transmitter sends caching information V_k to each receiver $k \in \mathcal{K}$, who then stores this information in its cache memory. The demand vector $\mathbf{d}$ is unknown to the transmitter and receivers during the caching phase, and, therefore, the cached information V_k cannot depend on the users' specific demands $\mathbf{d}$. Instead, V_k is a function of the entire library:

$$V_k := g_k(W_1,\dots,W_D) \qquad k \in \mathcal{K}$$

for some function

$$g_k \colon \left\{1,\dots,\lfloor 2^{nR}\rfloor\right\}^D \to \mathcal{V}_k, \qquad i \in \mathcal{K} \tag{5}$$

where $\mathcal{V}_k := \{1,\dots,\lfloor 2^{nM_k}\rfloor\}$. The caching phase occurs during a low-congestion period. We therefore assume that this phase incurs no erasures or other types of errors, and each receiver $i \in \mathcal{K}$ can store V_k in its cache memory.

After the caching phase and prior to the delivery phase, the transmitter and all receivers are provided with the demand vector $\mathbf{d}$.[1]

Depending on the demand vector $\mathbf{d}$, the transmitter chooses an encoding function

$$f_{\mathbf{d}} \colon \{1,\dots,\lfloor 2^{nR}\rfloor\}^D \to \mathcal{X}^n \tag{6}$$

and it sends

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

over the BC.

Each receiver $k \in \mathcal{K}$ observes Y_k^n according to the channel transition law $P_{Y_1Y_2\cdots Y_K|X}(y_1,\dots,y_K|x)$. It attempts to reconstruct its desired message from its channel outputs Y_k^n, its cache content V_k and the demand vector $\mathbf{d}$. More formally,

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

where

$$\varphi_{k,\mathbf{d}} \colon \mathcal{Y}_k^n \times \mathcal{V}_k \to \{1,\dots,\lfloor 2^{nR}\rfloor\} \qquad k \in \mathcal{K}. \tag{8}$$

An error is said to occur whenever

$$\hat{W}_k \neq W_{d_k} \quad \text{for some} \quad k \in \{1,\dots,K\}. \tag{9}$$

[1]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.

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

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

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

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

We say that a rate-memory tuple $(R, M_1,\dots,M_K)$ is *achievable* if for every $\epsilon > 0$ there exists a sufficiently large blocklength n and caching, encoding and decoding functions as in (5), (6) and (7) such that $\mathsf{P}_{\mathrm{e}}^{\text{worst}} < \epsilon$. The main problem of interest in this paper is to determine the following capacity versus cache memory tradeoff.

Definition 1: Given cache memory sizes $M_1,\dots,M_K$, we define the *capacity-memory tradeoff* $C(M_1,\dots,M_K)$ as the supremum of all rates R such that the rate-memory tuple $(R, M_1,\dots,M_K)$ is achievable.

III. Results

For each $\mathcal{S} \in \mathcal{K}$, let $R_{\text{sym},\mathcal{S}}$ denote the largest equal-rate that is achievable over a BC with receivers in $\mathcal{S}$ when there are no cache memories.

Theorem 1: The capacity-memory tradeoff $C(M_1,\dots,M_K)$ of a degraded BC is upper bounded as

$$C(M_1,\dots,M_K) \leq \min_{\mathcal{S}\subseteq\{1,\dots,K\}} \left(R_{\text{sym},\mathcal{S}} + \frac{M_{\mathcal{S}}}{D}\right),$$

where $M_{\mathcal{S}} = \sum_{k\in\mathcal{S}} M_k$ is the total cache size at receivers in $\mathcal{S}$.

Remark 1: Theorem 1 also holds for stochastically degraded BCs.

We specialize this theorem to the packet-erasure BC.

Corollary 1.1: The capacity-memory tradeoff $C(M_1,\dots,M_K)$ of the packet-erasure BC with packet size F, erasure probabilities $\delta_1,\dots,\delta_K \geq 0$, and cache memory sizes $M_1,\dots,M_K$, is upper bounded as

$$C(M_1,\dots,M_K) \leq \min_{\mathcal{S}\subseteq\{1,\dots,K\}} \left(\left(\sum_{k\in\mathcal{S}} \frac{1}{1-\delta_k}\right)^{-1} + \frac{M_{\mathcal{S}}}{D}\right). \tag{10}$$

As shown in [9], the upper bound in Corollary 1.1 is not always tight. It is however tight in some special cases. Figure 2 illustrates our upper bound on the capacity memory-tradeoff for a two-user packet erasure BC with packet size $F = 10$ bits, erasure probabilities $\delta_1 = 0.8$ and $\delta_1 = 0.2$ when receiver one has cache-memory size M_1 and receiver 2 has no cache. The upper bound is compared to the best known lower bound from [9] that uses superposition coding and piggyback coding for the delivery phase.

IV. Proof of Theorem 1

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

$$C(M_1,\dots,M_K) \leq \left(R_{\text{sym},\mathcal{K}} + \frac{1}{D}\sum_{k=1}^{K} M_k\right), \tag{11}$$

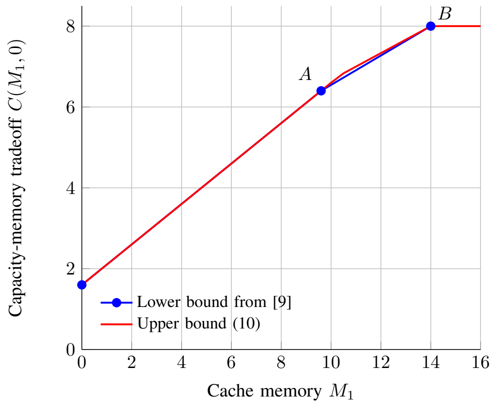

Fig. 2: Bounds on the capacity-memory tradeoff of a two-user packet-erasure BC.

where here $R_{\text{sym},\mathcal{K}}$ denotes the largest symmetric rate that is achievable over the BC $P_{Y_1Y_2\cdots Y_K|X}$ when there are no caches. The inequalities in the theorem that correspond to other subsets $\mathcal{S} \subseteq \{1,\ldots,K\}$ can be proved in an analogous way.

We start the proof of (11). 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 caching functions $\{g_i^{(n)}\}$, an encoding function $\{f_{\mathbf{d}}^{(n)}\}$, and K decoding functions $\{\varphi_{i,\mathbf{d}}^{(n)}\}$ so that the probability of worst-case error $\mathsf{P}_e(\mathbf{d})$ tends to 0 as $n \to \infty$. For each n let

$$V_k^{(n)} = g_k^{(n)}(W_1,\ldots,W_D), \qquad k \in \{1,\ldots,K\},$$

denote the cache contents for the chosen caching functions.

Lemma 2: 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 $(U_{1,\mathbf{d}},\ldots,U_{K,\mathbf{d}},X_{\mathbf{d}},Y_{1,\mathbf{d}},\ldots,Y_{K,\mathbf{d}})$ such that

$$U_{1,\mathbf{d}} - U_{2,\mathbf{d}} - \cdots - U_{K,\mathbf{d}} - X_{\mathbf{d}} - Y_{K,\mathbf{d}} - Y_{K-1,\mathbf{d}} \cdots - Y_{1,\mathbf{d}} \quad (12)$$

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

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

and so that the following K inequalities hold:

$$R - \epsilon \le \frac{1}{n} I\big(W_{d_1}; V_1^{(n)},\ldots,V_K^{(n)}\big) + I\big(U_{1,\mathbf{d}};Y_{1,\mathbf{d}}\big), \quad (13a)$$

$$R - \epsilon \le \frac{1}{n} I\big(W_{d_k}; V_1^{(n)},\ldots,V_K^{(n)}|W_{d_1},\ldots,W_{d_{k-1}}\big) + I\big(U_{k,\mathbf{d}};Y_{k,\mathbf{d}}|U_{k-1,\mathbf{d}}\big), \ \forall k \in \{2,\ldots,K\} \quad (13b)$$

Proof: The proof is similar to the converse proof of the capacity of degraded BCs without caching [14, Theorem 5.2]. It is omitted due to lack of space. ■

Fix $\epsilon > 0$ and a blocklength n (depending on this ϵ) so that Lemma 2 holds for all demand vectors $\mathbf{d}$ that have all different entries. We average the bound obtained in (19) over different demand vectors. Let $\mathcal{Q}$ be the set of all the $\binom{D}{K}K!$ 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 other random variables. Define: $U_1 := (U_{1,Q}, Q)$; $U_k := U_{k,Q}$, for $k \in \{2,\ldots,K\}$; $X_k := X_Q$; and $Y_k := Y_{k,Q}$ for $k \in \{1,\ldots,K\}$. Notice that they form the Markov chain

$$U_1 \to U_2 \to \cdots \to U_K \to X \to (Y_1,\ldots,Y_K) \quad (14)$$

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

$$(Y_1,Y_2,\ldots,Y_K) \sim P_{Y_1\cdots Y_K|X}(\cdots|x). \quad (15)$$

Averaging inequalities (19) over the demand vectors in $\mathcal{Q}$ and using standard arguments to take care of the time-sharing random variable Q, we obtain:

$$R - \epsilon \le \alpha_1 + I\big(U_1;Y_1\big), \quad (16a)$$

$$R - \epsilon \le \alpha_k + I\big(U_k;Y_k|U_{k-1}\big), \quad \forall k \in \{2,\ldots,K\}, \quad (16b)$$

where we defined $\alpha_1,\ldots,\alpha_K$ as follows:

$$\alpha_1 := \frac{1}{\binom{D}{K}K!} \sum_{\mathbf{d}\in\mathcal{Q}} \frac{1}{n} I(W_{d_1}; V_1^{(n)},\ldots,V_K^{(n)}), \quad (17a)$$

$$\alpha_k := \frac{1}{\binom{D}{K}K!} \sum_{\mathbf{d}\in\mathcal{Q}} \frac{1}{n} I(W_{d_k}; V_1^{(n)},\ldots,V_K^{(n)}|W_{d_1},\ldots,W_{d_{k-1}}). \quad (17b)$$

Lemma 3: Parameters α_k, $k = 1,\ldots,K$, defined in (17), satisfy the following constraints:

$$\alpha_k \ge 0, \qquad k \in \{1,\ldots,K\} \quad (18a)$$

$$\alpha_{k'} \le \alpha_k, \qquad k,k' \in \{1,\ldots,K\},\ k' \le k, \quad (18b)$$

$$\sum_{k\in\mathcal{K}} \alpha_k \le \frac{K}{D} \sum_{k\in\mathcal{K}} M_k. \quad (18c)$$

Proof: See Appendix A. ■

Taking $\epsilon \to 0$, by (16) and (17) and by Lemma 3, we conclude that the capacity-memory tradeoff $C(M_1,\ldots,M_K)$ is upper bounded by the following K inequalities:

$$C(M_1,\ldots,M_K) \le \alpha_1 + I\big(U_1;Y_1\big), \quad (19a)$$

$$C(M_1,\ldots,M_K) \le \alpha_k + I\big(U_k;Y_k|U_{k-1}\big), \quad \forall k \in \{2,\ldots,K\}, \quad (19b)$$

for some $\alpha_1,\ldots,\alpha_K$ satisfying (18) and some $U_1,\ldots,U_K,X,Y_1,\ldots,Y_K$ satisfying (14) and (15).

Lemma 4: Replacing each and every real number $\alpha_1,\ldots,\alpha_K$ in (19) by $\frac{1}{D}\sum_{k\in\{1,\ldots,K\}} M_k$ leads to a relaxed upper bound on $C(M_1,\ldots,M_K)$.

Proof: See Appendix B. ■

Thus,

$$C(M_1,\ldots,M_K) - \frac{1}{D} \sum_{k\in\{1,\ldots,K\}} M_k \le I\big(U_1;Y_1\big), \quad (20a)$$

$$C(M_1,\ldots,M_K) - \frac{1}{D} \sum_{k\in\{1,\ldots,K\}} M_k \le I\big(U_k;Y_k|U_{k-1}\big), \quad \forall k \in \{2,\ldots,K\}, \quad (20b)$$

for some $U_1,\ldots,U_K,X,Y_1,\ldots,Y_K$ satisfying (14) and (15).

All K constraints in (20) have the same LHS, and their RHSs coincide with the rate-constraints of a degraded BC without caches. Therefore, the choice of the auxiliaries

$(U_1, \ldots, U_K)$ 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 of the degraded BC without caches. This establishes the equivalence of (20) with the desired bound in (11), and thus concludes the proof.

Appendix A
Proof of Lemma 3

Constraint (18a) follows by the nonnegativity of mutual information. To prove Constraint (18b), 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 $k \in \{1, \ldots, K\}$, let $\overrightarrow{d}_k^{(\ell)}$ denote the k-th index of demand vector $\overrightarrow{\mathbf{d}}^{(\ell)}$. So,

$$\overrightarrow{d}_k^{(\ell)} = d_{(k-\ell) \bmod K} \tag{21}$$

where we define $(K) \bmod K = K$ (and not 0).

For each $\ell \in \{1, \ldots, K–1\}$ and $k, k' \in \{2, \ldots, K\}$, $k' \leq k$:

$$\begin{aligned}
I(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}) &\overset{(a)}{=} I(W_{\overrightarrow{d}_{k'}^{(k'-1)}}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\overset{(b)}{\leq} I(W_{\overrightarrow{d}_{k'}^{(k'-1)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\overrightarrow{d}_1^{(k'-1)}}, \ldots, W_{\overrightarrow{d}_{k'-1}^{(k'-1)}}) \\
&\overset{(a)}{=} I(W_{\overrightarrow{d}_k^{(k-1)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\overrightarrow{d}_2^{(k-1)}}, \ldots, W_{\overrightarrow{d}_{k-1}^{(k-1)}}) \\
&\overset{(b)}{\leq} I(W_{\overrightarrow{d}_k^{(k-1)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\overrightarrow{d}_1^{(k-1)}}, \ldots, W_{\overrightarrow{d}_{k-1}^{(k-1)}}),
\end{aligned} \tag{22}$$

where (a) follows by (21) and (b) is by the independence of messages.

Fix a demand vector $\mathbf{d} \in \mathcal{Q}$ and sum up the above inequality (22) over all K cyclic shifts $\mathbf{d}^{(0)}, \mathbf{d}^{(1)}, \ldots, \mathbf{d}^{(K)}$ of $\mathbf{d}$ to obtain:

$$\begin{aligned}
&\sum_{\ell=0}^{K-1} I(W_{\overrightarrow{d}_1^{(\ell)}}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\quad \leq \sum_{\ell=0}^{K-1} I(W_{\overrightarrow{d}_{k'}^{(\ell)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\overrightarrow{d}_1^{(\ell)}}, \ldots, W_{\overrightarrow{d}_{k'-1}^{(\ell)}}) \\
&\quad \leq \sum_{\ell=0}^{K-1} I(W_{\overrightarrow{d}_k^{(\ell)}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{\overrightarrow{d}_1^{(\ell)}}, \ldots, W_{\overrightarrow{d}_{k-1}^{(\ell)}}).
\end{aligned} \tag{23}$$

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 (23):

$$\begin{aligned}
&\sum_{\mathbf{d} \in \mathcal{Q}} I(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\quad \leq \sum_{\mathbf{d} \in \mathcal{Q}} I(W_{d_{k'}}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{k'-1}}) \\
&\quad \leq \sum_{\mathbf{d} \in \mathcal{Q}} I(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, \ldots, W_{d_{k-1}}).
\end{aligned} \tag{24}$$

This proves (18b).

We proceed to prove Constraint (18c). For each $\mathbf{d} \in \mathcal{Q}$:

$$\begin{aligned}
&I(W_{d_1}; V_1^{(n)}, \ldots, V_K^{(n)}) \\
&\quad + \sum_{k=2}^{K} I(W_{d_k}; V_1^{(n)}, \ldots, V_K^{(n)} | W_{d_1}, W_{d_2}, \ldots, W_{d_{k-1}}) \\
&\quad = I(W_{d_1}, W_{d_2}, \ldots, W_{d_{K-1}}; V_1^{(n)}, \ldots, V_K^{(n)}).
\end{aligned} \tag{25}$$

So,

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

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 5).

Proposition 5: 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 $\ell \in \{1, \ldots, L\}$:

$$\frac{1}{\binom{L}{\ell}} \sum_{\mathcal{S} \subseteq \{1, \ldots, L\} : |\mathcal{S}| = \ell} \frac{H(A_{\mathcal{S}})}{\ell} \geq \frac{1}{L} H(A_1, \ldots, A_L). \tag{26}$$

Proof: See [15, Theorem 17.6.1]. ∎

Appendix B
Proof of Lemma 4

Fix random variables $U_1, U_2, \ldots, U_K, X$ satisfying the Markov chain (14) and real numbers $\alpha_1, \ldots, \alpha_K$ satisfying (18). We will show that if $\alpha_{\tilde{k}} \neq \alpha_{\tilde{k}+1}$ for some $\tilde{k} \in \mathcal{K}$, then we can find new random variables $\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_K, \bar{X}$ satisfying the Markov chain (14) and real numbers $\bar{\alpha}_1, \ldots, \bar{\alpha}_K$ satisfying (18) so that the upper bound on $C(M_1, \ldots, M_K)$ in (19) is relaxed if we replace

$$(U_1, U_2, \ldots, U_K, X) \qquad \text{and} \qquad (\alpha_1, \ldots, \alpha_K)$$

by

$$(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_K, \bar{X}) \qquad \text{and} \qquad (\bar{\alpha}_1, \ldots, \bar{\alpha}_K).$$

This proves that we obtain a relaxed upper bound on $C(M_1, \ldots, M_K)$ if in (19) we replace all numbers $\alpha_1, \ldots, \alpha_K$ by the same number α. By (18c) this number $\alpha \leq \frac{1}{D}\sum_{k\in\{1,\ldots,K\}} M_k$, and by the monotonicity of the RHSs of (19) in $\alpha_1, \ldots, \alpha_K$ the choice $\alpha = \frac{1}{D}\sum_{k\in\{1,\ldots,K\}} M_k$ 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 (18b), the strict inequality

$$\alpha_{\tilde{k}} < \alpha_{\tilde{k}+1} \tag{27}$$

must hold. Choose

$$\bar{\alpha}_k = \alpha_k, \qquad k \in \mathcal{K},\ k \notin \{\tilde{k}, \tilde{k}+1\}, \tag{28}$$

$$\bar{\alpha}_{\tilde{k}} = \bar{\alpha}_{\tilde{k}+1} = \frac{1}{2}(\alpha_{\tilde{k}} + \alpha_{\tilde{k}+1}), \tag{29}$$

$$\bar{U}_k = U_k, \qquad k \in \mathcal{K},\ k \neq \tilde{k}. \tag{30}$$

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

$$I(U_{\tilde{k}}; Y_{\tilde{k}}|U_{\tilde{k}-1}) \leq I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}}), \tag{31a}$$

or

$$I(U_{\tilde{k}}; Y_{\tilde{k}}|U_{\tilde{k}-1}) > I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}}). \tag{31b}$$

If (31a) holds, choose

$$\bar{U}_{\tilde{k}} = U_{\tilde{k}}. \tag{32}$$

If (31b) holds, let $E \in \{0, 1\}$ be a Bernoulli-β random variable independent of everything else, where

$$\beta := \frac{1}{2} + \frac{1}{2}\frac{I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}})}{I(U_{\tilde{k}}; Y_{\tilde{k}}|U_{\tilde{k}-1})}. \tag{33}$$

Choose

$$\bar{U}_{\tilde{k}} = (U_{\tilde{k}-1+E}, E) \tag{34}$$

The proposed choice satisfies the Markov chain $\bar{U}_1 - \bar{U}_2 - \cdots \bar{U}_K - X$. Moreover, by (34) and (33):

$$\begin{aligned} &I(\bar{U}_{\tilde{k}}; Y_{\tilde{k}}|\bar{U}_{\tilde{k}-1}) \\ &= \frac{1}{2}\left(I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}}) + I(U_{\tilde{k}}; Y_{\tilde{k}}|U_{\tilde{k}-1})\right). \end{aligned} \tag{35}$$

Trivially, for $k \notin \{\tilde{k}, \tilde{k}+1\}$, constraint (19) is unchanged if we replace $(U_1, U_2, \ldots, U_K, X)$ by $(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_K, \bar{X})$ and $(\alpha_1, \ldots, \alpha_K)$ by $(\bar{\alpha}_1, \ldots, \bar{\alpha}_K)$.

If (31a) holds, then the proposed replacement relaxes constraint (19) 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 (31a) holds, the upper bound on $C(M_1, \ldots, M_K)$ in (19) is relaxed if everywhere one replaces $(U_1, U_2, \ldots, U_K, X)$ and $(\alpha_1, \ldots, \alpha_K)$ by $(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_K, \bar{X})$ and $(\bar{\alpha}_1, \ldots, \bar{\alpha}_K)$.

If (31b) holds, then the new constraint obtained for $k = \tilde{k}$ coincides with the average of the two original constraints for $k = \tilde{k}$ and for $k = \tilde{k}+1$, see (29) and (35). 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 (29) and because

$$\begin{aligned} &I(\bar{U}_{\tilde{k}+1}; Y_{\tilde{k}+1}|\bar{U}_{\tilde{k}}) \\ &\overset{(a)}{=} \beta I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}}) + (1-\beta) I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}-1}) \\ &\overset{(b)}{=} \beta I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}}) + (1-\beta) I(U_{\tilde{k}+1}, U_{\tilde{k}}; Y_{\tilde{k}+1}|U_{\tilde{k}-1}) \\ &\overset{(c)}{=} I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}}) + (1-\beta) I(U_{\tilde{k}}; Y_{\tilde{k}+1}|U_{\tilde{k}-1}) \\ &\overset{(d)}{\geq} I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}}) + (1-\beta) I(U_{\tilde{k}}; Y_{\tilde{k}}|U_{\tilde{k}-1}) \\ &\overset{(e)}{=} \frac{1}{2} I(U_{\tilde{k}+1}; Y_{\tilde{k}+1}|U_{\tilde{k}}) + \frac{1}{2} I(U_{\tilde{k}}; Y_{\tilde{k}}|U_{\tilde{k}-1}) \\ &\overset{(f)}{=} I(\bar{U}_{\tilde{k}}; Y_{\tilde{k}}|U_{\tilde{k}-1}), \end{aligned} \tag{36}$$

where (a) follows by the definition of $\bar{U}_{\tilde{k}}$ and $\bar{U}_{\tilde{k}+1}$; (b) by the Markov chain (14); (c) by the chain rule of mutual information; (d) by the degradedness of the channel (14); (e) by the definition of β in (33); and (f) by (35).

We can thus conclude that also when (31b) holds, the upper bound on $C(M_1, \ldots, M_K)$ in (19) is relaxed if one replaces $(U_1, U_2, \ldots, U_K, X)$ and $(\alpha_1, \ldots, \alpha_K)$ by $(\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_K, \bar{X})$ and $(\bar{\alpha}_1, \ldots, \bar{\alpha}_K)$.

References

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

[2] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," *proceedings IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, July, 2015.

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

[4] C.-Y. Wang, S. H. Lim, and M. Gastpar, "A new converse bound for coded caching," *arXiv*, 1601.05690, Jan. 2016.

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

[6] S. Wang, X. Tian and H. Liu, "Exploiting the unexploited of coded caching for wireless content distribution," *proceedings IEEE International Conference on Computing, Networking and Communications (ICNC)*, Feb. 2015.

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

[8] R. Timo and M. Wigger, "Joint cache-channel coding over erasure broadcast channels," *proceedings IEEE International Symposium on Wireless Communications Systems (ISWCS)*, Bruxelles, Belgium, Aug. 2015.

[9] S. Saeedi Bidokhti, M. Wigger and R. Timo, "Erasure Broadcast Networks with Receiver Caching," proceedings *IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain, July 2016.

[10] A. Ghorbel, M. Kobayashi, and S. Yang, "Cache-enabled broadcast packet erasure channels with state feedback," *proceedings Allerton Conference on Communications, Control and Computation*, Monticello (IL), USA, Oct. 2015.

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

[12] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," *proceedings IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, June 2015.

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

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

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