Erasure Broadcast Networks with Receiver Caching

†Shirin Saeedi Bidokhti, Roy Timo‡ and Michèle Wigger†‡
† Stanford University, saeedi@stanford.edu ‡Technische Universität München, roy.timo@tum.de
†‡LTCI CNRS, Telecom ParisTech, michele.wigger@telecom-paristech.fr

***Abstract*—A cache-aided broadcast erasure network is studied with a set of receivers that have access to individual cache memories and a set of receivers that have no cache memory. The erasure statistics of the channel are assumed to be symmetric with respect to all receivers in each set and users with no cache memory are assumed to have better channels, statistically.**

Lower and upper bounds are derived on the capacity of the network. The lower bounds are achieved by joint cache-channel coding schemes and are shown to be significantly larger than the bounds achievable by naive separate cache-channel coding schemes. For the case of two receivers, the capacity is characterized for interesting ranges of cache memory sizes.

I. Introduction and Problem Definition

A. Source & Channel Model

We consider a broadcast channel (BC) with a single transmitter and K receivers as depicted in Figure 1. The channel from the transmitter to the receivers is a memoryless *packet-erasure BC* with input alphabet

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

and common receiver output alphabets

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

Here $F \geq 0$ is a fixed positive integer, and each input symbol $x \in \mathcal{X}$ is an F-bit packet. The output erasure symbol Δ models loss of a packet at a given receiver, for example because of router buffers overload. The marginal transition laws[1] of the memoryless BC are described by

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

Let us assume that the receivers can be partitioned into a set of weak receivers $\mathcal{K}_\text{w} := \{1,\ldots,K_\text{w}\}$ and a set of strong receivers $\mathcal{K}_\text{s} := \{K_\text{w}+1,\ldots,K\}$. Let $K_\text{s} = K - K_\text{w}$ and

$$\begin{aligned} \delta_i &= \delta_\text{w}, \qquad i \in \mathcal{K}_\text{w}, \\ \delta_j &= \delta_\text{s}, \qquad j \in \mathcal{K}_\text{s}, \end{aligned} \tag{2}$$

where

$$0 \leq \delta_\text{s} \leq \delta_\text{w} \leq 1. \tag{3}$$

Let us also assume that the receivers in $\mathcal{K}_\text{w}$ are provided with cache memories of size nM bits, see Figure 1. This problem setup is motivated by [2], which illustrated the benefits of allocating caches to the weak receivers.

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

[1]It turns out that only the marginal transition law is relevant in our problem [18]

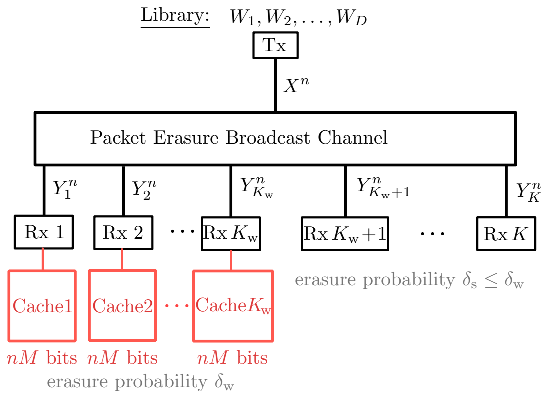

Fig. 1: K user packet-erasure BC with K_w weak receivers that have cache memories.

The transmitter has a library of D independent messages $W_1,\ldots,W_D$, where each W_d is independent and uniformly distributed on $\{1,\ldots,\lfloor 2^{nR} \rfloor\}$. Here, $R \geq 0$ is the rate of each message and n is the blocklength of transmission. We assume $D \geq K$, so there are more messages than receivers. Suppose that each receiver will demand (that is, request and download) exactly one message from the library. We denote the message demanded by receiver k by W_{d_k}, where $d_k \in \{1,\ldots,D\}$. Let

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

denote the receivers' demand vector. The communications process takes place in two phases: a caching phase and a delivery phase.

B. Caching Phase

During a period of low network-congestion and before the receivers' demand vector $\mathbf{d}$ is known, the transmitter sends an individual cache message $V_i \in \{1,\ldots,\lfloor 2^{nM} \rfloor\}$, $i \in \mathcal{K}_\text{w}$, to each of the weak receivers. Since $\mathbf{d}$ is unknown at this time, the cache messages will be functions of the entire library:

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

C. Delivery Phase

Transmitter: The transmitter is given the receivers' demands $\mathbf{d}$, and it sends the corresponding messages $W_{d_1},\ldots,W_{d_K}$ over the BC to the appropriate receivers. We also assume that $\mathbf{d}$ is known to all receivers (e.g., $\mathbf{d}$ can be communicated to the receivers with zero transmission rate). For a given $\mathbf{d}$, let

$$X^n := f_\mathbf{d}(W_1,\ldots,W_D), \quad f_\mathbf{d} \colon \{1,\ldots,\lfloor 2^{nR} \rfloor\}^D \to \mathcal{X}^n,$$

denote the channel input at the transmitter. Each receiver $k \in \{1, \ldots, K\}$ observes Y_k^n according to the memoryless transition law (1).

Weak receivers: Each receiver $i \in \mathcal{K}_\mathrm{w}$ outputs

$$\hat{W}_i := \varphi_{i,\mathbf{d}}(Y_i^n, V_i), \quad \varphi_{i,\mathbf{d}} \colon \mathcal{Y}^n \times \mathcal{Z} \to \big\{1, \ldots, \lfloor 2^{nR} \rfloor\big\}.$$

Strong receivers: Each receiver $j \in \mathcal{K}_\mathrm{s}$ outputs

$$\hat{W}_j := \varphi_{j,\mathbf{d}}(Y_j^n), \quad \varphi_{j,\mathbf{d}} \colon \mathcal{Y}^n \to \big\{1, \ldots, \lfloor 2^{nR} \rfloor\big\}.$$

Communications error: An error is declared for $\mathbf{d}$ if

$$\hat{W}_k \neq W_{d_k} \text{ for some } k \in \{1, \ldots, K\}. \tag{4}$$

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

$$\mathsf{P}_\mathrm{e}^\mathrm{worst} := \max_{\mathbf{d} \in \{1,\ldots,D\}^K} \mathbb{P}\left[\bigcup_{k=1}^{K} \hat{W}_k \neq W_{d_k}\right].$$

A rate-memory pair $(R,\ M)$ is said to be *achievable* if for every $\epsilon > 0$ and sufficiently large blocklength n there exist encoding and decoding functions such that $\mathsf{P}_\mathrm{e}^\mathrm{worst} < \epsilon$. The main focus of this paper is on the *capacity-memory tradeoff.*

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

D. Previous Works and New Contributions

The first information-theoretic work on cache-aided communication systems by Maddah-Ali and Niesen [1] considered the scenario where all receivers have equal cache sizes and the delivery phase takes place over noise-free bit-pipes. This is a special case of our setup when $K_\mathrm{s} = 0$ and $\delta_\mathrm{w} = 0$. They show that a smart design of the cached messages $\{V_i\}$ creates possibilities for sending coded (XOR-ed) data during the delivery phase that can simultaneously serve multiple receivers. This way, the delivery rate is reduced beyond the obvious *local caching gain* of $\frac{M}{D}$ bits, i.e., beyond the rate of the data that each receiver can retrieve from its local cache. The additional reduction is due to the fact that other receivers also have caches, and was thus termed [1] *global caching gain*.

In this work, we present two new coding ideas that permit further global caching gains for erasure BCs.

Our first idea applies to general parameters $K_\mathrm{w}, K_\mathrm{s} \geq 1$ and $D \geq 2$, but requires that $\delta_\mathrm{s} < \delta_\mathrm{w}$. The main novelty is to use a *joint cache-channel coding scheme* for the delivery phase. This means that the contents in the receivers' caches do not only determine *what* to transmit in the delivery phase, but also *how* to transmit. Specifically, we propose a scheme where information that is intended for strong receivers and cached at weak receivers can be freely *piggybacked* on messages sent to the weak receivers. Numerical simulations show that our new joint cache-channel coding scheme significantly outperforms naive separate cache-channel coding schemes.

Our second idea concerns a small network with $K_\mathrm{w} = K_\mathrm{s} = 1$ and $D = 2$, and is interesting also for equal erasure probabilities $\delta_\mathrm{w} = \delta_\mathrm{s}$. The idea is to cache XOR data similar to [1, Appendix] and to deliver uncoded information using our new piggyback coding idea.

We conclude that modelling the delivery phase communication by a noisy channel is important because it facilitates joint cache-channel coding schemes. The importance of a noisy channel model for the delivery phase was also observed in [2]–[11]. For example, [7] and [8] illustrate interesting interplays between feedback or channel state information with the idea of caching, and [9] shows that caches at the transmitter-side allows for load-balancing and interference mitigation in noisy interference networks.

We also present converse results, i.e., upper bounds on the capacity-memory tradeoff. The upper bounds match our obtained lower bounds in the following three cases:

- $K_\mathrm{w} = K_\mathrm{s} = 1$ and $M \leq FD\frac{(1-\delta_\mathrm{s})(\delta_\mathrm{w}-\delta_\mathrm{s})}{(1-\delta_\mathrm{w})+(1-\delta_\mathrm{s})}$;
- $K_\mathrm{w} = K_\mathrm{s} = 1$, $D = 2$ and $\delta_\mathrm{w} = \delta_\mathrm{s}$;
- $K_\mathrm{w} = K_\mathrm{s} = 1$, $D = 2$ and $M \geq F((1-\delta_\mathrm{s}) + (\delta_\mathrm{w} - \delta_\mathrm{s}))$.

Converse results for the Maddah-Ali & Niesen noise-free bit-pipe model were presented in [1], [12]–[15].

II. Main Results

Our main results are in the form of a general lower bound and a general upper bound on the capacity-memory tradeoff $C(M)$. Furthermore, we show the tightness of the bounds in certain regimes of M in the special case of $K_\mathrm{w} = K_\mathrm{s} = 1$.

A. Lower bound

Define $K_\mathrm{w} + 2$ rate-memory pairs $\{(R_t, M_t);\ t = 0, 1, \ldots, K_\mathrm{w} + 1\}$ as follows:

(i)

$$R_0 := F\left(\frac{K_\mathrm{w}}{1-\delta_\mathrm{w}} + \frac{K_\mathrm{s}}{1-\delta_\mathrm{s}}\right)^{-1}, \qquad M_0 := 0; \tag{5}$$

(ii) For each $t \in \{1, \ldots, K_\mathrm{w}\}$:

$$R_t := \frac{F(1-\delta_\mathrm{w})\left(1 + \frac{K_\mathrm{w}-t+1}{tK_\mathrm{s}}\ \frac{\delta_\mathrm{w}-\delta_\mathrm{s}}{1-\delta_\mathrm{w}}\right)}{\frac{K_\mathrm{w}-t+1}{t}\left(1 + \frac{K_\mathrm{w}-t}{(t+1)K_\mathrm{s}}\ \frac{\delta_\mathrm{w}-\delta_\mathrm{s}}{1-\delta_\mathrm{w}}\right) + K_\mathrm{s}\frac{1-\delta_w}{1-\delta_\mathrm{s}}},$$

$$M_t := R_t\ \frac{D}{K_\mathrm{w}}\left(t - \left(1 + \frac{K_\mathrm{w}-t+1}{tK_\mathrm{s}}\ \frac{\delta_\mathrm{w}-\delta_\mathrm{s}}{1-\delta_\mathrm{w}}\right)^{-1}\right); \tag{6}$$

(iii)

$$R_{K_\mathrm{w}+1} := F\frac{1-\delta_\mathrm{s}}{K_\mathrm{s}}, \qquad M_{K_\mathrm{w}+1} := DF\ \frac{1-\delta_\mathrm{s}}{K_\mathrm{s}}. \tag{7}$$

Theorem 1 (Direct Part): The upper convex hull of the $K_\mathrm{w}+2$ rate-memory pairs $\{(R_t, M_t);\ t = 0, 1, \ldots, K_\mathrm{w}+1\}$ in (5)–(7) forms a lower bound on the capacity-memory tradeoff:

$$C(M) \geq \text{upper hull}\big\{(R_t, M_t)\colon\ t = 0, \ldots, K_\mathrm{w}+1\big\}. \tag{8}$$

Proof Outline: The pair $(R_0,\ M_0 = 0)$ corresponds to the case without caches, and the achievability of R_0 follows from the usual capacity result for packet-erasure BCs [16].

The pair $(R_{K_\mathrm{w}+1},\ M_{K_\mathrm{w}+1})$ corresponds to the case where M is large enough so that each receiver in $\mathcal{K}_\mathrm{w}$ can store the entire library in its cache memory. The delivery phase thus only needs to serve receivers in $\mathcal{K}_\text{s}$ and achievability of $(R_{K_\text{w}+1},\ M_{K_\text{w}+1})$ follows again from the usual capacity of packet-erasure BCs (where we only consider strong receivers).

The remaining pairs (R_t, M_t), $t = 1, \ldots, K_\text{w}$, are more interesting and are achieved by the joint cache-channel coding scheme in section III. The upper convex hull of $\{(R_t, M_t); t = 0, 1, \ldots, K_\text{w}+1\}$, finally, is achieved by time-sharing. ■

Remark 1: Consider the following naive separate cache-channel coding scheme. *Step 1:* Apply Maddah-Ali & Niesen coded caching [1, Algorithm 1] to the messages that are demanded by receivers $\mathcal{K}_\text{w}$. *Step 2:* Send the XOR messages from step 1 together with the messages that are demanded by receivers in $\mathcal{K}_\text{s}$ using a capacity-achieving scheme for the packet-erasure BC. This scheme achieves the upper convex hull of the rate-memory pairs $\{(R_{t,\text{sep}}, M_{t,\text{sep}}); t = 0, 1, \ldots, K_\text{w}\}$, where

$$R_{t,\text{sep}} := F\left(\frac{K_\text{w}-t}{(t+1)(1-\delta_\text{w})} + \frac{K_\text{s}}{(1-\delta_\text{s})}\right)^{-1}, \tag{9a}$$

$$M_{t,\text{sep}} := R_{t,\text{sep}}\ t\ \frac{D}{K_\text{w}}. \tag{9b}$$

B. Upper bound

For each pair of integers (k_w, k_s) satisfying

$$0 \le k_\text{w} \le K_\text{w}; \quad 0 \le k_\text{s} \le K_\text{s}; \quad \text{and} \quad 1 \le k_\text{w} + k_\text{s}; \tag{10}$$

define

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

Theorem 2 (Converse): The capacity-memory tradeoff $C(M)$ is upper bounded as

$$C(M) \le \text{lower hull}\left\{R_{k_\text{w},k_\text{s}}(M)\colon\ k_\text{w}, k_\text{s} \text{ satisfying (10)}\right\}.$$

The proof of Theorem 2 is deferred to [18]. Figures 2 and 3 illustrate the upper bound from Theorem 2 together with the two lower bounds on $C(M)$ from Theorem 1 and Remark 1 for two setups.

C. Special case of $K_\text{w} = 1$ and $K_\text{s} = 1$

We evaluate our bounds for a setup with a single weak receiver and a single strong receiver and we assume for simplicity of exposition that D is even. Let

$$\Gamma_1 := F\frac{(1-\delta_\text{s})(\delta_w - \delta_\text{s})}{(1-\delta_w)+(1-\delta_\text{s})}, \tag{11}$$

$$\Gamma_2 := (1-\delta_\text{s})F, \tag{12}$$

$$\Gamma_3 := F\frac{(1-\delta_\text{s})^2}{(1-\delta_w)+(1-\delta_\text{s})}. \tag{13}$$

Notice that $0 \le \Gamma_1 \le \Gamma_3 \le \Gamma_2$. From Theorems 1 and 2 we obtain the following corollary.

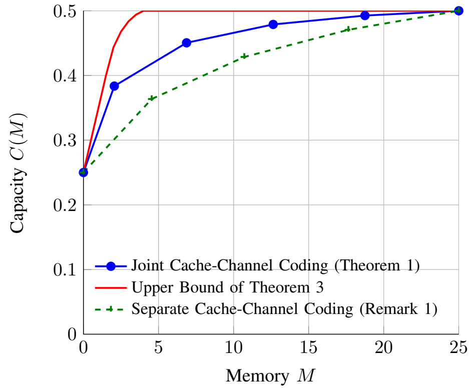

Fig. 2: Bounds on capacity-memory tradeoff $C(M)$ for $K_\text{w} = 4$, $K_\text{s} = 16$, $D = 50$, $\delta_\text{w} = 0.8$, $\delta_\text{s} = 0.2$, and $F = 50$.

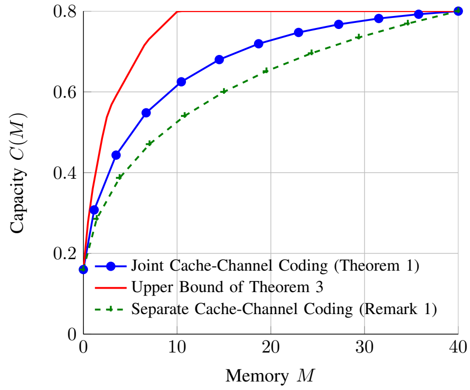

Fig. 3: Bounds on capacity-memory tradeoff $C(M)$ for $K_\text{w} = K_\text{s} = 10$, $D = 50$, $\delta_\text{w} = 0.8$, $\delta_\text{s} = 0.2$, and $F = 50$.

Corollary 2.1: If $K_\text{w} = 1$ and $K_\text{s} = 1$, the capacity-memory tradeoff is lower bounded by:

$$C(M) \ge \begin{cases} F\frac{(1-\delta_\text{w})(1-\delta_\text{s})}{(1-\delta_\text{w})+(1-\delta_\text{s})} + \frac{M}{D}, & \text{if } \frac{M}{D} \in [0, \Gamma_1] \\ F\frac{1}{2}(1-\delta_\text{s}) + \frac{M}{2D}, & \text{if } \frac{M}{D} \in (\Gamma_1, \Gamma_2] \\ F(1-\delta_\text{s}) & \text{if } \frac{M}{D} > \Gamma_2, \end{cases} \tag{14}$$

and upper bounded by:

$$C(M) \le \begin{cases} F\frac{(1-\delta_\text{w})(1-\delta_\text{s})}{(1-\delta_\text{w})+(1-\delta_\text{s})} + \frac{M}{D}, & \text{if } \frac{M}{D} \in [0, \Gamma_3] \\ F(1-\delta_\text{s}) & \text{if } \frac{M}{D} > \Gamma_3. \end{cases} \tag{15}$$

Figure 4 shows these two bounds and the bound in Remark 1 for $D = 10$, $\delta_\text{w} = 0.8$, $\delta_\text{s} = 0.2$, and $F = 10$.

We identify three regimes. In the first regime $0 \le \frac{M}{D} \le \Gamma_1$, the cache memory allows reducing the rate R to each receiver by $\frac{M}{D}$. This is the same performance as when a naive uncoded caching strategy is used in a setup where *both receivers* have cache memories of rate M. The single cache at the weak receiver thus seems to serve all receivers in the network. In this sense, in the first regime, our joint cache-channel coding scheme enables the best possible global caching gain. In the second regime $\Gamma_1 < \frac{M}{D} \le \Gamma_2$ the gains are not as significant as in the first regime, but increasing the cache sizes still

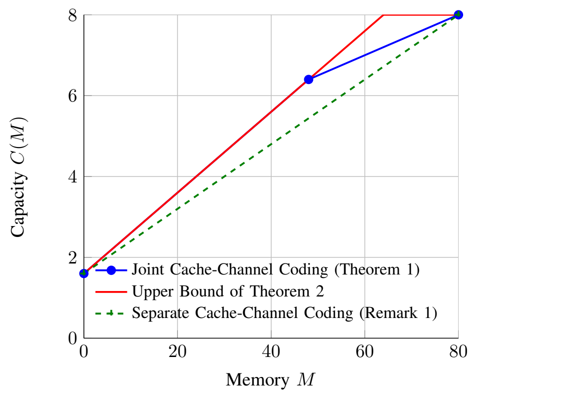

Fig. 4: Bounds on the capacity-memory tradeoff for $K_{\text{w}} = 1$, $K_{\text{s}} = 1$, $D = 10$, $\delta_{\text{w}} = 0.8$, $\delta_{\text{s}} = 0.2$, and $F = 10$.

results in an improved performance. Finally, in the last regime $\frac{M}{D} > \Gamma_2$ the weak receivers have all relevant information in their caches and thus the communication is only restricted by the communication to the stronger receivers. In this regime the performance is independent of the cache sizes.

In the first and last regime, $0 \leq \frac{M}{D} \leq \Gamma_1$ and $\frac{M}{D} > \Gamma_2$, our joint cache-channel coding scheme of section III achieves the capacity-memory tradeoff $C(M)$.

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 in some special cases. Let

$$\tilde{\Gamma}_1 := F\frac{(1-\delta_{\text{w}})^2 + (1-\delta_{\text{s}})^2 - (1-\delta_{\text{w}})(1-\delta_{\text{s}})}{(1-\delta_{\text{w}}) + (1-\delta_{\text{s}})}, \quad (16)$$

$$\tilde{\Gamma}_2 := \frac{1}{2}F\left((1-\delta_{\text{s}}) + (\delta_{\text{w}} - \delta_{\text{s}})\right). \quad (17)$$

Theorem 3: If $K_{\text{w}} = K_{\text{s}} = 1$ and $D = 2$:

$$C(M) \leq \begin{cases} F\frac{(1-\delta_{\text{w}})(1-\delta_{\text{s}})}{(1-\delta_{\text{w}})+(1-\delta_{\text{s}})} + \frac{M}{2}, & \text{if } \frac{M}{2} \in [0, \tilde{\Gamma}_1] \\ F\frac{1}{3}(2-\delta_{\text{s}}-\delta_{\text{w}}) + \frac{M}{3}, & \text{if } \frac{M}{2} \in (\tilde{\Gamma}_1, \tilde{\Gamma}_2] \\ F(1-\delta_{\text{s}}) & \text{if } \frac{M}{2} > \tilde{\Gamma}_2. \end{cases} \quad (18)$$

and lower bounded as:

$$C(M) \geq \begin{cases} F\frac{(1-\delta_{\text{w}})(1-\delta_{\text{s}})}{(1-\delta_{\text{w}})+(1-\delta_{\text{s}})} + \frac{M}{2}, & \frac{M}{2} \in [0, \Gamma_1] \\ \frac{(1-\delta_s)}{3(1-\delta_{\text{s}})-(1-\delta_{\text{w}})}\left(F(1-\delta_{\text{s}}) + M\right), & \frac{M}{2} \in (\Gamma_1, \tilde{\Gamma}_2] \\ F(1-\delta_{\text{s}}) & \frac{M}{2} > \tilde{\Gamma}_2. \end{cases} \quad (19)$$

Proof: Upper bound (18) is proved in [18].

Lower bound (19) coincides with the upper convex hull of the three rate-memory pairs: $(R_0,\ M_0)$ in (5); $(R_1,\ M_1)$ in (6); and $(F(1-\delta_{\text{s}}),\ 2\tilde{\Gamma}_2)$. Achievability of the former two pairs follows from Theorem 1. Achievability of the last pair follows from the following scheme. Store the XOR message $W_1 \bigoplus W_2$[2] and the first $F(\delta_w - \delta_s)$ bits of W_1 in the weak receiver's cache. Use the *piggyback coding* idea described in section III for phase 2 (see also Figure 6 and [2], [17]), to deliver the first $nF(\delta_{\text{w}} - \delta_{\text{s}})$ bits of W_{d_2} to the strong receiver only and the remaining $n(R - F(\delta_{\text{w}} - \delta_{\text{s}}))$ bits of W_{d_2} to both receivers. The weak receiver reconstructs W_{d_1} from its decoded bits of W_{d_2} and the content in its cache. ■

Figure 5 shows the bounds of Theorem 3 for $\delta_{\text{w}} = 0.8$, $\delta_{\text{s}} = 0.2$, and $F = 10$. Upper and lower bounds of Theorem 3

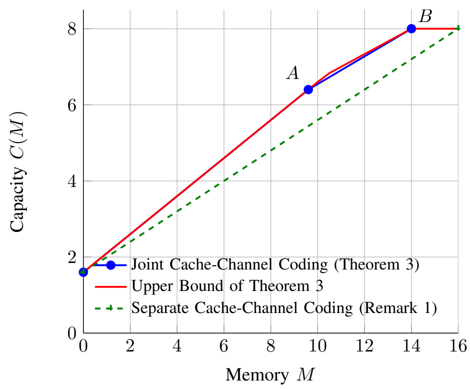

Fig. 5: Bounds on the capacity-memory tradeoff for $K_{\text{w}} = 1$, $K_{\text{s}} = 1$, $D = 2$, $\delta_{\text{w}} = 0.8$, $\delta_{\text{s}} = 0.2$, and $F = 10$.

coincide in the case of equal erasure probabilities $\delta_{\text{w}} = \delta_{\text{s}}$:

Corollary 3.1: If $K_{\text{w}} = K_{\text{s}} = 1$, $D = 2$ and $\delta_{\text{w}} = \delta_{\text{s}} = \delta$:

$$C(M) = \begin{cases} F\frac{1}{2}(1-\delta) + \frac{M}{2}, & \text{if } \frac{M}{2} \in \left[0, \frac{1}{2}F(1-\delta)\right] \\ F(1-\delta) & \text{if } \frac{M}{2} > \frac{1}{2}F(1-\delta). \end{cases} \quad (20)$$

Proof: Follows from Theorem 3 because for $\delta_{\text{w}} = \delta_{\text{s}}$: $\tilde{\Gamma}_1 = \tilde{\Gamma}_2 = \frac{1}{2}F(1-\delta)$, and in the regime $\frac{M}{2} \in (\Gamma_1, \tilde{\Gamma}_2]$ lower bound (19) specialises to $C(M) \geq F\frac{1}{2}(1-\delta) + \frac{M}{2}$. ■

III. Joint Cache-Channel Coding

The new joint cache-channel scheme outlined in this section will extensively use Maddah-Ali and Niesen's scheme from [1, Algorithm 1]. This scheme has two parameters: The number of receivers $\tilde{K}$, and an index $\tilde{t}$ that varies from 0 to $\tilde{K}$. It consists of three methods:

- `Ca`: Split each message W_d into $\binom{\tilde{K}}{\tilde{t}}$ submessages, and store each of them in a different subset of $\tilde{t}$ receivers.
- `En`: The delivery encoder sends XORs of submessages (see footnote 2), where each XOR is a common message intended to a group of $\tilde{t}+1$ receivers.
- `De`: Each receiver k reconstructs the $\binom{\tilde{K}}{\tilde{t}}$ submessages of W_{d_k} by XOR-ing the received XOR messages with appropriate submessages in its cache.

We now describe a scheme, parameterised by $t \in \{1, \ldots, K_{\text{w}}\}$,[3] that achieves rate-memory pairs $\{(R_t,\ M_t)\}$ in (6).

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

$$W_d = (W_d^{(t)},\ W_d^{(t-1)})$$

[2] $\bigoplus$ always means XOR on the binary representations.

[3] For $K_{\text{w}} = K_{\text{s}} = 1$, the scheme is described with more details in [2].

with respective rates $R^{(t)} = R - R^{(t-1)}$ and

$$R^{(t-1)} = R\left(1 + \frac{K_\mathrm{w} - t + 1}{tK_\mathrm{s}} \cdot \frac{\delta_\mathrm{w} - \delta_\mathrm{s}}{1 - \delta_\mathrm{w}}\right)^{-1}. \tag{21}$$

1) Caching phase: Apply the cache placement method `Ca`, with $\tilde{K} = K_\mathrm{w}$ and $\tilde{t} = t$, to cache $W_1^{(t)}, W_2^{(t)}, \ldots, W_D^{(t)}$ at the weak receivers. Reapply method `Ca`, now with $\tilde{t} = t-1$, to cache $W_1^{(t-1)}$, $W_2^{(t-1)}, \ldots, W_D^{(t-1)}$.

2) Delivery phase: Delivery takes place in three phases. *Phase 1:* The transmitter applies method `En`, with $\tilde{K} = K_\mathrm{w}$ and $\tilde{t} = t$, to messages

$$\{W_{d_i}^{(t)} : \; i \in \mathcal{K}_\mathrm{w}\}, \tag{22}$$

which are demanded by the weak receivers. It then uses a capacity-achieving code for the packet-erasure BC to send the produced XORs to their intended weak receivers.

Each weak receiver $i \in \mathcal{K}_\mathrm{w}$ applies a good channel decoder for the BC to decode the XORs and subsequently reconstructs $W_{d_i}^{(t)}$ by means of method `De`.

Phase 2: The transmitter conveys submessages

$$\{W_{d_i}^{(t-1)} : \; i \in \mathcal{K}_\mathrm{w}\}, \tag{23}$$

to the weak receivers, and submessages

$$\{W_{d_j}^{(t)} : \; j \in \mathcal{K}_\mathrm{s}\}, \tag{24}$$

to the strong receivers.

Communication in Phase 2 is split into $\tau := \binom{K_w}{t}$ subphases, where each subphase is associated with a different set

$$\mathcal{S} \subseteq \mathcal{K}_\mathrm{w} \quad \text{with} \quad |\mathcal{S}| = t. \tag{25}$$

The transmitter first applies method `En`, with $\tilde{K} = K_\mathrm{w}$ and $\tilde{t} = t-1$, to messages in (23). Let $W_\mathcal{S}^{(t-1)}$ denote the XOR produced for group $\mathcal{S}$.

In subphase $\mathcal{S}$ the transmitter sends

$$\mathbf{x}\big(W_\mathcal{S}^{(t-1)},\; W_{d_{K_\mathrm{w}+1}}^{(t),\mathcal{S}},\; \ldots,\; W_{d_K}^{(t),\mathcal{S}}\big) \tag{26}$$

where $W_{d_j}^{(t),\mathcal{S}}$, for $j \in \mathcal{K}_\mathrm{s}$, is the submessage of $W_{d_j}^{(t)}$ cached at each receiver in $\mathcal{S}$, and where $\mathbf{x}(\cdots)$ is a codeword from the codebook $\mathcal{C}_\mathcal{S}$ illustrated in Figure 6. Codewords of codebook

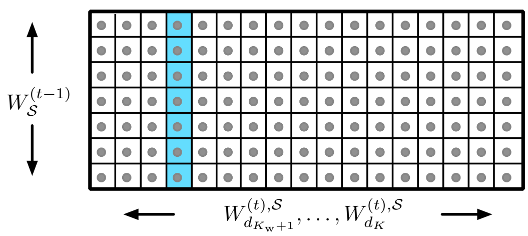

Fig. 6: Codebook $\mathcal{C}_\mathcal{S}$ where dots represent codewords.

$\mathcal{C}_\mathcal{S}$ are arranged in an array: the vertical dimension encodes $W_\mathcal{S}^{(t-1)}$ and the horizontal dimension $(W_{d_{K_\mathrm{w}+1}}^{(t),\mathcal{S}}, \ldots, W_{d_K}^{(t),\mathcal{S}})$.

Each strong receiver decodes the entire tuple $(W_\mathcal{S}^{(t-1)}, W_{d_{K_\mathrm{w}+1}}^{(t),\mathcal{S}}, \ldots, W_{d_K}^{(t),\mathcal{S}})$ for which it has to consider the full codebook $\mathcal{C}_\mathcal{S}$. It then reconstructs its desired message in (24).

Each weak receiver in $\mathcal{S}$ decodes $W_\mathcal{S}^{(t-1)}$ *where it can restrict to the column of codebook* $\mathcal{C}_\mathcal{S}$ that corresponds to the realisations of $W_{d_{K_\mathrm{w}+1}}^{(t)\mathcal{S}}, \ldots, W_{d_K}^{(t),\mathcal{S}}$ stored in its cache.

Weak receivers in $\mathcal{S}$ decode $W_\mathcal{S}^{(t-1)}$ with the same probability of error as if submessages $W_{d_{K_\mathrm{w}+1}}^{(t),\mathcal{S}}, \ldots, W_{d_K}^{(t),\mathcal{S}}$ had not been sent at all. Since we chose the rates $R^{(t-1)}$ and $R^{(t)}$ in (21) so that the decodings at the weak receivers and at the strong receivers yield the same constraints on the rate R, our *piggybacking* of submessages $W_{d_{K_\mathrm{w}+1}}^{(t),\mathcal{S}}, \ldots, W_{d_K}^{(t),\mathcal{S}}$ on the XOR $W_\mathcal{S}^{(t-1)}$ is free (unchanged constraints on R).

As opposed to phases 1 and 3, in this phase 2 *joint source-channel coding* is used: the weak receivers' decodings depend on the content stored in their caches.

Phase 3: The transmitter uses a capacity-achieving code for the packet-erasure BC to send

$$\{W_{d_j}^{(t-1)} : \; j \in \mathcal{K}_\mathrm{s}\},$$

to the strong receivers.

References

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

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

[3] W. Huang, S. Wang, L. Ding, F. Yang, and W. Zhang, "The performance analysis of coded cache in wireless fading channel." Online: `ArXiv:1504.01452`.

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

[5] Y. Ugur, Z. H. Awan and A. Sezgin, "Cloud radio access networks with coded caching." Online: `ArXiv:1512.02385`.

[6] 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.

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

[8] J. Zhang and P. Elia, "Fundamental limits of cache-aided wireless BC: interplay of coded-caching and CSIT feedback." Online: `ArXiv:1511.03961`.

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

[10] 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`

[11] 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.*

[12] Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: Improved bounds for small buffer users." Online: `ArXiv:1407.1935`.

[13] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching." Online: `Arxiv:1501.06003`.

[14] M. Ji, A. M. Tulino, J. Llorca, and G. Caire, "Order-optimal rate of caching and coded multicasting with random demands." Online: `ArXiv:11502.03124`.

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

[16] R. Urbanke and A. Wyner"Packetizing for the erasure broadcast channel with an internet application," 1999. Online: `lthcwww.epfl.ch/~ruediger/papers`.

[17] Y. Wu and M. Wigger, "Coding schemes with rate-limited feedback that improve over the nofeedback capacity for a large class of broadcast channels." Online: `ArXiv.org:1401.6219`.

[18] S. Saeedi Bidokhti, R. Timo and M. Wigger, "Noisy Broadcast Networks with Receiver Caching." Online: `stanford.edu/~saeedi/jrnlcache.pdf`.