Improved Converses and Gap-Results for Coded Caching

†Chien-Yi Wang, ‡Shirin Saeedi Bidokhti, and †Michèle Wigger
†LTCI, Telecom ParisTech, Université Paris-Saclay, 75013 Paris, France, {chien-yi.wang, michele.wigger}@telecom-paristech.fr
‡ Stanford University, saeedi@stanford.edu

***Abstract*—Improved lower bounds on the worst-case and the average-case rate-memory tradeoffs for the Maddah-Ali&Niesen coded-caching scenario are presented. For any number of users and files and for arbitrary cache sizes, the multiplicative gap between the exact rate-memory tradeoff and the new lower bound is less than 2.315 in the worst-case scenario and less than 2.507 in the average-case scenario.**

I. Introduction

We consider the canonical coded-caching scenario by Maddah-Ali and Niesen [1] with a single transmitter and K receivers, where each receiver is equipped with a cache memory of equal size. Communication takes place in two phases. In a first *caching phase* the transmitter stores contents (arbitrary functions of files) at the receivers' cache memories. In a subsequent *delivery phase* it conveys one of the files stored in its library to each of the receivers. The main challenge in this configuration is that during the caching phase it is not known which receiver demands which specific file from the library. The cache contents thus need to be designed so as to be useful for many possible demands.

Traditional caching systems store the same most popular files in each and every cache memory. This allows the receivers to retrieve these files locally without burdening the common communication link from the transmitter to the receivers. Recently [1], it has been shown that much larger gains, so called *global caching gains*, are possible if the various receivers store different parts of the files in their cache memories. In this case, the transmitter can simultaneously serve multiple receivers during the delivery phase by sending coded data, thus reducing the rate on the communication link from the transmitter to the receivers.

The main quantity of interest in this work is the *rate-memory tradeoff*—i.e., the minimum required delivery rate, as a function of the available cache memories, so that all receivers reliably recover their demanded files. Several recent works [1]–[15] have presented upper and lower bounds on the rate-memory tradeoff. The related scenario where delivery communication takes place over a noisy broadcast channel was considered in [16]–[20].

More specifically, the works in [1]–[11], [14], [15] investigated the worst-case rate-memory tradeoff where the communication rate *cannot* depend on the specific demand. The previously best lower and upper bounds for this worst-case rate-memory tradeoff have been shown to match the exact rate-memory tradeoff up to a multiplicative gap of 4 [9]. In this paper we provide a new lower bound and show that it matches the exact rate-memory tradeoff up to a multiplicative gap of 2.315.

The works in [12]–[14] investigated the average rate-memory tradeoff when the communication rate *can* depend on the specific demand, and the rate of interest is the average rate over all possible demands. The previously best lower and upper bounds for this average rate-memory tradeoff have been shown to match up to a multiplicative gap of 4.7 [13]. Here we present a new lower bound and show that it matches the exact rate-memory tradeoff up to a multiplicative gap of 2.507.

II. Detailed Problem Setup

The transmitter has a library of N independent messages $W_1, \ldots, W_N$, where each W_d is uniformly distributed over $\{1, \ldots, 2^F\}$ for F a positive integer. Each of the K receivers is provided with a cache memory of size $F\mathsf{M}$ bits. Suppose that each receiver will demand exactly one message from the library. We denote the demand of Receiver k by $d_k \in \mathcal{N} := \{1, \ldots, N\}$. I.e., Receiver k demands message W_{d_k}.

Communication takes place in two phases: a caching phase and a delivery phase. In the first cache phase, the transmitter sends an individual cache message $\mathbb{V}_k \in \{1, \ldots, \lfloor 2^{F\mathsf{M}} \rfloor\}$, to each of the K receivers. Since the demands $d_1, \cdots, d_K$ are unknown at this time, the cache messages will be functions of the entire library:[1]

$$\mathbb{V}_k := g_k(W_1, \ldots, W_N), \quad k \in \{1, \ldots, K\},$$

where $g_k \colon \{1, \ldots, 2^F\}^N \to \{1, \ldots, \lfloor 2^{F\mathsf{M}} \rfloor\}$.

In the delivery phase, the transmitter is given the demand vector $\mathbf{d} = (d_1, \ldots, d_K)$, and it generates the delivery-symbol

$$X := f_{\mathbf{d}}(W_1, \ldots, W_N), \quad f_{\mathbf{d}} \colon \left\{1, \ldots, 2^F\right\}^N \to \mathcal{X},$$

where $\mathcal{X}$ is the delivery alphabet that we will specify shortly.

We assume that $\mathbf{d}$ is also revealed to all receivers (this communication takes zero rate), and that a common noise-free bit-pipe connects the transmitter to the receivers. Each Receiver $k \in \{1, \ldots, K\}$ thus observes the delivery-symbol X, and can recover its desired message as

$$\hat{W}_k := \varphi_{k,\mathbf{d}}(X, \mathbb{V}_k),$$

[1] Allowing for randomized caching functions does not change the rate-memory tradeoff studied in this paper.

where $\varphi_{k,\mathbf{d}}\colon \mathcal{X} \times \{1,\ldots,\lfloor 2^{F\mathsf{M}}\rfloor\} \to \{1,\ldots,2^F\}$.

We are left with specifying the delivery alphabet $\mathcal{X}$. We distinguish two scenarios:

- In the *worst-case scenario*, the delivery alphabet $\mathcal{X}$ does not depend on the demand vector $\mathbf{d}$.
  In this scenario, a rate-memory pair $(R,\ \mathsf{M})$ is *achievable* if for every $\epsilon > 0$ and sufficiently large message lengths F, there exist caching, encoding, and decoding functions with delivery alphabet
$$\mathcal{X} = \{1,\ldots,\lfloor 2^{F(R+\epsilon)}\rfloor\}, \tag{1}$$
  so that for each demand vector $\mathbf{d} \in \mathcal{N}^K$, each Receiver k can perfectly reconstruct its desired message:
$$\hat{W}_k = W_{d_k}. \tag{2}$$
- In the *average-case scenario*, the delivery alphabet $\mathcal{X}$ depends on the demand vector $\mathbf{d}$.
  In this scenario, a rate-memory pair $(R,\ \mathsf{M})$ is *achievable* if for each demand vector $\mathbf{d} \in \mathcal{N}^K$, each $\epsilon > 0$, and sufficiently large message lengths F, there exist caching, encoding, and decoding functions with delivery alphabet
$$\mathcal{X}_{\mathbf{d}} = \{1,\ldots,\lfloor 2^{FR_{\mathbf{d}}}\rfloor\}, \tag{3}$$
  so that each Receiver $k \in \{1,\ldots,K\}$ can perfectly reconstruct its desired message, (2), and
$$\frac{1}{N^K}\sum_{\mathbf{d}\in\mathcal{N}^K} R_{\mathbf{d}} \le R + \epsilon. \tag{4}$$

The main focus of this paper is on the *rate-memory tradeoffs* of the worst-case and the average-case scenarios:

Definition 1: Given the cache memory size M, we define the *rate-memory tradeoffs* $R^\star_{\text{worst}}(\mathsf{M})$ and $R^\star_{\text{avg}}(\mathsf{M})$ as the infimum of all rates R such that the rate-memory pair $(R,\ \mathsf{M})$ is achievable for the worst-case and the average-case scenarios, respectively.

The problem is interesting for $\mathsf{M} < N$. For $\mathsf{M} \ge N$ each receiver can store all the library in its cache memory and there is no need to transmit anything during the delivery phase. We henceforth assume $\mathsf{M} < N$.

III. Main Results

Define $\bar{N} := \min\{K, N\}$ and $\bar{\mathcal{N}} := \{1,2,\ldots,\bar{N}\}$.

A. Worst-Case Scenario

Theorem 1: For all $\mathsf{M} \in [0, N)$,
$$R^\star_{\text{worst}}(\mathsf{M}) \ge R^{\text{low}}_{\text{worst}}(\mathsf{M}), \tag{5}$$
where
$$R^{\text{low}}_{\text{worst}}(\mathsf{M}) := \max\left\{ \max_{\ell\in\bar{\mathcal{N}}}\left[\ell - \mathsf{M}\frac{\ell^2}{N}\right],\ \max_{\ell\in\bar{\mathcal{N}}}\left[\ell - \mathsf{M}\sum_{j=1}^{\ell}\frac{j}{N-j+1}\right]\right\}.$$

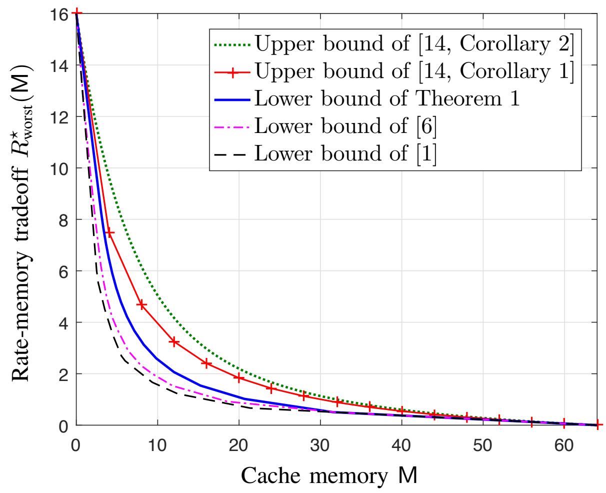

Fig. 1: Upper and lower bounds of $R^\star_{\text{worst}}(\mathsf{M})$ for $K = 16$ and $N = 64$.

Proof: Can be obtained from the lower bound on the capacity-memory tradeoff in degraded noisy broadcast channels with receiver caching in [17]. A direct proof is presented in the extended version [21]. ■

Figure 1 compares this new lower bound on $R^\star_{\text{worst}}(\mathsf{M})$ with the existing lower bounds in [1] and [6]. The figure also shows upper bounds from [14]. The red solid upper bound is for centralized caching, as considered in this paper. The green dashed upper bound is for decentralized caching. For simplicity, the latter upper bound is used to derive the gap results in the following Theorem 2.

The lower bound $R^\star_{\text{worst}}(\mathsf{M})$ matches the exact rate-memory tradeoff up to a constant factor of 2.315, see the following Theorem 2 and Remark 1.

Theorem 2: Irrespective of the number of users K, the library size N, and the memory size $\mathsf{M} \in [0, N)$:
$$\frac{R^\star_{\text{worst}}(\mathsf{M})}{R^{\text{low}}_{\text{worst}}(\mathsf{M})} \le \max_{\ell\in\mathbb{Z}^+}\max_{a\in(0,1)} \phi(a,\ell),$$
where
$$\phi(a,\ell) := \frac{\frac{a(\ell+1)}{(1-a)\ell}\left(1-\left(\frac{\ell+a}{\ell+1}\right)^{\ell/a}\right)}{1-(1-a)\sum_{j=0}^{\ell-1}\frac{1}{\ell-aj}}.$$

Proof: See Section V. ■

Remark 1: For any $\ell \in \mathbb{Z}^+$, the function $a \mapsto \phi(a,\ell)$ is continuous and bounded over $(0,1)$, see also Figure 2. Numerical evaluation[2] shows that for $\ell \in \{1,\ldots,10^4\}$:
$$\max_{a\in(0,1)} \phi(a,\ell) \le 2.315.$$
Moreover,
$$\max_{\ell\in\mathbb{Z}^+:\ \ell>10^4}\ \max_{a\in(0,1)} \phi(a,\ell) \le \max_{b\in(0,10^{-4})}\ \max_{a\in(0,1)} \psi(a,b), \tag{6}$$

[2] All numerical evaluations in this paper are performed by applying the MATLAB function fmincon with the sequential quadratic programming (SQP) method.

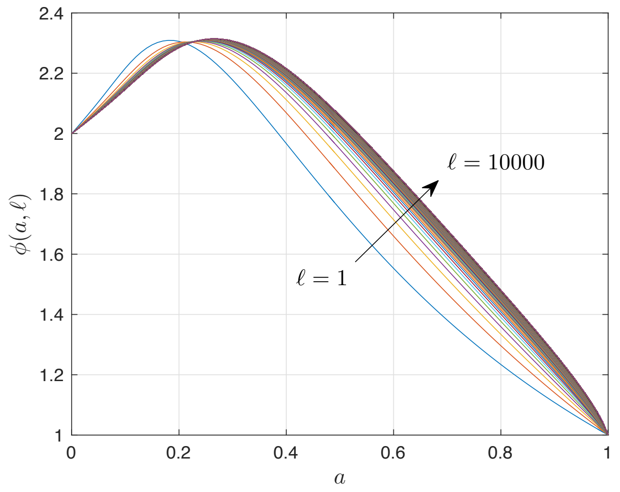

Fig. 2: The functions $\phi(a, \ell)$ for $a \in (0, 1)$ and $\ell = 1, \ldots, 10^4$.

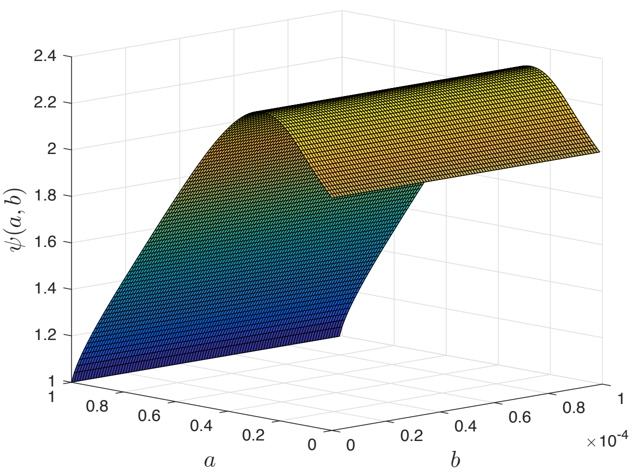

Fig. 3: The function $\psi(a, b)$ over $(a, b) \in (0, 1) \times (0, 10^{-4})$.

where

$$\psi(a, b) := \frac{\frac{a(1+b)}{1-a}\left(1 - \left(\frac{1+ab}{1+b}\right)^{\frac{1}{ab}}\right)}{1 - \frac{(1-a)b}{1-a+ab} + \frac{1-a}{a}\ln(1 - a + ab)}. \tag{7}$$

The function $(a, b) \mapsto \psi(a, b)$ is continuous and bounded over $(0, 1) \times (0, 10^{-4})$, see also Figure 3. Numerical evaluation shows that

$$\max_{b \in (0, 10^{-4})} \max_{a \in (0,1)} \psi(a, b) \leq 2.315. \tag{8}$$

Inequality (6) is derived because the objective function on the right-hand side can be easily plotted over the entire optimization domain, which is not true for the left-hand side.

B. Average-Case Scenario

Theorem 3: For all $\mathsf{M} \in [0, N)$,

$$R^\star_{\text{avg}}(\mathsf{M}) \geq R^{\text{low}}_{\text{avg}}(\mathsf{M}),$$

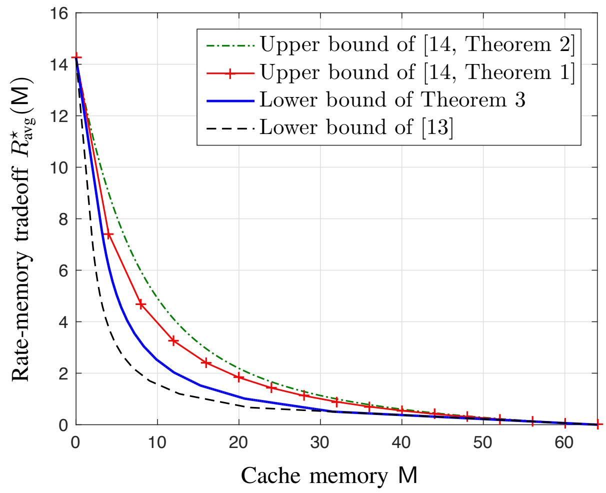

Fig. 4: Upper and lower bounds on $R^\star_{\text{avg}}(\mathsf{M})$ for $K = 16$ and $N = 64$.

where

$$\begin{aligned} &R^{\text{low}}_{\text{avg}}(\mathsf{M}) \\ &:= \max\left\{ \max_{\ell \in \{1,\ldots,K\}} \left[\left(1 - \left(1 - \frac{1}{N}\right)^{\ell}\right)(N - \ell\mathsf{M})\right], \right. \\ &\qquad \left. \max_{\ell \in \{1,\ldots,K\}} \left[\left(1 - \left(1 - \frac{1}{N}\right)^{\ell}\right)N - \frac{\ell(\ell+1)}{2N}\mathsf{M}\right]\right\}. \end{aligned}$$

Proof: See Section IV. ■

Figure 4 compares this new lower bound on $R^\star_{\text{avg}}(\mathsf{M})$ with the existing lower bounds in [13] and the upper bounds in [14]. The red solid upper bound is for centralized caching, as considered in this paper. The green dashed upper bound is for decentralized caching and also from [14].

Lower bound $R^\star_{\text{avg}}(\mathsf{M})$ matches the exact rate-memory tradeoff up to a multiplicative gap of 2.507, see the following Theorem 4 and Remark 2.

Theorem 4: Irrespective of the number of users K, the library size N, and the memory size $\mathsf{M} \in [0, N)$:

$$\frac{R^\star_{\text{avg}}(\mathsf{M})}{R^{\text{low}}_{\text{avg}}(\mathsf{M})} \leq \max_{u \in (0,1]} \max_{v \in (0,1/2]} \eta(u, v),$$

where

$$\begin{aligned} &\eta(u, v) \\ &:= \frac{\left(u + v - v(1-v)^{\frac{u}{v}}\right)\left(1 - \left(1 - \frac{v}{u+v}(1-v)^{\frac{u}{v}}\right)^{\frac{1}{v}}\right)}{(1-v)^{\frac{u}{v}}\left(1 - \left(1 + \frac{u}{2}\right)(1-v)^{\frac{u}{v}}\right)}. \end{aligned}$$

Proof: Omitted. See [21]. ■

Remark 2: The function η is continuous and bounded over $(0, 1] \times (0, 1/2]$. Numerical evaluation shows that

$$\max_{u \in (0,1]} \max_{v \in (0,1/2]} \eta(u, v) \leq 2.507. \tag{9}$$

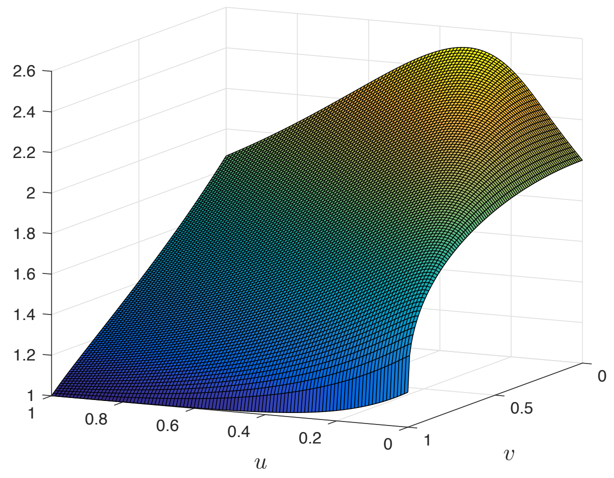

Fig. 5: The function $\eta(u,v)$ over $(u,v) \in (0,1) \times (0,1)$.

IV. Proof of Theorem 3

Lemma 5: Fix $\ell \in \bar{\mathcal{N}}$ and a demand vector $\mathbf{d} \in \mathcal{N}^K$ whose first ℓ entries are $d_1, \ldots, d_\ell$. Fix also $\epsilon > 0$ and assume a sufficiently large F with caching, encoding, and decoding functions so that (2) holds for all $k \in \{1, \ldots, K\}$. Then,

$$R_{\mathbf{d}}+\epsilon \geq \kappa_{\mathbf{d}}(\ell)-\frac{1}{F}\sum_{k=1}^{\ell} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, \ldots, W_{d_{k-1}}), \tag{10}$$

where $\kappa_{\mathbf{d}}(\ell)$ denotes the number of distinct demands at Receivers $1, \ldots, \ell$:

$$\kappa_{\mathbf{d}}(\ell) := |\{d_1, \ldots, d_\ell\}|.$$

Proof: Omitted. See [21]. ■

For any $\ell \in \{1, \ldots, K\}$, let $\mathcal{Q}_\ell^{\text{rep}}$ be the set of all ordered length-ℓ vectors $(d_1, \ldots, d_\ell) \in \mathcal{N}^\ell$, where repetitions are allowed. In the average-case scenario, the random demand vector $\mathbf{D} := (D_1, \ldots, D_K)$ is uniform over $\mathcal{Q}_K^{\text{rep}}$.

Fix now an $\ell \in \{1, \ldots, K\}$, and average Inequality (10) over all demand vectors $\mathbf{d} \in \mathcal{Q}_K^{\text{rep}}$. This yields:

$$\mathbb{E}_{\mathbf{D}}[R_{\mathbf{D}} + \epsilon] \geq \mathbb{E}_{\mathbf{D}}[\kappa_{\mathbf{D}}(\ell)] - \sum_{k=1}^{\ell} \beta_k, \tag{11}$$

where

$$\beta_1 := \frac{1}{F} I(W_{D_1}; \mathbb{V}_1 | \mathbf{D}), \tag{12a}$$

and for $k = 2, \ldots, \ell$:

$$\beta_k := \frac{1}{F} I(W_{D_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{D_1}, \ldots, W_{D_{k-1}}, \mathbf{D}). \tag{12b}$$

The following two lemmas and letting $\epsilon \to 0$ and thus $F \to \infty$, conclude the proof.

Lemma 6:

$$\mathbb{E}_{\mathbf{D}}\big[\kappa_{\mathbf{D}}(\ell)\big] = N\Big(1 - \Big(1 - \frac{1}{N}\Big)^\ell\Big). \tag{13}$$

Proof: Omitted. See [21]. ■

Lemma 7: Parameters $\beta_1, \ldots, \beta_\ell$ satisfy

$$\sum_{k=1}^{\ell} \beta_k \leq \min\Big\{\mathbb{E}_{\mathbf{D}}\big[\kappa_{\mathbf{D}}(\ell)\big] \cdot \frac{\ell \mathsf{M}}{N}, \ \sum_{k=1}^{\ell} \frac{k\mathsf{M}}{N}\Big\}. \tag{14}$$

Proof: We only prove

$$\sum_{k=1}^{\ell} \beta_k \leq \sum_{k=1}^{\ell} \frac{k\mathsf{M}}{N}. \tag{15}$$

See [21] for the proof of

$$\sum_{k=1}^{\ell} \beta_k \leq \mathbb{E}_{\mathbf{D}}\big[\kappa_{\mathbf{D}}(\ell)\big] \cdot \frac{\ell \mathsf{M}}{N}. \tag{16}$$

To prove (15), we show that for each $k \in \{1, \ldots, \ell\}$:

$$\beta_k \leq \frac{k\mathsf{M}}{N}. \tag{17}$$

In fact, defining $\mathbf{D}_k := (D_1, \ldots, D_k)$, we have:

$$\begin{aligned}
F\beta_k &= I(W_{D_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{D_1}, \ldots, W_{D_{k-1}}, \mathbf{D}) \\
&= I(W_{D_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{D_1}, \ldots, W_{D_{k-1}}, \mathbf{D}_k) \\
&= \frac{1}{N^k} \sum_{\mathbf{d} \in \mathcal{Q}_k^{\text{rep}}} I(W_{d_k}; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{d_1}, \ldots, W_{d_{k-1}}) \\
&= \frac{1}{N^k} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\text{rep}}} \sum_{j=1}^{N} I(W_j; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}}) \\
&\overset{(a)}{\leq} \frac{1}{N^k} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\text{rep}}} I(W_1, \ldots, W_N; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}}) \\
&\overset{(b)}{\leq} \frac{1}{N^k} \sum_{\tilde{\mathbf{d}} \in \mathcal{Q}_{k-1}^{\text{rep}}} kF\mathsf{M} \\
&= \frac{kF\mathsf{M}}{N},
\end{aligned}$$

where (a) holds because the messages are independent and because $H(A_1, \ldots, A_L) \leq \sum_{l=1}^{L} H(A_l)$ for any random L-tuple $(A_1, \ldots, A_L)$; and (b) holds because $I(W_1, \ldots, W_N; \mathbb{V}_1, \ldots, \mathbb{V}_k | W_{\tilde{\mathbf{d}}})$ cannot exceed $kF\mathsf{M}$. ■

V. Proof of Theorem 2

We wish to uniformly bound the gap

$$\xi(K, N, \mathsf{M}) := \frac{R^\star_{\text{worst}}(\mathsf{M})}{R^{\text{low}}_{\text{worst}}(\mathsf{M})}, \tag{18}$$

irrespective of $K, N \geq 1$ and $\mathsf{M} \in [0, N)$.

Recall the achievable rate-memory tradeoff from [14, Corollary 2]. For any pair of positive integers $K, N \geq 1$, define

$$R_{\mathsf{YMA}}(K, N, \mathsf{M}) := \begin{cases} \bar{N} & \text{if } \mathsf{M} = 0, \\ \frac{N-\mathsf{M}}{\mathsf{M}}\Big(1 - \big(1 - \frac{\mathsf{M}}{N}\big)^{\bar{N}}\Big) & \text{if } \mathsf{M} \in (0, N). \end{cases}$$

Since $R_{\mathsf{YMA}}(K, N, \mathsf{M})$ upper bounds the worst-case rate-memory tradeoff under a *decentralized caching* assumption

[14], it must also upper bound the rate-memory tradeoff under *centralized caching* as considered here. (In fact, decentralized caching imposes additional constraints on the caching functions g_k compared to our setup here.) Thus, for all K, N:

$$R^{\star}_{\text{worst}}(\mathsf{M}) \leq R_{\mathsf{YMA}}(K,N,\mathsf{M}), \quad \mathsf{M} \in [0,N). \tag{19}$$

We have

$$\xi(K,N,\mathsf{M}) \leq \frac{R_{\mathsf{YMA}}(K,N,\mathsf{M})}{R^{\text{low}}_{\text{worst}}(\mathsf{M})} \leq \frac{R_{\mathsf{YMA}}(K,N,\mathsf{M})}{\underline{R}^{\text{low}}_{\text{worst}}(K,N,\mathsf{M})}, \tag{20}$$

where

$$\underline{R}^{\text{low}}_{\text{worst}}(K,N,\mathsf{M}) := \max_{\ell \in \bar{\mathcal{N}}} \sum_{j=1}^{\ell} \left(1 - \frac{j\mathsf{M}}{N-j+1}\right) \tag{21}$$

and where the second inequality holds because $R^{\text{low}}_{\text{worst}}(\mathsf{M}) \geq \underline{R}^{\text{low}}_{\text{worst}}(K,N,\mathsf{M})$ for all K,N,M.

Define

$$\mathsf{M}_\ell := \begin{cases} \frac{N-\ell}{\ell+1} & \text{if } \ell \in \{0,1,\ldots,\bar{N}-1\}, \\ 0 & \text{if } \ell = \bar{N}. \end{cases}$$

The function $\underline{R}^{\text{low}}_{\text{worst}}(K,N,\mathsf{M})$ is piecewise-linear with $\bar{N}$ line segments over the intervals

$$[\mathsf{M}_{\ell+1}, \mathsf{M}_\ell], \qquad \ell \in \{1,\ldots,\bar{N}-1\} \tag{22a}$$

$$[\mathsf{M}_1, \mathsf{M}_0), \tag{22b}$$

where the last interval is half-open.

We next upper bound $R_{\mathsf{YMA}}(K,N,\mathsf{M})$ by a function $\overline{R}_{\mathsf{YMA}}(K,N,\mathsf{M})$ that is piecewise-linear over the same intervals (22). For every $\ell \in \bar{\mathcal{N}}$ and $\mathsf{M} \in [\mathsf{M}_\ell, \mathsf{M}_{\ell-1})$, define

$$\begin{aligned}\overline{R}_{\mathsf{YMA}}(K,N,\mathsf{M}) :=& \frac{\mathsf{M}-\mathsf{M}_\ell}{\mathsf{M}_{\ell-1}-\mathsf{M}_\ell} \cdot R_{\mathsf{YMA}}(K,N,\mathsf{M}_\ell) \\ &+ \frac{\mathsf{M}_{\ell-1}-\mathsf{M}}{\mathsf{M}_{\ell-1}-\mathsf{M}_\ell} \cdot R_{\mathsf{YMA}}(K,N,\mathsf{M}_{\ell-1}).\end{aligned}$$

Notice that since $R_{\mathsf{YMA}}(K,N,\mathsf{M})$ is convex,

$$\overline{R}_{\mathsf{YMA}}(K,N,\mathsf{M}) \geq R_{\mathsf{YMA}}(K,N,\mathsf{M}), \qquad \mathsf{M} \in [0,N). \tag{23}$$

Plugging (23) into (20), we obtain:

$$\xi(K,N,\mathsf{M}) \leq \frac{\overline{R}_{\mathsf{YMA}}(K,N,\mathsf{M})}{\underline{R}^{\text{low}}_{\text{worst}}(K,N,\mathsf{M})} =: \Xi(K,N,\mathsf{M}). \tag{24}$$

Now, since the upper bound $\Xi(K,N,\mathsf{M})$ is continuous and bounded in $\mathsf{M} \in [0,N)$, and because it is quasiconvex [22] in M over each of the $\bar{N}$ intervals (22),[3] the maximum of $\Xi(K,N,\mathsf{M})$ over each interval is attained at the boundary. Moreover, since $\Xi(K,N,\mathsf{M})$ is constant over the last interval $[\mathsf{M}_1, \mathsf{M}_0)$, we obtain that

$$\max_{\mathsf{M}\in[0,N)} \xi(K,N,\mathsf{M}) \leq \max_{\ell\in\bar{\mathcal{N}}} \Xi(K,N,\mathsf{M}_\ell). \tag{25}$$

Notice that irrespective of $K, N \in \mathbb{Z}^+$:

$$\Xi(K,N,\mathsf{M}_{\bar{N}}) = \frac{\bar{N}}{\bar{N}} = 1. \tag{26}$$

[3] A linear-fractional function is always quasiconvex.

When $\bar{N} = 1$ (i.e., only one file or only one user), Inequalities (25) and (26) imply that $\xi(K,N,\mathsf{M}) = 1$ for all $\mathsf{M} \in [0,N)$, and hence our lower bound is exact. When $\bar{N} \geq 2$, then defining $a := \ell/N$, it can be shown [21] that for all $\ell \in \{1,\ldots,\bar{N}-1\}$:

$$\Xi(K,N,\mathsf{M}_\ell) \leq \frac{\frac{a(\ell+1)}{(1-a)\ell}\left(1-\left(\frac{\ell+a}{\ell+1}\right)^{\ell/a}\right)}{1-(1-a)\sum_{j=0}^{\ell-1}\frac{1}{\ell-aj}} =: \phi(a,\ell).$$

Since $a \in [\frac{1}{N}, 1)$, combined with (26) this concludes the proof.

References

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

[2] Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: Improved bounds for small buffer users," in *arXiv:1407.1935v2[cs.IT]*, Nov. 2015.

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

[4] 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.

[5] 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.

[6] 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.

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

[8] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," in *arXiv:1501.06003v2[cs.IT]*, Feb. 2016.

[9] 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.

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

[11] 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.

[12] 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.

[13] 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.

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

[15] J. Gómez-Vilardebó, "Fundamental limits of caching: improved bounds with coded prefetching," in *arXiv:1612.09071v2[cs.IT]*, Jan. 2017.

[16] S. Saeedi Bidokhti, M. Wigger, and R. Timo, "Noisy broadcast networks with receiver caching," in *arXiv:1605.02317[cs.IT]*, May 2016.

[17] S. Saeedi Bidokhti, M. Wigger, and A. Yener, "Benefits of cache-assignment on noisy broadcast channels," submitted to *ISIT'2017*.

[18] P. Hassanzadeh, E. Erkip, J. Llorca,and A. Tulino, "Distortion-memory tradeoffs in Cache-aided wireless video delivery," in *in Proc. 53rd Annu. Allerton Conference,* Monticello, IL, Sep.-Oct. 2015.

[19] A. Ghorbel, M. Kobayashi, and S. Yang, "Cache-enabled broadcast packet erasure channels with state feedback," in *in Proc. 53rd Annu. Allerton Conference,* Monticello, IL, Sep.-Oct. 2015.

[20] J. Zhang, F. Engelmann, and P. Elia, "Coded caching for reducing CSIT-feedback in wireless communications," in *in Proc. 53rd Annu. Allerton Conference,* Monticello, IL, Sep.-Oct. 2015.

[21] C.-Y. Wang, S. Saeedi Bidokhti, and M. Wigger, "Improved converses and gap-results for coded caching," to be posted on *arXiv*.

[22] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004.