

A New Converse Bound for Coded Caching

Chien-Yi Wang, Sung Hoon Lim, and Michael Gastpar

School of Computer and Communication Sciences
EPFL

Lausanne, Switzerland

{chien-yi.wang, sung.lim, michael.gastpar}@epfl.ch

Abstract—An information-theoretic lower bound is developed for the caching system studied by Maddah-Ali and Niesen. By comparing the proposed lower bound with the decentralized coded caching scheme of Maddah-Ali and Niesen, the optimal memory–rate tradeoff is characterized to within a multiplicative gap of 4.7 for the worst case, improving the previous analytical gap of 12. Furthermore, for the case when users’ requests follow the uniform distribution, the multiplicative gap is tightened to 4.7, improving the previous analytical gap of 72. As an independent result of interest, for the single-user average case in which the user requests multiple files, it is proved that caching the most requested files is optimal.

I. INTRODUCTION

Recently, Maddah-Ali and Niesen considered the following problem setup of *caching*: A file server has access to a database of N files. There are K users, each equipped with an individual cache of the same size. Each user wishes to retrieve one of the N files. During the placement phase, some information is stored in the users’ caches. During the delivery phase, the server sends updates through a shared link so that each user can recover the desired file from the cache content and the received update message. The goal is to design the cache contents and the update such that given a fixed cache memory, the update rate is minimized. For a given cache size, the maximum update rate required over all possible requests (the worst case) was studied in [1]. For the case when the users’ requests follow some probability distribution, the averaged update rate (the average case) was studied in [2].

For the worst case, Maddah-Ali and Niesen gave an analytical characterization of the optimal memory–rate tradeoff to within a multiplicative gap of 12 [1, Theorem 3]. For the average case in which the requests are distributed independently and uniformly (the uniform case), they also gave an analytical characterization of the optimal memory–rate tradeoff to within a multiplicative gap of 72 [2, Claim 1]. Later on, Zhang, Lin, and Wang [3] improved their arguments and gave a universal multiplicative-plus-additive gap ($87R^*+2$) for the general average case. Other improved converse bounds can be found in [4]–[6].

In this work, we propose an information-theoretic lower bound for the average case. By comparing with the achievable memory–rate tradeoff in [7, Theorem 1], we tighten the multiplicative gap from 12 to 4.7 for the worst case (Theorem 3). Moreover, for the uniform case we tighten the multiplicative gap from 72 to 4.7 (Theorem 4). As an essential lemma for

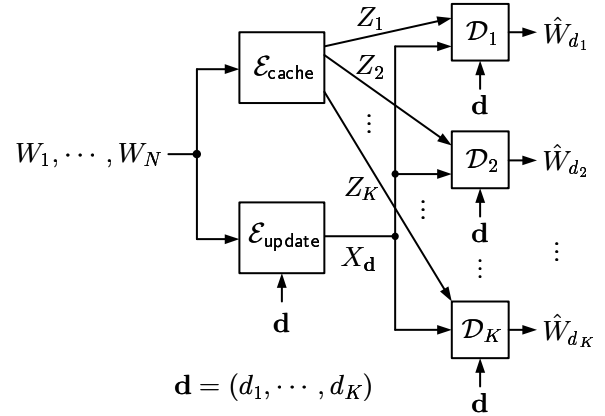


Fig. 1. The K -user caching problem with a database of N files.

the developed lower bound, and as an independent result of interest, we prove that caching the most requested files is an optimal caching strategy for the single-user average case where the user may request *multiple* files, i.e., any subset of the N files (Theorem 6).

Notation: We use calligraphic symbols (e.g., \mathcal{X}) to denote sets. Denote by $|\cdot|$ the cardinality of a set. Random variables and their realizations are represented by uppercase letters (e.g., X) and lowercase letters (e.g., x), respectively. The probability distribution of a random variable X is denoted by p_X . We say that $X \text{---} Y \text{---} Z$ form a Markov chain if $p_{X,Y,Z} = p_X p_{Y|X} p_{Z|Y}$.

We denote $x^+ := \max\{x, 0\}$ for all $x \in \mathbb{R}$ and $[a] := \{1, 2, \dots, a\}$ for all $a \in \mathbb{N}$. Also, we denote $[0 : a] = \{0\} \cup [a]$. Given any sequence or tuple (x_1, x_2, \dots, x_k) and any subset $\mathcal{J} \subset [k]$, we use two short-hand notations $x^{\mathcal{J}}$ and $x_{\mathcal{J}}$ for the subsequence $(x_i : i \in \mathcal{J})$. For the case $\mathcal{J} = [k]$, $x^{[k]}$ is simply denoted by x^k or by \mathbf{x} .

II. PROBLEM STATEMENT

Denote by N the number of files and by K the number of users. Let W_1, W_2, \dots, W_N be N random variables independently and uniformly drawn from $[2^F]$, where F is a positive integer. Each W_n represents a file of size F bits. On the other hand, we denote by d_k the request of user $k \in [K]$. For notational convenience, we denote $W = (W_1, \dots, W_N)$, $\mathbf{d} = (d_1, \dots, d_K)$, and $\mathbf{R} = (R_{\mathbf{d}} : \mathbf{d} \in [N]^K)$. The K -user caching system is depicted in Figure 1.

An (M, \mathbf{R}) caching scheme consists of

- one cache encoder, which assigns K indices $z_k(w) \in [2^{\lfloor FM \rfloor}]$, where $k \in [K]$, to each tuple $w \in [2^F]^N$;
- one update encoder, which assigns an index $x_{\mathbf{d}}(w) \in [2^{\lfloor FR_{\mathbf{d}} \rfloor}]$ to each tuple $(w, \mathbf{d}) \in [2^F]^N \times [N]^K$;
- K decoders, where decoder $k \in [K]$ maps the received messages and the requests, i.e., $(z_k, x_{\mathbf{d}}, \mathbf{d})$, into an estimate $\hat{w}_{d_k} \in [2^F]$.

During the placement phase, the cache encoder maps the files W_1, \dots, W_N into the cache contents Z_1, \dots, Z_K and places Z_k in the cache of user $k \in [K]$. Then, during the delivery phase, the update encoder broadcasts the update message $X_{\mathbf{d}}$ to all users through the shared link. Finally, each user $k \in [K]$ recovers the desired file \hat{W}_{d_k} from the received messages $(Z_k, X_{\mathbf{d}})$ and the requests \mathbf{d} . The probability of error is defined as

$$P_e := \max_{\mathbf{d} \in [N]^K} \max_{k \in [K]} \mathbb{P}(\hat{W}_{d_k} \neq W_{d_k}).$$

We say that a rate tuple (M, \mathbf{R}) is *achievable* if for every $\epsilon > 0$, there exists an (M, \mathbf{R}) caching scheme with large enough file size F such that $P_e < \epsilon$. The optimal rate region \mathcal{R}^* is the closure of the set of achievable rate tuples. Given a fixed memory size $M \geq 0$, we restrict attention to the following projections of \mathbf{R} . The first projection is the maximum of update rates (thus the worst case):

$$R_{\text{worst}}^*(M) := \min_{\mathbf{R}: (M, \mathbf{R}) \in \mathcal{R}^*} \max_{\mathbf{d} \in [N]^K} R_{\mathbf{d}}. \quad (1)$$

Denote by $p_{\mathbf{D}}$ a probability distribution of users' requests. The second projection is the weighted sum of update rates with weights $p_{\mathbf{D}}$ (thus the average case):

$$R_{\text{avg}}^*(M) := \min_{\mathbf{R}: (M, \mathbf{R}) \in \mathcal{R}^*} \sum_{\mathbf{d} \in [N]^K} p_{\mathbf{D}}(\mathbf{d}) R_{\mathbf{d}}. \quad (2)$$

In this work we assume that the requests D_1, \dots, D_K are i.i.d. drawn from the distribution p_D , i.e., $p_{\mathbf{D}}(\mathbf{d}) = \prod_{k=1}^K p_D(d_k)$. When we specialize (2) to the uniform case, i.e., $p_D(d) = \frac{1}{N}$ for all $d \in [N]$, we denote the corresponding optimal memory–rate tradeoff by $R_{\text{uniform}}^*(M)$. Note that the uniform case models the scenario where the requests of different users are independent and the files are equally popular. Clearly, it holds that $R_{\text{uniform}}^*(M) \leq R_{\text{worst}}^*(M)$ for all $M \geq 0$.

III. MAIN RESULTS

Our first contribution is the following closed-form converse bound for the average case. Without loss of generality, we assume that $p_D(1) \geq p_D(2) \geq \dots \geq p_D(N)$.

Theorem 1: Consider the caching problem for the average case with request distribution p_D . For all $M \geq 0$,

$$R_{\text{avg}}^*(M) \geq \max_{k \in [K]} \sum_{n=1}^N (s_n(k) - s_{n+1}(k))(n - kM)^+,$$

where $s_{N+1}(k) = 0$ and $s_n(k) = 1 - (1 - p_D(n))^k$, $n \in [N]$.

The proof of this theorem is deferred to Section IV. By setting $p_D(n) = \frac{1}{N}$, $n \in [N]$, in Theorem 1, we have the converse

bound for the uniform case which also serves as a converse bound for the worst case.

Corollary 1: Consider the caching problem for the uniform case. For all $M \geq 0$,

$$R_{\text{uniform}}^*(M) \geq \max_{k \in [K]} (1 - (1 - 1/N)^k) (N - kM)^+. \quad (3)$$

We will compare the proposed lower bound in Corollary 1 with the decentralized coded caching scheme of Maddah-Ali and Niesen [7]. The achievable memory–rate tradeoff is stated in the following theorem.

Theorem 2 (Maddah-Ali–Niesen [7]): Consider the caching problem for the worst case. It holds that $R_{\text{worst}}^*(0) \leq \min\{K, N\}$, and for all $M \in (0, N]$,

$$\begin{aligned} R_{\text{worst}}^*(M) &\leq (N - M) \cdot \min \left\{ \frac{1}{M} \left(1 - \left(1 - \frac{M}{N} \right)^K \right), 1 \right\} \\ &=: R_{\text{MN}}(K, N, M). \end{aligned} \quad (4)$$

The achievable memory–rate tradeoff described in (4) is not convex. Thus, by time sharing among the achievable points, the achievable memory–rate tradeoff can be improved. We denote by $\check{R}_{\text{MN}}(K, N, M)$ the corresponding convexified bound.

Our second contribution is in showing that given a fixed memory size $M \geq 0$, the decentralized coded caching with time sharing achieves an update rate to within a constant multiplicative factor from the optimum memory–rate tradeoff $R_{\text{worst}}^*(M)$. The main result is the following theorem.

Theorem 3: For all $N \geq 1$, $K \geq 1$, and $M \in [0, N]$,

$$\frac{\check{R}_{\text{MN}}(K, N, M)}{R_{\text{worst}}^*(M)} < 4.7.$$

Since our information-theoretic converse bound is developed for the general average case, we can also establish the following stronger claim.

Theorem 4: For all $N \geq 1$, $K \geq 1$, and $M \in [0, N]$,

$$\frac{\check{R}_{\text{MN}}(K, N, M)}{R_{\text{uniform}}^*(M)} < 4.7.$$

Remark 1: In [1], Maddah-Ali and Niesen proposed the following lower bound for the worst case:

$$R_{\text{worst}}^*(M) \geq \max_{k \in \{\min\{K, N\}\}} \left(k - \frac{k}{\lfloor N/k \rfloor} M \right)^+, \quad (5)$$

for all $M \geq 0$. Numerical evaluation reveals that not surprisingly, the lower bound (5) (which applies only to the worst case) is often tighter than the proposed lower bound (3) (which applies both to the worst and to the uniform case). Nevertheless, there exist cases where (3) is tighter than (5), e.g., when $N = K = 5$ and $M = 1$.

IV. CONVERSE BOUND FOR THE AVERAGE CASE

In this section, we present the average case converse bound. When attributing a distribution $p_{\mathbf{D}}$ on the requests, we further assume that W and \mathbf{D} are independent.

For $n \in [N]$, we denote by $(B_{n1}, B_{n2}, \dots, B_{nF})$ the binary representation of W_n . Since W_n is uniformly distributed over $[2^F]$, $B_{n[F]}$ are i.i.d. Bernoulli(1/2) random variables. For notational convenience, we denote $B = (B_1, \dots, B_N)$, where the entries are i.i.d. Bernoulli(1/2) random variables. Then, we have the following converse bound.

Theorem 5: Consider the caching problem for the average case with request distribution $p_{\mathbf{D}}$. For all $M \geq 0$,

$$R_{\text{avg}}^*(M) \geq \min_{\mathcal{A} \subseteq [K]} \max H(B_{D_{\mathcal{A}}}|V_{\mathcal{A}}, \mathbf{D}),$$

where $B_{D_{\mathcal{A}}} = (B_{D_k} : k \in \mathcal{A})$ and the minimum is over all conditional pmfs $p_{V_{[K]}|B}$ such that $V_{[K]} \text{---} B \text{---} \mathbf{D}$ form a Markov chain and

$$I(B; V_{\mathcal{A}}) \leq |\mathcal{A}|M,$$

for all subsets $\mathcal{A} \subseteq [K]$.

Proof: Consider any subset $\mathcal{A} \subseteq [K]$. Recall that B^F is the binary representation of W . Denote $V_{ki} = (Z_k, B^{i-1})$, $k \in [K]$, $i \in [F]$. Since W and \mathbf{D} are independent by assumption, the Markov chain $V_{[K]i} \text{---} B_i \text{---} \mathbf{D}$ holds for all $i \in [F]$. Then, since $H(Z_k) \leq FM$ for all $k \in [K]$, we have

$$\begin{aligned} |\mathcal{A}|FM &\geq \sum_{k \in \mathcal{A}} H(Z_k) \\ &\geq H(Z_{\mathcal{A}}) \\ &= I(W; Z_{\mathcal{A}}) \\ &= I(B^F; Z_{\mathcal{A}}) \\ &= \sum_{i=1}^F I(B_i; Z_{\mathcal{A}}|B^{i-1}) \\ &= \sum_{i=1}^F I(B_i; V_{\mathcal{A}i}). \end{aligned}$$

Next, we have

$$\begin{aligned} &F \sum_{\mathbf{d} \in [N]^K} p_{\mathbf{D}}(\mathbf{d}) R_{\mathbf{d}} \\ &\geq \sum_{\mathbf{d} \in [N]^K} p_{\mathbf{D}}(\mathbf{d}) H(X_{\mathcal{D}}|\mathbf{D} = \mathbf{d}) \\ &= H(X_{\mathbf{D}}|\mathbf{D}) \\ &\geq H(X_{\mathbf{D}}|Z_{\mathcal{A}}, \mathbf{D}) \\ &= H(W_{D_{\mathcal{A}}}, X_{\mathbf{D}}|Z_{\mathcal{A}}, \mathbf{D}) - H(W_{D_{\mathcal{A}}}|X_{\mathbf{D}}, Z_{\mathcal{A}}, \mathbf{D}) \\ &\stackrel{(a)}{\geq} H(W_{D_{\mathcal{A}}}|Z_{\mathcal{A}}, \mathbf{D}) - F\epsilon_F \\ &= H(B_{D_{\mathcal{A}}}^F|Z_{\mathcal{A}}, \mathbf{D}) - F\epsilon_F \\ &= \sum_{i=1}^F H(B_{D_{\mathcal{A}i}}|B_{D_{\mathcal{A}}}^{i-1}, Z_{\mathcal{A}}, \mathbf{D}) - F\epsilon_F \\ &\geq \sum_{i=1}^F H(B_{D_{\mathcal{A}i}}|V_{\mathcal{A}i}, \mathbf{D}) - F\epsilon_F \end{aligned}$$

where (a) follows from the data processing inequality and Fano's inequality, and ϵ_F tends to zero as $F \rightarrow \infty$. The rest of the proof follows from the standard time sharing argument and then letting $F \rightarrow \infty$. ■

Now let us restrict attention to the case of i.i.d. requests, i.e., $p_{\mathbf{D}}(\mathbf{d}) = \prod_{k=1}^K p_D(d_k)$ for some distribution p_D defined on $[N]$. Then, by symmetry, the bound in Theorem 5 only depends on the cardinality of \mathcal{A} . Furthermore, to facilitate the analysis, we relax the lower bound by swapping the positions of minimum and maximum. Then, we have the following corollary.

Corollary 2: Consider the caching problem for the average case with $p_{\mathbf{D}}(\mathbf{d}) = \prod_{k=1}^K p_D(d_k)$ for some distribution p_D . For all $M \geq 0$,

$$R_{\text{uniform}}^*(M) \geq \max_{k \in [K]} \min H(B_{D_{[k]}}|V_k, \mathbf{D}), \quad (6)$$

where given a fixed $k \in [K]$, the minimum is over all conditional pmfs $p_{V_k|B}$ such that $V_k \text{---} B \text{---} \mathbf{D}$ form a Markov chain and $I(B; V_k) \leq kM$.

Next, we give a closed-form expression of (6) by relating it to a single-user caching problem. The caching network that we consider for this task is a generalization of our caching problem with $K = 1$. In particular, we formulate a single-user caching system in which the user may request *multiple* files, namely, any subset of the N files, with request distribution p_Y , where Y is an element of the power set $\mathcal{P}([N])$. We refer to this caching network as the single-user multiple request caching network.

The relation between the multi-user single request caching network (our primary problem of interest) and the single-user multiple request caching network is as follows. In the multi-user single request setup, users $1, \dots, K$ wish to recover files W_{D_1}, \dots, W_{D_K} , respectively. Following a cut-set based argument in which we assume that users in some subset $\mathcal{A} \subseteq [K]$ cooperate, the cache memories are combined resulting in a single cache of size $|\mathcal{A}|M$. Moreover, the (cooperative) decoder wishes to recover multiple files $(W_{D_k} : k \in \mathcal{A})$. Thus, the optimal memory–rate tradeoff for the single-user multiple request network with memory size $|\mathcal{A}|M$ and request $Y = (D_k : k \in \mathcal{A})$ serves as a lower bound on the multi-user single request network with memory size M and requests (D_1, \dots, D_K) .

A. Single-user multiple request caching

For each $n \in [N]$, we denote

$$s_n := \mathbb{P}(Y \ni n) = \sum_{\substack{y \in \mathcal{P}([N]) \\ \text{s.t. } n \in y}} p_Y(y). \quad (7)$$

Without loss of generality, we assume that $s_1 \geq s_2 \geq \dots \geq s_N$. We establish the following theorem for the single-user multiple request caching network.

Theorem 6: Consider the single-user multiple request caching problem with request distribution p_Y . The optimal memory–rate tradeoff for the average case is

$$R_{\text{avg}}^*(M) = \sum_{n=1}^N (s_n - s_{n+1}) (n - M)^+,$$

where s_n is defined in (7) for all $n \in [N]$ and $s_{N+1} = 0$.

The proof of Theorem 6 is deferred to Appendix.

Remark 2: Theorem 6 indicates that an optimal caching strategy for the single-user multiple request caching network is to cache the most popular files, where the popularity is measured by how often they are requested.

B. Proof of Theorem 1

Next, we observe that given any fixed $k \in [K]$, it holds that for all $n \in [N]$,

$$\begin{aligned} s_n &= \mathbb{P}(n \in \{D_1, \dots, D_k\}) \\ &= 1 - \mathbb{P}(n \notin \{D_1, \dots, D_k\}) \\ &= 1 - \prod_{j=1}^k \mathbb{P}(D_j \neq n) \\ &= 1 - (1 - P_D(n))^k. \end{aligned}$$

Finally, by applying Theorem 6 to Corollary 2, we have the closed-form converse bound in Theorem 1.

V. THE GAP ANALYSIS: PROOF OF THEOREMS 3 AND 4

If $N = 1$, it can be easily checked that $\check{R}_{\text{MN}}(K, 1, M) = 1 - M = R_{\text{uniform}}^*(M)$. For $N \in \{2, 3, 4\}$, we have

$$\begin{aligned} \frac{\check{R}_{\text{MN}}(K, N, M)}{R_{\text{uniform}}^*(M)} &\leq \frac{N - M}{\max_{k \in [K]} (1 - (1 - 1/N)^k)(N - kM)} \\ &\stackrel{k=1}{\leq} \frac{N - M}{(1 - (1 - 1/N))(N - M)} \\ &= N < 4.7. \end{aligned}$$

For the rest of analysis, we assume that $N \geq 5$. To facilitate the gap analysis, we consider the following relaxed upper bound of Theorem 2:

$$\begin{aligned} R_{\text{worst}}(M) &\leq (N - M) \cdot \min \left\{ \frac{1}{M}, 1 \right\} \\ &=: R_{\text{upper}}(K, N, M), \end{aligned}$$

for all $M \in (0, N]$, and we define $R_{\text{upper}}(K, N, 0) := \min\{K, N\}$. We remark that $R_{\text{upper}}(K, N, M)$ is quite sub-optimal as an upper bound and is not continuous at $M = 0$ when $K < N$. However, the corresponding convexified bound $\check{R}_{\text{upper}}(K, N, M)$ is sufficient for our analysis. On the other hand, we consider the following relaxed lower bound

$$\begin{aligned} R_{\text{uniform}}^*(M) &\geq \max_{k \in [\min\{K, \lceil N/4 \rceil\}]} (1 - (1 - 1/N)^k)(N - kM)^+ \\ &=: R_{\text{lower}}(K, N, M). \end{aligned}$$

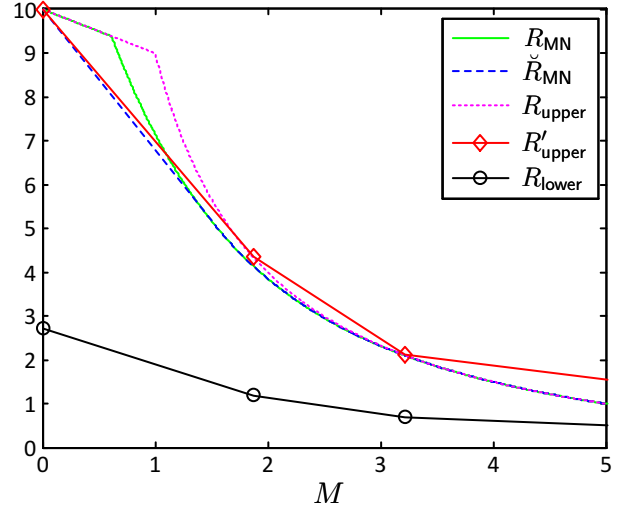


Fig. 2. Plots of various bounds for $(K, N) = (15, 10)$ and $M \in [0, 5]$.

Since $R_{\text{lower}}(K, N, M) \leq R_{\text{uniform}}^*(M) \leq R_{\text{worst}}^*(M) \leq \check{R}_{\text{upper}}(K, N, M)$, it suffices to show

$$\frac{\check{R}_{\text{upper}}(K, N, M)}{R_{\text{lower}}(K, N, M)} < 4.7, \quad M \in [0, N).$$

For notational convenience, we denote $\overline{K} = \min\{K, \lceil N/4 \rceil\}$ and $\kappa = \min\{K, N/4\}$.

The lower bound $R_{\text{lower}}(K, N, M)$ is an intersection of half planes. The corner points of $R_{\text{lower}}(K, N, M)$ are characterized by the set $\Omega = \{(\omega_k, R_{\text{lower}}(K, N, \omega_k)) : k \in [0 : \overline{K}]\}$, where

$$\omega_k = \begin{cases} N & \text{if } k = 0, \\ \frac{N(1 - \frac{1}{N})^k}{N + (k+1 - N)(1 - \frac{1}{N})^k} & \text{if } k \in [\overline{K} - 1], \\ 0 & \text{if } k = \overline{K}. \end{cases}$$

It can be checked that for all $k \in [\overline{K} - 1]$, the two lines

$$\begin{aligned} y &= (1 - (1 - 1/N)^k)(N - kx), \\ y &= (1 - (1 - 1/N)^{k+1})(N - (k+1)x) \end{aligned}$$

intersect at $x = \omega_k$.

Next, we relax the upper bound $\check{R}_{\text{upper}}(K, N, M)$ by the following piecewise-linear bound resulting from $\{\omega_k : k \in [0 : \overline{K}]\}$:

$$\begin{aligned} R'_{\text{upper}}(K, N, M) &:= (1 - \theta)R_{\text{upper}}(K, N, \omega_k) + \theta R_{\text{upper}}(K, N, \omega_{k-1}), \end{aligned}$$

where $M = (1 - \theta)\omega_k + \theta\omega_{k-1}$ for some $\theta \in [0, 1)$, $k \in [\overline{K}]$. Note that $R'_{\text{upper}}(K, N, M) = R_{\text{upper}}(K, N, M)$ for all $M \in \{\omega_k : k \in [0 : \overline{K}]\}$. In Figure 2 we provide an example with $(K, N) = (15, 10)$ summarizing the various bounds used in the analysis.

Then, for each segment $[\omega_k, \omega_{k-1})$, $k \in [\overline{K}]$, the ratio $\frac{R'_{\text{upper}}(K, N, M)}{R_{\text{lower}}(K, N, M)}$ is a linear-fractional function with respect to M , and thus it is quasiconvex [8]. A quasiconvex function has the

$$\begin{aligned}
&= \frac{\left[N + (k - N) \left(1 - \frac{1}{N} \right)^k \right] \left[N + (k - N + 1) \left(1 - \frac{1}{N} \right)^k \right]}{\left[N \left(1 - \frac{1}{N} \right)^k \right] \left[1 - \left(1 - \frac{1}{N} \right)^k \right] \left[N + (-N + 1) \left(1 - \frac{1}{N} \right)^k \right]} \\
&= \frac{\left[1 - \left(1 - \frac{1}{N} \right)^k + \frac{k}{N} \left(1 - \frac{1}{N} \right)^k \right] \left[1 - \left(1 - \frac{1}{N} \right)^{k+1} + \frac{k}{N} \left(1 - \frac{1}{N} \right)^k \right]}{\left(1 - \frac{1}{N} \right)^k \left[1 - \left(1 - \frac{1}{N} \right)^k \right] \left[1 - \left(1 - \frac{1}{N} \right)^{k+1} \right]} \\
&= \frac{1}{\left(1 - \frac{1}{N} \right)^k} \left[1 + \frac{\frac{k}{N} \left(1 - \frac{1}{N} \right)^k}{1 - \left(1 - \frac{1}{N} \right)^k} + \frac{\frac{k}{N} \left(1 - \frac{1}{N} \right)^k}{1 - \left(1 - \frac{1}{N} \right)^{k+1}} + \frac{\left(\frac{k}{N} \left(1 - \frac{1}{N} \right)^k \right)^2}{\left[1 - \left(1 - \frac{1}{N} \right)^k \right] \left[1 - \left(1 - \frac{1}{N} \right)^{k+1} \right]} \right]
\end{aligned}$$

property that the value of the function on a segment does not exceed the maximum of its values at the endpoints. That is to say, it suffices to check whether $\frac{R'_{\text{upper}}(K, N, M)}{R'_{\text{lower}}(K, N, M)} < 4.7$ for all $M \in \{\omega_0, \omega_1, \dots, \omega_{\overline{K}}\}$.

First, it is clear that we have

$$\lim_{M \uparrow N} \frac{R'_{\text{upper}}(K, N, M)}{R'_{\text{lower}}(K, N, M)} = 1.$$

Next, we have

$$\begin{aligned}
\frac{R'_{\text{upper}}(K, N, 0)}{R'_{\text{lower}}(K, N, 0)} &= \frac{\min\{K, N\}}{(1 - (1 - 1/N)^{\overline{K}})N} \\
&\leq \frac{4\kappa/N}{1 - (1 - 1/N)^\kappa} \\
&\stackrel{(a)}{\leq} 4 \cdot \frac{\kappa/N}{1 - e^{-\kappa/N}} \\
&\stackrel{(b)}{\leq} \frac{1}{1 - e^{-1/4}} \approx 4.521,
\end{aligned}$$

where (a) follows since $(1 - 1/z)^z \leq e^{-1}$ for all $z > 1$ and (b) follows since $\psi(z) = \frac{z}{1 - e^{-z}}$ is an increasing function and $\kappa/N \leq 1/4$.

As for $k \in [\overline{K} - 1]$, we have

$$\begin{aligned}
&\frac{R'_{\text{upper}}(K, N, \omega_k)}{R'_{\text{lower}}(K, N, \omega_k)} \\
&\leq \frac{\frac{N - \omega_k}{\omega_k}}{(1 - (1 - 1/N)^k)(N - k\omega_k)} \\
&= (\text{see the top of the current page}) \\
&\leq \frac{1}{\left(1 - \frac{1}{N} \right)^k} \left(1 + \frac{\frac{k}{N} \left(1 - \frac{1}{N} \right)^k}{1 - \left(1 - \frac{1}{N} \right)^k} \right)^2 \\
&\stackrel{(a)}{=} e^z \left(1 + \frac{1}{\ln(1 - 1/N)^{-N}} \frac{z}{e^z - 1} \right)^2 \\
&\stackrel{(b)}{\leq} e^z \left(1 + \frac{z}{e^z - 1} \right)^2 \\
&\stackrel{(c)}{\leq} e^z \left(1 + \frac{z}{e^z - 1} \right)^2 \Bigg|_{z = -\frac{N}{4} \ln(1 - 1/N)},
\end{aligned}$$

where (a) follows by a change of variable $z = -k \ln(1 - 1/N)$, (b) follows since $z \geq 0$ and $(1 - 1/N)^{-N} \geq e$ for

all $N > 1$, (c) follows since $\phi(z) = e^z \left(1 + \frac{z}{e^z - 1} \right)^2$ is an increasing function[§] and $z \leq -\frac{N}{4} \ln(1 - 1/N)$ (since $k \leq \overline{K} - 1 \leq N/4$). Finally, since $-N \ln(1 - 1/N)$ is a decreasing function of N and $N \geq 5$, we have

$$\begin{aligned}
\frac{R'_{\text{upper}}(K, N, \omega_k)}{R'_{\text{lower}}(K, N, \omega_k)} &\leq e^z \left(1 + \frac{z}{e^z - 1} \right)^2 \Bigg|_{z = -\frac{5}{4} \ln(1 - 1/5)} \\
&= \nu^\nu \left(1 + \frac{\nu \ln \nu}{\nu^\nu - 1} \right)^2 \Bigg|_{\nu = \frac{5}{4}} \\
&\approx 4.607.
\end{aligned}$$

APPENDIX: PROOF OF THEOREM 6

For the single-user caching problem, a single-letter characterization of the optimal memory–rate tradeoff for the average case can be found in [9, Chapter 3.7.3]. For the considered setup in Theorem 6, the optimal memory–rate tradeoff can be expressed as

$$R_{\text{avg}}^*(M) = \min H(B_Y | V, Y),$$

where the minimum is over all conditional pmfs $p_{V|B}$ such that $V \text{---} B \text{---} Y$ form a Markov chain and $I(B; V) \leq M$.

(Converse.) Let $M \geq 0$ be fixed. Consider any conditional pmf $p_{V|B}$ such that $V \text{---} B \text{---} Y$ form a Markov chain and $I(B; V) \leq M$. Recall that $B = (B_1, \dots, B_N)$. Then, for all $n \in [N]$, we have

$$\begin{aligned}
M &\geq I(B; V) \\
&\geq I(B_{[n]}; V | B_{[n+1:N]}) \\
&= H(B_{[n]}) - H(B_{[n]} | V, B_{[n+1:N]}) \\
&= n - H(B_n | V, B_{[n+1:N]}) - H(B_{[n-1]} | V, B_{[n:N]}). \quad (8)
\end{aligned}$$

Now we show that

$$H(B_Y | V, Y) \geq \sum_{n=1}^N (s_n - s_{n+1}) (n - M)^+.$$

[§] It can be verified that the function $\phi(z) = e^z \left(1 + \frac{z}{e^z - 1} \right)^2$, $z \geq 0$, is an increasing function by showing that its first derivative is nonnegative.

First, we have

$$\begin{aligned}
& H(B_Y|V, Y) \\
&= \sum_{y \in \mathcal{P}([N])} p_Y(y) H(B_y|V) \\
&\stackrel{(a)}{\geq} \sum_{n=1}^N s_n H(B_n|V, B_{[n+1:N]}),
\end{aligned}$$

where (a) follows by recursively applying the inequality

$$\begin{aligned}
& \sum_{y \in \mathcal{P}([N])} p_Y(y) H(B_y|V, B_{[n+1:N]}) \\
&\geq s_n H(B_n|V, B_{[n+1:N]}) \\
&\quad + \sum_{y \in \mathcal{P}([N])} p_Y(y) H(B_y|V, B_n, B_{[n+1:N]}),
\end{aligned}$$

in the order $N, N-1, \dots, 1$. Next, $H(B_Y|V, Y)$ can be further lower bounded as

$$\begin{aligned}
& H(B_Y|V, Y) \\
&\geq \sum_{n=1}^N s_n H(B_n|V, B_{[n+1:N]}) \\
&= s_N H(B_N|V) + \sum_{n=1}^{N-1} s_n H(B_n|V, B_{[n+1:N]}) \\
&\stackrel{(a)}{\geq} s_N (N - M - H(B_{[N-1]}|V, B_N))^+ \\
&\quad + \sum_{n=1}^{N-1} s_n H(B_n|V, B_{[n+1:N]}) \\
&\stackrel{(b)}{\geq} s_N (N - M)^+ - s_N H(B_{[N-1]}|V, B_N) \\
&\quad + \sum_{n=1}^{N-1} s_n H(B_n|V, B_{[n+1:N]}) \\
&= s_N (N - M)^+ + \sum_{n=1}^{N-1} (s_n - s_N) H(B_n|V, B_{[n+1:N]}) \\
&= s_N (N - M)^+ + (s_{N-1} - s_N) H(B_{N-1}|V, B_N) \\
&\quad + \sum_{n=1}^{N-2} (s_n - s_N) H(B_n|V, B_{[n+1:N]}) \\
&\stackrel{(c)}{\geq} s_N (N - M)^+ \\
&\quad + (s_{N-1} - s_N) (N - 1 - M - H(B_{[N-2]}|V, B_{[N-1:N]}))^+ \\
&\quad + \sum_{n=1}^{N-2} (s_n - s_N) H(B_n|V, B_{[n+1:N]}) \\
&\stackrel{(d)}{\geq} s_N (N - M)^+ + (s_{N-1} - s_N) (N - 1 - M)^+ \\
&\quad - (s_{N-1} - s_N) H(B_{[N-2]}|V, B_{[N-1:N]}) \\
&\quad + \sum_{n=1}^{N-2} (s_n - s_N) H(B_n|V, B_{[n+1:N]}) \\
&= s_N (N - M)^+ + (s_{N-1} - s_N) (N - 1 - M)^+
\end{aligned}$$

$$+ \sum_{n=1}^{N-2} (s_n - s_{N-1}) H(B_n|V, B_{[n+1:N]}),$$

where (a) and (c) follow from (8) and $H(B_n|V, B_{[n+1:N]}) \geq 0$ with $n = N$ and $n = N - 1$, respectively, and (b) and (d) follow since $(u - v)^+ \geq (u)^+ - v$ for all $v \geq 0$. At this point, it is clear that we can apply the same argument for another $N - 2$ times and arrive at

$$H(B_Y|V, Y) \geq \sum_{n=1}^N (s_n - s_{n+1}) (n - M)^+, \quad (9)$$

where $s_{N+1} = 0$.

(Achievability.) Note that the lower bound (9) is equivalent to saying that

- 1) if $M \geq N$, then $H(B_Y|V, Y) \geq 0$, and
- 2) if $n - 1 \leq M < n$ for some $n \in [N]$, then

$$\begin{aligned}
& H(B_Y|V, Y) \\
&\geq s_n (n - M) + \sum_{j=n+1}^N s_j \\
&= \sum_{\substack{y \in \mathcal{P}[N] \\ \text{s.t. } n \in y}} p_Y(y) (n - M) + \sum_{y \in \mathcal{P}[N]} p_Y(y) H(B_y|B_{[n]}).
\end{aligned}$$

Therefore, for all $M \in \{0\} \cup [N]$, setting $V = (B_1, \dots, B_M)$ makes (9) hold with equality. Since the rest of the memory-rate tradeoff can be achieved by time sharing, the achievability is established.

ACKNOWLEDGEMENT

This work has been supported in part by the European ERC Starting Grant 259530-ComCom.

REFERENCES

- [1] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, pp. 2856–2867, May 2014.
- [2] U. Niesen and M. A. Maddah-Ali, "Coded caching with nonuniform demands," in *arXiv:1308.0178[cs.IT]*, Mar. 2014.
- [3] J. Zhang, X. Lin, and X. Wang, "Coded caching under arbitrary popularity distributions," in *Proc. Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb. 2015.
- [4] Z. Chen, P. Fan, and K. B. Letaief, "Fundamental limits of caching: Improved bounds for small buffer users," in *arXiv:1407.1935[cs.IT]*, Nov. 2015.
- [5] H. Ghasemi and A. Ramamoorthy, "Improved lower bounds for coded caching," in *arXiv:1501.06003[cs.IT]*, Jan. 2015.
- [6] M. Ji, A. M. Tulino, J. Llorca, and G. Caire, "Order-optimal rate of caching and coded multicasting with random demands," in *arXiv:1502.03124[cs.IT]*, Feb. 2015.
- [7] M. A. Maddah-Ali and U. Niesen, "Decentralized coded caching attains order-optimal memory-rate tradeoff," *IEEE/ACM Trans. Networking*, vol. 23, pp. 1029–1040, Aug. 2015.
- [8] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge Univ. Press, 2004.
- [9] C.-Y. Wang, "Function computation over networks: Efficient information processing for cache and sensor applications," Ph.D. dissertation, EPFL, 2015.