Caching and Coded Delivery over Gaussian Broadcast Channels for Energy Efficiency

Mohammad Mohammadi Amiri, *Student Member, IEEE*, and Deniz Gündüz, *Senior Member, IEEE*

Abstract—A cache-aided K-user Gaussian broadcast channel (BC) is considered. The transmitter has a library of N equal-rate files, from which each user demands one. The impact of the equal-capacity receiver cache memories on the minimum required transmit power to satisfy all user demands is studied. Considering uniformly random demands across the library, both the minimum average power (averaged over all demand combinations) and the minimum peak power (minimum power required to satisfy all demand combinations) are studied. Upper bounds are presented on the minimum required average and peak transmit power as a function of the cache capacity considering both *centralized* and *decentralized* caching. The lower bounds on the minimum required average and peak power values are also derived assuming uncoded cache placement. The bounds for both the peak and average power values are shown to be tight in the centralized scenario through numerical simulations. The results in this paper show that proactive caching and coded delivery can provide significant energy savings in wireless networks.

Index Terms—Gaussian broadcast channel, centralized caching, decentralized caching, joint cache-channel coding.

I. INTRODUCTION

We study proactive content caching to user devices followed by coded delivery [1], assuming that the delivery phase takes place over a Gaussian broadcast channel (BC) from the server to the users. Several other recent papers have studied coded delivery over noisy channels. Fading and interference channel models are considered in [2] and [3], respectively. In [4] and [5], centralized caching is considered while the delivery phase takes place over a packet-erasure BC. The capacity-memory trade-off is investigated in this setting assuming that only the weak users have caches, which requires knowledge about the channel qualities in the delivery phase in advance. Proactive caching and coded delivery over a Gaussian BC is studied in [6] in the high power regime. In [7] cache-aided data delivery is studied over a degraded BC in a centralized setting, while [8] focuses on the delay analysis in a cache-aided wireless channel with coded delivery by utilizing superposition coding. A multi-antenna server is considered in [9]–[11] with different assumptions regarding the channel state information available at the transmitter.

While most of the previous works on coded delivery over noisy wireless broadcast channels focus on the high signal-to-noise ratio (SNR) regime; in this work, our goal is to highlight the benefits of proactive caching and coded delivery from an energy efficiency perspective. The impact of proactive caching on energy efficiency was previously studied in [12] and [13] in an offline scenario, assuming known demands and channel conditions. However, these works consider only proactive caching and do not benefit from coded delivery.

In [3]–[7], users' channel conditions in the delivery phase are assumed to be known in advance during the placement phase. However, in practical wireless networks, it is often not possible to know the identities of users that will participate in a particular delivery phase, let alone their channel conditions. Therefore, our goal here is to study the benefits of proactive caching in reducing the transmit power, assuming that the noiseless cache placement phase is carried out without the knowledge of channel conditions during the delivery phase.

Assuming uniform popularity across the files, we first study the minimum required *average power* to serve all the users, averaged over all the user demand combinations. Note that we allow the transmitter to change its power depending on the demand combination in order to minimize the average power consumption. We then consider the transmit power required to satisfy the worst-case demand combination, called the *peak power*. Building upon our previous work in [14], we first provide upper bounds on the minimum average and peak power values as a function of the rate of the files in the library and the capacity of the user caches, for centralized cache placement. We then extend the proposed scheme by considering *decentralized* cache placement. The proposed delivery strategy employs superposition coding and power allocation, and the achievable transmit power for any demand combination is derived thanks to the degradedness of a Gaussian BC. We further derive lower bounds on the performance assuming uncoded cache placement.

The main novelty of the proposed proactive caching and coded delivery scheme is the way the coded packets designed for each user are generated for any demand combination, particularly when a file may be requested by more than one user, and the way these coded packets are delivered over a Gaussian BC in order to minimize the transmit power. We show that the proposed achievable scheme reduces the transmit power significantly, even with the availability of only a small cache capacity at each receiver, in both the centralized and decentralized scenarios. It is also shown that the power loss between the centralized and the more practical decentralized scenario is quite small. Furthermore, numerical results show that the gaps between the peak and average transmit powers of the proposed achievable scheme for the centralized scenario and the corresponding lower bounds are negligible. In particular, we have observed numerically that the multiplicative gap between the two bounds for the centralized caching is below 2 for the examples considered. Numerical results

also illustrate that adjusting the transmit power based on the demand vector can significantly reduce the average power consumption compared to the worst-case demand combination.

II. SYSTEM MODEL AND PRELIMINARIES

We study cache-aided content delivery over a K-user Gaussian BC. The server has a library of N files, $\mathbf{W} \triangleq (W_1, ..., W_N)$, each distributed uniformly over the set[1] $\left[\left\lceil 2^{nR}\right\rceil\right]$, where R is the rate of the files and n denotes the blocklength, referring to n uses of the BC. Each user has a cache of size nMR bits. We define the *normalized global cache capacity* as $t \triangleq MK/N$.

Data delivery from the server to the users takes place in two phases. Caches of the users are filled during the initial *placement phase*, which takes place over a period of low traffic and high energy efficiency; and therefore, data delivery in the placement phase is assumed to be error-free and at a negligible energy cost[2]; however, without either the knowledge of the user demands, or the users' future channel gains when they place their requests. The caching function for user $k \in [K]$ is

$$\phi_k : \left[\left\lceil 2^{nR}\right\rceil\right]^N \to \left[\left\lfloor 2^{nMR}\right\rfloor\right], \tag{1}$$

which maps the library to the cache contents U_k of user k, i.e., $U_k = \phi_k(\mathbf{W})$, for $k \in [K]$.

During the peak traffic period, each user requests a single file from the library, where $d_k \in [N]$ denotes the index of the file requested by user $k \in [K]$. We assume that the user demands are independent and uniformly distributed over the file library; that is $\Pr\{d_k = i\} = 1/N$, $\forall i \in [N], \forall k \in [K]$, and $\mathbf{d} \triangleq (d_1, ..., d_K)$ denotes the demand vector. The requests must be satisfied simultaneously during the *delivery phase*. As opposed to the placement phase, the delivery phase takes place over a noisy BC, characterized by

$$Y_{k,i}(\mathbf{W}, \mathbf{d}) = h_k X_i(\mathbf{W}, \mathbf{d}) + Z_{k,i}, \text{ for } i \in [n], k \in [K], \tag{2}$$

where $X_i(\mathbf{W}, \mathbf{d})$ denotes the transmitted signal from the server at time i, h_k is a real channel gain between the server and user k, $Z_{k,i}$ is the zero-mean unit-variance real Gaussian noise at user k at time i, i.e., $Z_{k,i} \sim \mathcal{N}(0, 1)$, and $Y_{k,i}(\mathbf{W}, \mathbf{d})$ is the signal received at user k. The noise components are assumed independent and identically distributed (i.i.d.) across time and users. Without loss of generality, we assume that $h_1^2 \leq h_2^2 \leq \cdots \leq h_K^2$. The channel input vector $X^n(\mathbf{W}, \mathbf{d})$ is generated by

$$\psi : \left[\left\lceil 2^{nR}\right\rceil\right]^N \times [N]^K \to \mathbb{R}^n, \tag{3}$$

and its average power is given by $P(\mathbf{W}, \mathbf{d}) \triangleq \frac{1}{n}\sum_{i=1}^n X_i^2(\mathbf{W}, \mathbf{d})$. We define the average power of this encoding function for demand vector $\mathbf{d}$ as

$$P(\mathbf{d}) \triangleq \max_{W_1, ..., W_N} P(\mathbf{W}, \mathbf{d}), \tag{4}$$

where the maximization is over all possible realizations of the file library.

User $k \in [K]$ reconstructs $\hat{W}_{d_k}$ using its channel output $Y_k^n(\mathbf{W}, \mathbf{d})$, local cache contents U_k, channel vector $\mathbf{h} \triangleq (h_1, ..., h_K)$, and demand vector $\mathbf{d}$ through the function

$$\mu_k : \mathbb{R}^n \times \left[\left\lfloor 2^{nMR}\right\rfloor\right] \times \mathbb{R}^K \times [N]^K \to \left[\left\lceil 2^{nR}\right\rceil\right], \tag{5}$$

where $\hat{W}_{d_k} = \mu_k\left(Y_k^n(\mathbf{W}, \mathbf{d}), U_k, \mathbf{h}, \mathbf{d}\right)$. The probability of error is defined as

$$P_e \triangleq \Pr\left\{\bigcup_{\mathbf{d} \in [N]^K} \bigcup_{k=1}^K \left\{\hat{W}_{d_k} \neq W_{d_k}\right\}\right\}. \tag{6}$$

An (n, R, M) *code* consists of K caching functions $\phi_1, \ldots, \phi_K$, channel encoding function ψ, and K decoding functions $\mu_1, \ldots, \mu_K$. We say that an $\left(R, M, \bar{P}, \hat{P}\right)$ tuple is *achievable* if for every $\varepsilon > 0$, there exists an (n, R, M) code with sufficiently large n, which satisfies $P_e < \varepsilon$, $\mathrm{E}_{\mathbf{d}}[P(\mathbf{d})] \leq \bar{P}$, and $P(\mathbf{d}) \leq \hat{P}$, $\forall \mathbf{d}$. For given rate R and normalized cache capacity M, the average and peak power-memory trade-offs are defined, respectively, as

$$\bar{P}^*(R, M) \triangleq \inf\left\{\bar{P} : \left(R, M, \bar{P}, \infty\right) \text{ is achievable}\right\}, \tag{7a}$$

$$\hat{P}^*(R, M) \triangleq \inf\left\{\hat{P} : \left(R, M, \hat{P}, \hat{P}\right) \text{ is achievable}\right\}. \tag{7b}$$

Remark 1. *In the above definition, $\bar{P}^*$ is evaluated by allowing a different transmission power for each demand combination, and minimizing the average power across demands; while $\hat{P}^*$ characterizes the worst-case transmit power, which can also be considered as the minimum transmit power required to satisfy all possible demands if the transmitter is not allowed to adapt its transmit power according to user demands.*

We will consider both centralized and decentralized caching, and assume, in both scenarios that the placement phase is performed without any information about the channel gains during the delivery phase.

Proposition 1. *[15] Consider a K-user Gaussian BC presented above with $M = 0$, where a distinct message of rate R_k is targeted for user k, $k \in [K]$. The minimum total power P that is required to deliver all K messages reliably can be achieved by superposition coding with Gaussian codewords of power $\alpha_k P$ allocated for user k, $\forall k \in [K]$, where[3]*

$$\alpha_k P = \left(\frac{2^{2R_k} - 1}{h_k^2}\right)$$
$$\left(1 + h_k^2 \sum_{i=k+1}^K \left(\frac{2^{2R_i} - 1}{h_i^2}\right) \prod_{j=k+1}^{i-1} 2^{2R_j}\right), \tag{8}$$

and the total transmitted power is

$$P = \sum_{k=1}^K \left(\frac{2^{2R_k} - 1}{h_k^2}\right) \prod_{i=1}^{k-1} 2^{2R_i}. \tag{9}$$

For a demand vector $\mathbf{d}$ in the delivery phase, we denote the number of distinct demands by $N_{\mathbf{d}}$, where $N_{\mathbf{d}} \leq \min\{N, K\}$.

---

[1] For any positive integer i, $[i]$ denotes the set $\{1, ..., i\}$.

[2] We assume that the placement phase takes place over a significantly longer period of time and over orthogonal high-quality links; which, in theory, allows the server to achieve the minimum energy per bit required to send the cache contents to the users.

[3] For two integers i and j, if $i > j$, we assume that $\sum_{n=i}^j a_n = 0$, and $\prod_{n=i}^j a_n = 1$, where a_n is an arbitrary sequence.

Let $\mathcal{U}_{\mathbf{d}}$ denote the set of users with distinct requests, which have the worst channel qualities; that is, $\mathcal{U}_{\mathbf{d}}$ consists of $N_{\mathbf{d}}$ indices corresponding to users with distinct requests, where a user is included in set $\mathcal{U}_{\mathbf{d}}$ iff it has the worst channel quality among all the users with the same demand, i.e., if $k \in \mathcal{U}_{\mathbf{d}}$ and $d_k = d_m$ for some $m \in [K]$, then $h_k^2 \leq h_m^2$, or equivalently, $k \leq m$. Note that, for any demand vector $\mathbf{d}$, $1 \in \mathcal{U}_{\mathbf{d}}$. For given $\mathbf{d}$ and user k, $k \in [K]$, let $\mathcal{U}_{\mathbf{d},k}$ denote the set of users in $\mathcal{U}_{\mathbf{d}}$ which have better channels than user k:

$$\mathcal{U}_{\mathbf{d},k} \triangleq \{i \in \mathcal{U}_{\mathbf{d}} : i > k\}, \quad k \in [K]. \tag{10}$$

We denote the cardinality of $\mathcal{U}_{\mathbf{d},k}$ by $N_{\mathbf{d},k}$, i.e., $N_{\mathbf{d},k} \triangleq |\mathcal{U}_{\mathbf{d},k}|$.

III. CENTRALIZED CACHING AND DELIVERY

Here we present our centralized caching and coded delivery scheme. We follow the placement phase in [16] since the users' channel gains are not known in advance. In the delivery phase, our goal is to identify the coded packets targeted to each user in order to minimize the transmit power. We will use superposition coding in sending multiple coded packets, and benefit heavily from the degradedness of the underlying Gaussian BC.

Theorem 1. *In centralized caching followed by delivery over a Gaussian BC, we have*

$$\bar{P}^*(R,M) \leq \bar{P}^{\mathrm{C}}_{\mathrm{UB}}(R,M) \triangleq$$

$$\frac{1}{N^K} \sum_{\mathbf{d} \in [N]^K} \left[\sum_{i=1}^{K} \left(\frac{2^{2R^{\mathrm{C}}_{\mathbf{d},i}} - 1}{h_i^2} \right) \prod_{j=1}^{i-1} 2^{2R^{\mathrm{C}}_{\mathbf{d},j}} \right], \tag{11a}$$

where

$$R^{\mathrm{C}}_{\mathbf{d},k} \triangleq \begin{cases} \frac{\binom{K-k}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}} (\lfloor t \rfloor + 1 - t) R + \\ \qquad \frac{\binom{K-k}{\lfloor t \rfloor + 1}}{\binom{K}{\lfloor t \rfloor + 1}} (t - \lfloor t \rfloor) R, & \text{if } k \in \mathcal{U}_{\mathbf{d}}, \\ \frac{\binom{K-k}{\lfloor t \rfloor} - \binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}} (\lfloor t \rfloor + 1 - t) R + \\ \qquad \frac{\binom{K-k}{\lfloor t \rfloor + 1} - \binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor + 1}}{\binom{K}{\lfloor t \rfloor + 1}} (t - \lfloor t \rfloor) R, & \text{otherwise,} \end{cases} \tag{11b}$$

and

$$\hat{P}^*(R,M) \leq \hat{P}^{\mathrm{C}}_{\mathrm{UB}}(R,M) \triangleq \sum_{i=1}^{K} \left(\frac{2^{2\hat{R}^{\mathrm{C}}_{\mathbf{d},i}} - 1}{h_i^2} \right) \prod_{j=1}^{i-1} 2^{2\hat{R}^{\mathrm{C}}_{\mathbf{d},j}}, \tag{12a}$$

where

$$\hat{R}^{\mathrm{C}}_{\mathbf{d},k} \triangleq \begin{cases} \frac{\binom{K-k}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}} (\lfloor t \rfloor + 1 - t) R + \\ \qquad \frac{\binom{K-k}{\lfloor t \rfloor + 1}}{\binom{K}{\lfloor t \rfloor + 1}} (t - \lfloor t \rfloor) R, & \text{if } k \in [\min\{N,K\}], \\ 0, & \text{otherwise.} \end{cases} \tag{12b}$$

Proof. For simplicity, we assume that both nR and nMR are integers. The proposed scheme is first presented for integer normalized global cache capacities, that is $t \in [0 : K]$. The scheme is then extended to any $t \in [0, K]$.

A. Integer t Values

Here, we assume that $t \in [0 : K]$. We denote the t-element subsets of K users by $\mathcal{C}_1^t, \mathcal{C}_2^t, \ldots, \mathcal{C}_{\binom{K}{t}}^t$.

Placement phase: A centralized cache placement phase is performed without the knowledge of the future user demands or the channel gains in the delivery phase. Each file W_i, $i \in [N]$, is split into $\binom{K}{t}$ equal-length subfiles $W_{i,\mathcal{C}_1^t}, W_{i,\mathcal{C}_2^t}, \ldots, W_{i,\mathcal{C}_{\binom{K}{t}}^t}$, each of rate $R/\binom{K}{t}$. User k, $k \in [K]$, caches subfile $W_{i,\mathcal{C}_l^t}$, if $k \in \mathcal{C}_l^t$, for $i \in [N]$ and $l \in \left[\binom{K}{t}\right]$. Hence, the cache contents of user k is given by

$$U_k = \bigcup_{i \in [N]} \bigcup_{l \in \left[\binom{K}{t}\right]: k \in \mathcal{C}_l^t} W_{i,\mathcal{C}_l^t}, \quad \text{for } k \in [K], \tag{13}$$

where the cache capacity constraint is satisfied with equality.

Delivery phase: For an arbitrary $\mathbf{d}$, we will deliver the following coded message to the users in $\mathcal{C}_l^{t+1}$, $\forall l \in \left[\binom{K}{t+1}\right]$:

$$Q_{\mathcal{C}_l^{t+1}} \triangleq \bigoplus_{k \in \mathcal{C}_l^{t+1}} W_{d_k, \mathcal{C}_l^{t+1} \setminus \{k\}}, \tag{14}$$

where $\oplus$ represents the bitwise XOR operation. Then, each user $i \in \mathcal{C}_l^{t+1}$ can recover subfile $W_{d_i, \mathcal{C}_l^{t+1} \setminus \{i\}}$, since it has cached all the other subfiles $W_{d_j, \mathcal{C}_l^{t+1} \setminus \{j\}}$, $\forall j \in \mathcal{C}_l^{t+1} \setminus \{i\}$. Note that each coded message $Q_{\mathcal{C}_l^{t+1}}$, $l \in \left[\binom{K}{t+1}\right]$, is of rate $R/\binom{K}{t}$. Note also that, for $k \in [K]$, sending $\bigcup_{l: k \in \mathcal{C}_l^{t+1}} Q_{\mathcal{C}_l^{t+1}}$ to user k enables that user to obtain all the subfiles $W_{d_k, \mathcal{C}_l^t}$, $\forall l \in \left[\binom{K}{t}\right]$ and $k \notin \mathcal{C}_l^t$. Thus, together with its cache contents, the demand of user k, $k \in [K]$ would be fully satisfied after receiving $\bigcup_{l: k \in \mathcal{C}_l^{t+1}} Q_{\mathcal{C}_l^{t+1}}$. As observed in [16], for a demand vector $\mathbf{d}$ with $N_{\mathbf{d}}$ distinct requests, if $K - N_{\mathbf{d}} \geq t+1$, not all the coded messages $Q_{\mathcal{C}_l^{t+1}}$, $\forall l \in \left[\binom{K}{t}\right]$, need to be delivered.

Following [16, Lemma 1], for a demand vector $\mathbf{d}$ with $N_{\mathbf{d}}$ distinct requests, let $\mathcal{U}_{\mathbf{d}} \subset \mathcal{B} \subset [K]$. We define $\mathcal{G}_{\mathcal{B}}$ as the set consisting of all the subsets of $\mathcal{B}$ with cardinality $N_{\mathbf{d}}$, such that all $N_{\mathbf{d}}$ users in each subset request distinct files. For any $\mathcal{B}$, we have $\bigoplus_{\mathcal{G} \in \mathcal{G}_{\mathcal{B}}} Q_{\mathcal{B} \setminus \mathcal{G}} = \mathbf{0}$, where $\mathbf{0}$ denotes the zero vector.

Remark 2. *Given a demand vector $\mathbf{d}$ with $N_{\mathbf{d}} < K$, and any set $\mathcal{S} \subset [K] \setminus \mathcal{U}_{\mathbf{d}}$ of users, by setting $\mathcal{B} = \mathcal{S} \cup \mathcal{U}_{\mathbf{d}}$, we have*

$$\bigoplus_{\mathcal{G} \in \mathcal{G}_{\mathcal{B}}} Q_{\mathcal{B} \setminus \mathcal{G}} = \left(\bigoplus_{\mathcal{G} \in \mathcal{G}_{\mathcal{B}} \setminus \mathcal{U}_{\mathbf{d}}} Q_{\mathcal{B} \setminus \mathcal{G}} \right) \oplus Q_{\mathcal{B} \setminus \mathcal{U}_{\mathbf{d}}}$$

$$= \left(\bigoplus_{\mathcal{G} \in \mathcal{G}_{\mathcal{B}} \setminus \mathcal{U}_{\mathbf{d}}} Q_{\mathcal{B} \setminus \mathcal{G}} \right) \oplus Q_{\mathcal{S}} = \mathbf{0}, \tag{15}$$

which leads to

$$Q_{\mathcal{S}} = \bigoplus_{\mathcal{G} \in \mathcal{G}_{\mathcal{B}} \setminus \mathcal{U}_{\mathbf{d}}} Q_{\mathcal{B} \setminus \mathcal{G}}. \tag{16}$$

Thus, having received all the coded messages $Q_{\mathcal{B} \setminus \mathcal{G}}$, $\forall \mathcal{G} \in \mathcal{G}_{\mathcal{B}} \setminus \mathcal{U}_{\mathbf{d}}$, $Q_{\mathcal{S}}$ can be recovered through (16). Note that, for any $\mathcal{G} \in \mathcal{G}_{\mathcal{B}} \setminus \mathcal{U}_{\mathbf{d}}$, we have

$$|\mathcal{B} \setminus \mathcal{G}| = |\mathcal{S}|, \tag{17a}$$

$$(\mathcal{B} \setminus \mathcal{G}) \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, \tag{17b}$$

that is, each coded message on the right hand side (RHS) of (16) is targeted for a set of $|\mathcal{S}|$ users, at least one of which is

in set $\mathcal{U}_{\mathbf{d}}$. Furthermore, for each $k \in \mathcal{S}$, there is a user $k' \in \mathcal{U}_{\mathbf{d}}$ with $h_{k'}^2 \leq h_k^2$, such that $d_{k'} = d_k$. Note that, since no two users with the same demand are in any of the sets $\mathcal{G} \in \mathcal{G}_{\mathcal{B}}$, for any set $\mathcal{G} \in \mathcal{G}_{\mathcal{B}} \backslash \mathcal{U}_{\mathbf{d}}$, we have either $k \in \mathcal{B} \backslash \mathcal{G}$ or $k' \in \mathcal{B} \backslash \mathcal{G}$.

Given a demand vector $\mathbf{d}$, the delivery phase is designed such that only the coded messages $Q_{\mathcal{C}_l^{t+1}}$, $\forall l \in \left[\binom{K}{t+1} \right]$ such that $\mathcal{C}_l^{t+1} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset$, are delivered, i.e., the coded messages that are targeted for at least one user in $\mathcal{U}_{\mathbf{d}}$ are delivered, and the remaining coded messages can be recovered through (16). To achieve this, for any such set $\mathcal{C}_l^{t+1}$ with $\mathcal{C}_l^{t+1} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset$, the transmission power is adjusted such that the worst user in $\mathcal{C}_l^{t+1}$ can decode $Q_{\mathcal{C}_l^{t+1}}$; and so can all the other users in $\mathcal{C}_l^{t+1}$ due to the degradedness of the Gaussian BC. As a result, the demand of every user in $\mathcal{U}_{\mathbf{d}}$ will be satisfied.

We aim to find the coded packets targeted for each user that will minimize the transmitted power, while guaranteeing that all the user demands are satisfied. In delivering the coded messages, we start from the worst user, i.e., user 1, and first transmit all the coded messages targeted for user 1. We then target the second worst user, and transmit the coded messages targeted for it that have not been already delivered, keeping in mind that only the coded messages $Q_{\mathcal{C}_l^{t+1}}$ for which $\mathcal{C}_l^{t+1} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset$ are delivered. We continue similarly until we deliver the messages targeted for the best user in $\mathcal{U}_{\mathbf{d}}$.

We denote the contents targeted for user k by $\tilde{Q}_k$, and their total rate by $R_{\mathbf{d},t,k}$, $k \in [K]$. For a demand vector $\mathbf{d}$, contents

$$\tilde{Q}_1 = \bigcup_{l:1 \in \mathcal{C}_l^{t+1}} Q_{\mathcal{C}_l^{t+1}} \tag{18}$$

are targeted for user 1. Note that there are $\binom{K-1}{t}$ different $(t+1)$-element subsets $\mathcal{C}_l^{t+1}$, in which user 1 is included. Thus, the total rate of the messages targeted for user 1 is

$$R_{\mathbf{d},t,1} = \frac{\binom{K-1}{t}}{\binom{K}{t}} R = \frac{K-t}{K} R. \tag{19}$$

For $k \in [2:K]$, the coded contents targeted for user k that have not been sent through the transmissions to the previous $k-1$ users and are targeted for at least one user in $\mathcal{U}_{\mathbf{d}}$ are delivered. Thus, for $k \in [2:K]$, we deliver

$$\tilde{Q}_k = \bigcup_{l:\mathcal{C}_l^{t+1} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, \mathcal{C}_l^{t+1} \cap [k-1] = \emptyset, k \in \mathcal{C}_l^{t+1}} Q_{\mathcal{C}_l^{t+1}}, \tag{20}$$

which is equivalent to

$$\tilde{Q}_k = \left(\bigcup_{l:\mathcal{C}_l^{t+1} \cap [k-1] = \emptyset, k \in \mathcal{C}_l^{t+1}} Q_{\mathcal{C}_l^{t+1}} \right) - \left(\bigcup_{l:\mathcal{C}_l^{t+1} \cap \mathcal{U}_{\mathbf{d}} = \emptyset, \mathcal{C}_l^{t+1} \cap [k-1] = \emptyset, k \in \mathcal{C}_l^{t+1}} Q_{\mathcal{C}_l^{t+1}} \right). \tag{21}$$

For each user $k \in [2:K]$, there are $\binom{K-k}{t}$ different $(t+1)$-element subsets $\mathcal{C}_l^{t+1}$, such that $\mathcal{C}_l^{t+1} \cap [k-1] = \emptyset$ and $k \in \mathcal{C}_l^{t+1}$, for $l = 1, \ldots, \binom{K}{t+1}$. On the other hand, for each user $k \in [2:K] \backslash \mathcal{U}_{\mathbf{d}}$, since there are $N_{\mathbf{d},k}$ users among the set of users $[k:K]$ that belong to set $\mathcal{U}_{\mathbf{d}}$, there are $\binom{K-k-N_{\mathbf{d},k}}{t}$ different $(t+1)$-element subsets $\mathcal{C}_l^{t+1}$, such that $\mathcal{C}_l^{t+1} \cap (\mathcal{U}_{\mathbf{d}} \cup [k-1]) = \emptyset$ and $k \in \mathcal{C}_l^{t+1}$, for $l = 1, \ldots, \binom{K}{t+1}$. Note that, if $k \in \mathcal{U}_{\mathbf{d}}$, the second term on the RHS of (21)

Algorithm 1 Centralized Caching and Coded Packet Generation

1: **procedure** PLACEMENT PHASE
2: $\quad W_i = \bigcup_{l=1}^{\binom{K}{t}} W_{i,\mathcal{C}_l^t}, \quad$ for $i = 1, \ldots, N$

3: $\quad U_k = \bigcup_{i \in [N]} \bigcup_{l \in [\binom{K}{t}]:k \in \mathcal{C}_l^t} W_{i,\mathcal{C}_l^t}, \quad$ for $k = 1, \ldots, K$
4: **end procedure**

5: **procedure** CODED PACKET GENERATION
6: $\quad Q_{\mathcal{C}_l^{t+1}} = \bigoplus_{k \in \mathcal{C}_l^{t+1}} W_{d_k, \mathcal{C}_l^{t+1} \backslash \{k\}}, \quad$ for $l = 1, \ldots, \binom{K}{t+1}$

7: $\quad \tilde{Q}_k = \bigcup_{l:\mathcal{C}_l^{t+1} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, \mathcal{C}_l^{t+1} \cap [k-1] = \emptyset, k \in \mathcal{C}_l^{t+1}} Q_{\mathcal{C}_l^{t+1}}, \quad$ for $k = 1, \ldots, K$
8: **end procedure**

---

includes no content. Thus, if $k \in [2:K] \cap \mathcal{U}_{\mathbf{d}}$, total rate targeted for user k is

$$R_{\mathbf{d},t,k} = \frac{\binom{K-k}{t}}{\binom{K}{t}} R = \left(\prod_{i=0}^{k-1} \frac{K-t-i}{K-i} \right) R, \tag{22}$$

while the total rate targeted for user k, $k \in [2:K] \backslash \mathcal{U}_{\mathbf{d}}$, is

$$R_{\mathbf{d},t,k} = \frac{\binom{K-k}{t} - \binom{K-k-N_{\mathbf{d},k}}{t+1}}{\binom{K}{t}} R. \tag{23}$$

In summary, the proposed achievable scheme intends to deliver contents of total rate $R_{\mathbf{d},t,k}$ to user k, for $k \in [K]$, where

$$R_{\mathbf{d},t,k} \triangleq \begin{cases} \frac{\binom{K-k}{t}}{\binom{K}{t}} R, & \text{if } k \in \mathcal{U}_{\mathbf{d}}, \\ \frac{\binom{K-k}{t} - \binom{K-k-N_{\mathbf{d},k}}{t}}{\binom{K}{t}} R, & \text{otherwise.} \end{cases} \tag{24}$$

The centralized caching and coded packet generation explained above is summarized in Algorithm 1.

Once the coded packets targeted to each user are determined, the next step is to design the physical layer coding scheme to deliver these packets over the Gaussian BC. Given $\mathbf{d}$, we generate K codebooks, where the codebook $k \in [K]$ is designed to deliver the contents $\tilde{Q}_k$ to user k. The k-th codebook consists of $2^{nR_{\mathbf{d},t,k}}$ i.i.d. Gaussian codewords $x_k^n(\mathbf{W}, \mathbf{d})$, generated according to the normal distribution $\mathcal{N}\left(0, \alpha_k P_{\text{UB}}^{\text{C}}(R, M, \mathbf{d})\right)$, where $\alpha_k \geq 0$ and $\sum_{i=1}^K \alpha_i = 1$. The transmission is performed through superposition coding; that is, the channel input is given by $\sum_{k=1}^K x_k^n(\mathbf{W}, \mathbf{d}, \tilde{q}_k)$, where $\tilde{q}_k \in \left[2^{nR_{\mathbf{d},t,k}^{\text{C}}}\right]$. User k decodes codewords $x_1^n(\mathbf{W}, \mathbf{d}), \ldots, x_k^n(\mathbf{W}, \mathbf{d})$ by successive decoding, considering all the codewords in higher levels as noise. If all the k codewords are decoded successfully, user $k \in [K]$ can recover contents $\tilde{Q}_1, \ldots, \tilde{Q}_k$. Accordingly, any user $k \in [K]$ will receive all the coded contents targeted for it except those that are not intended for at least one user in $\mathcal{U}_{\mathbf{d}}$ (which have not been delivered); that is, user k receives all the coded contents

$$\bigcup_{l:\mathcal{C}_l^{t+1} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, k \in \mathcal{C}_l^{t+1}} Q_{\mathcal{C}_l^{t+1}}. \tag{25}$$

Thanks to the degradedness of the underlying Gaussian BC, it can also obtain all the coded contents targeted for users in

$[k-1]$, which are also intended for at least one user in $\mathcal{U}_{\mathbf{d}}$, i.e., all the coded contents in

$$\bigcup_{l:\mathcal{C}_l^{t+1}\cap\mathcal{U}_{\mathbf{d}}\neq\emptyset,[k-1]\cap\mathcal{C}_l^{t+1}\neq\emptyset} Q_{\mathcal{C}_l^{t+1}}. \tag{26}$$

Note that, if $k\in\mathcal{U}_{\mathbf{d}}$, (25) reduces to $\bigcup_{l:k\in\mathcal{C}_l^{t+1}} Q_{\mathcal{C}_l^{t+1}}$, which shows that the demand of user $k\in\mathcal{U}_{\mathbf{d}}$ is satisfied. Next, we illustrate that the users in $[K]\backslash\mathcal{U}_{\mathbf{d}}$ can obtain their requests without being delivered any extra messages. Given any set of users $\mathcal{C}_l^{t+1}$ such that $\mathcal{C}_l^{t+1}\cap\mathcal{U}_{\mathbf{d}}=\emptyset$, we need to show that every user in $\mathcal{C}_l^{t+1}$ can decode all the coded messages $Q_{\mathcal{B}\backslash\mathcal{G}}$, $\forall\mathcal{G}\in\mathcal{G}_{\mathcal{B}}\backslash\mathcal{U}_{\mathbf{d}}$, where $\mathcal{B}=\mathcal{C}_l^{t+1}\cup\mathcal{U}_{\mathbf{d}}$. In this case, they can also decode $Q_{\mathcal{C}_l^{t+1}}$ through (16).

Assume that there exists a set of users $\mathcal{C}_l^{t+1}$ such that $\mathcal{C}_l^{t+1}\cap\mathcal{U}_{\mathbf{d}}=\emptyset$; set $\mathcal{B}=\mathcal{C}_l^{t+1}\cup\mathcal{U}_{\mathbf{d}}$. According to (17b), there is at least one user in $\mathcal{U}_{\mathbf{d}}$ in any set of users $\mathcal{B}\backslash\mathcal{G}$, $\forall\mathcal{G}\in\mathcal{G}_{\mathcal{B}}\backslash\mathcal{U}_{\mathbf{d}}$. Thus, all the coded messages $Q_{\mathcal{B}\backslash\mathcal{G}}$ have been delivered by the proposed delivery scheme. Remember that, for each user $k\in\mathcal{C}_l^{t+1}$ and $\forall\mathcal{G}\in\mathcal{G}_{\mathcal{B}}\backslash\mathcal{U}_{\mathbf{d}}$, either $k\in\mathcal{B}\backslash\mathcal{G}$ or $\exists k'\in\mathcal{B}\backslash\mathcal{G}$, where $d_{k'}=d_k$ and $k'\in\mathcal{U}_{\mathbf{d}}$, i.e., $h_{k'}^2\leq h_k^2$. If $k\in\mathcal{B}\backslash\mathcal{G}$, since $\mathcal{B}\backslash\mathcal{G}\cap\mathcal{U}_{\mathbf{d}}\neq\emptyset$, according to (25), user k can obtain $Q_{\mathcal{B}\backslash\mathcal{G}}$. If $\exists k'\in\mathcal{B}\backslash\mathcal{G}$ with $d_{k'}=d_k$ and $k'\in\mathcal{U}_{\mathbf{d}}$, then user k' can decode $Q_{\mathcal{B}\backslash\mathcal{G}}$, and since $h_{k'}^2\leq h_k^2$, user k can also decode $Q_{\mathcal{B}\backslash\mathcal{G}}$. Thus, each user $k\in\mathcal{C}_l^{t+1}$ can decode $Q_{\mathcal{C}_l^{t+1}}$ successfully, for any set of users $\mathcal{C}_l^{t+1}$ that satisfies $\mathcal{C}_l^{t+1}\cap\mathcal{U}_{\mathbf{d}}=\emptyset$. This fact confirms that the demands of all the users in $[K]\backslash\mathcal{U}_{\mathbf{d}}$ are satisfied.

We remark here that, with the proposed coded delivery scheme, the coded packets targeted to user k are delivered only to users in $[k]$, and not to the any of the users in $[k+1:K]$, for $k\in[K]$. Next, we provide an example to illustrate the proposed joint cache and channel coding scheme.

Example: Consider $K=5$ users, each equipped with a cache of normalized size $M=N/5$, i.e., $t=1$. File W_i is partitioned into 5 equal-length subfiles $W_{i,\{1\}},\ldots,W_{i,\{5\}}$, each of which is of rate $R/5$, for $i\in[N]$. Cache contents of user k after the placement phase is given by

$$U_k=\bigcup_{i\in[N]} W_{i,\{k\}},\quad k\in[5]. \tag{27}$$

Assuming that $N\geq 3$, let the user demands be $\mathbf{d}=(1,2,1,1,3)$, where we have $N_{\mathbf{d}}=3$, and $\mathcal{U}_{\mathbf{d}}=\{1,2,5\}$. We generate the following coded packets:

$$Q_{\{1,2\}}=W_{1,\{2\}}\oplus W_{2,\{1\}}, \tag{28a}$$
$$Q_{\{1,3\}}=W_{1,\{3\}}\oplus W_{1,\{1\}}, \tag{28b}$$
$$Q_{\{1,4\}}=W_{1,\{4\}}\oplus W_{1,\{1\}}, \tag{28c}$$
$$Q_{\{1,5\}}=W_{1,\{5\}}\oplus W_{3,\{1\}}, \tag{28d}$$
$$Q_{\{2,3\}}=W_{2,\{3\}}\oplus W_{1,\{2\}}, \tag{28e}$$
$$Q_{\{2,4\}}=W_{2,\{4\}}\oplus W_{1,\{2\}}, \tag{28f}$$
$$Q_{\{2,5\}}=W_{2,\{5\}}\oplus W_{3,\{2\}}, \tag{28g}$$
$$Q_{\{3,5\}}=W_{1,\{5\}}\oplus W_{3,\{3\}}, \tag{28h}$$
$$Q_{\{4,5\}}=W_{1,\{5\}}\oplus W_{3,\{4\}}. \tag{28i}$$

The coded contents $\tilde{Q}_k$ targeted to user k, $k\in[5]$, are:

$$\tilde{Q}_1=Q_{\{1,2\}},Q_{\{1,3\}},Q_{\{1,4\}},Q_{\{1,5\}}, \tag{29a}$$
$$\tilde{Q}_2=Q_{\{2,3\}},Q_{\{2,4\}},Q_{\{2,5\}}, \tag{29b}$$

$$\tilde{Q}_3=Q_{\{3,5\}}, \tag{29c}$$
$$\tilde{Q}_4=Q_{\{4,5\}}, \tag{29d}$$
$$\tilde{Q}_5=\mathbf{0}. \tag{29e}$$

We perform superposition coding to deliver these contents, and send $\sum_{k=1}^4 x_k^n(\mathbf{W},\mathbf{d},\tilde{q}_k)$, where $\tilde{q}_1\in[2^{4nR/5}]$, $\tilde{q}_2\in[2^{3nR/5}]$, $\tilde{q}_3\in[2^{nR/5}]$, and $\tilde{q}_4\in[2^{nR/5}]$. User k, $k\in[5]$, decodes the codewords $x_1^n(\mathbf{W},\mathbf{d}),\ldots,x_k^n(\mathbf{W},\mathbf{d})$ by successive decoding, considering the higher level codewords as noise. If successful, user k can recover $\tilde{Q}_1,\ldots,\tilde{Q}_k$, $k\in[5]$. Now, note that user 1 can recover W_1 from $\tilde{Q}_1$ together with its cache contents; user 2 can decode W_2 having received the coded packets $Q_{\{1,2\}}$, $Q_{\{2,3\}}$, $Q_{\{2,4\}}$, and $Q_{\{2,5\}}$ along with its cache contents; coded packets $Q_{\{1,5\}}$, $Q_{\{2,5\}}$, $Q_{\{3,5\}}$, and $Q_{\{4,5\}}$, and its cache contents U_5 enable user 5 to decode W_3; user 3 can directly obtain $Q_{\{1,3\}}$, $Q_{\{2,3\}}$, and $Q_{\{3,5\}}$, and it can generate $Q_{\{3,4\}}$ as follows:

$$Q_{\{3,4\}}=Q_{\{1,3\}}\oplus Q_{\{1,4\}}, \tag{30}$$

so it can decode W_1 together with its cache contents. Note that both $Q_{\{1,3\}}$ and $Q_{\{1,4\}}$ are delivered within $\tilde{Q}_1$, which is decoded by user 3. Similarly, user 4 will be delivered $Q_{\{1,4\}}$, $Q_{\{2,4\}}$, and $Q_{\{4,5\}}$, and it can also recover $Q_{\{3,4\}}$ through (30), and together with its cache contents it can recover W_1. □

B. Non-integer t Values

Here we extend the proposed scheme to non-integer t values. We divide the whole database as well as the cache memory of the users into two, such that the corresponding t parameters for both parts are integer. This way we can employ the placement and delivery schemes introduced in Section III-A for each part separately.

Placement phase: Each file W_i, for $i=1,...,N$, is divided into two non-overlapping subfiles, W_i^1 of rate R_1 and W_i^2 of rate R_2. We set

$$R_1=(\lfloor t\rfloor+1-t)R, \tag{31a}$$
$$R_2=(t-\lfloor t\rfloor)R, \tag{31b}$$

such that $R_1+R_2=R$. For subfiles $\{W_1^1,\ldots,W_N^1\}$, we perform the placement phase proposed in Section III-A corresponding to the normalized global cache capacity $\lfloor t\rfloor$, which requires a cache capacity of $n(\lfloor t\rfloor N/K)R_1$ bits. While, for subfiles $\{W_1^2,\ldots,W_N^2\}$, we perform the placement phase proposed in Section III-A corresponding to the normalized global cache capacity $\lfloor t\rfloor+1$, which requires a cache capacity of $n((\lfloor t\rfloor+1)N/K)R_2$ bits. By summing up, the total cache capacity is found to be nMR bits, which shows that the cache capacity constraint is satisfied.

Delivery phase: For any demand vector $\mathbf{d}$, we perform the proposed coded delivery phase in Section III-A corresponding to the normalized global cache capacity $\lfloor t\rfloor$ to deliver the missing bits of subfiles $\{W_1^1,\ldots,W_N^1\}$ to the intended users. Moreover, the missing bits of subfiles $\{W_1^2,\ldots,W_N^2\}$ are delivered to the intended users by performing the coded

delivery scheme proposed in Section III-A corresponding to the normalized global cache capacity $\lfloor t \rfloor + 1$.

For an arbitrary demand vector $\mathbf{d} = (d_1, \ldots, d_K)$, we define

$$Q_{\mathcal{C}_l^{t_i+1}}^i \triangleq \bigoplus_{k \in \mathcal{C}_l^{t_i+1}} W_{d_k, \mathcal{C}_l^{t_i+1} \backslash \{k\}}^i, \quad \text{for } i = 1, 2, \quad (32)$$

where $t_1 \triangleq \lfloor t \rfloor$ and $t_2 \triangleq \lfloor t \rfloor + 1$. According to Algorithm 1, by performing the centralized caching and coded delivery scheme proposed in Section III-A for the normalized global cache capacity $\lfloor t \rfloor$ to serve the subfiles $\{W_1^1, \ldots, W_N^1\}$, each of rate R_1, user $k \in [K]$ should receive

$$\tilde{Q}_k^1 = \bigcup_{l: \mathcal{C}_l^{\lfloor t \rfloor + 1} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, \mathcal{C}_l^{\lfloor t \rfloor + 1} \cap [k-1] = \emptyset, k \in \mathcal{C}_l^{\lfloor t \rfloor + 1}} Q_{\mathcal{C}_l^{\lfloor t \rfloor + 1}}^1, \quad (33)$$

of the following total rate obtained according to (24)

$$R_{\mathbf{d}, \lfloor t \rfloor, k} = \begin{cases} \dfrac{\binom{K-k}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}} R_1, & \text{if } k \in \mathcal{U}_{\mathbf{d}}, \\ \dfrac{\binom{K-k}{\lfloor t \rfloor} - \binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}} R_1, & \text{otherwise.} \end{cases} \quad (34)$$

Similarly, to serve the subfiles $\{W_1^2, \ldots, W_N^2\}$, each of rate R_2, user $k \in [K]$ should receive

$$\tilde{Q}_k^2 = \bigcup_{l: \mathcal{C}_l^{\lfloor t \rfloor + 2} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, \mathcal{C}_l^{\lfloor t \rfloor + 2} \cap [k-1] = \emptyset, k \in \mathcal{C}_l^{\lfloor t \rfloor + 2}} Q_{\mathcal{C}_l^{\lfloor t \rfloor + 2}}^2, \quad (35)$$

of total rate

$$R_{\mathbf{d}, \lfloor t \rfloor + 1, k} = \begin{cases} \dfrac{\binom{K-k}{\lfloor t \rfloor + 1}}{\binom{K}{\lfloor t \rfloor + 1}} R_2, & \text{if } k \in \mathcal{U}_{\mathbf{d}}, \\ \dfrac{\binom{K-k}{\lfloor t \rfloor + 1} - \binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor + 1}}{\binom{K}{\lfloor t \rfloor + 1}} R_2, & \text{otherwise.} \end{cases} \quad (36)$$

Thus, $\tilde{Q}_k \triangleq \left(\tilde{Q}_k^1, \tilde{Q}_k^2 \right)$, at a total rate of $R_{\mathbf{d},k}^{\mathrm{C}} \triangleq R_{\mathbf{d}, \lfloor t \rfloor, k} + R_{\mathbf{d}, \lfloor t \rfloor + 1, k}$ given by

$$R_{\mathbf{d},k}^{\mathrm{C}} = \begin{cases} \dfrac{\binom{K-k}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}} (\lfloor t \rfloor + 1 - t) R + \\ \quad \dfrac{\binom{K-k}{\lfloor t \rfloor + 1}}{\binom{K}{\lfloor t \rfloor + 1}} (t - \lfloor t \rfloor) R, & \text{if } k \in \mathcal{U}_{\mathbf{d}}, \\[2mm] \dfrac{\binom{K-k}{\lfloor t \rfloor} - \binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}} (\lfloor t \rfloor + 1 - t) R + \\ \quad \dfrac{\binom{K-k}{\lfloor t \rfloor + 1} - \binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor + 1}}{\binom{K}{\lfloor t \rfloor + 1}} (t - \lfloor t \rfloor) R, & \text{otherwise} \end{cases} \quad (37)$$

is delivered to user k, for $k = 1, \ldots, K$.

C. Transmit Power Analysis

In the proposed delivery scheme, user $k \in [K]$ decodes codewords $x_1^n(\mathbf{W}, \mathbf{d}), \ldots, x_k^n(\mathbf{W}, \mathbf{d})$ successively considering all the other codewords in higher levels as noise. User k can decode its intended message successfully if, for $k = 1, \ldots, K$,

$$R_{\mathbf{d},k}^{\mathrm{C}} \leq \frac{1}{2} \log_2 \left(1 + \frac{\alpha_k h_k^2 P_{\mathrm{UB}}^{\mathrm{C}}(R, M, \mathbf{d})}{h_k^2 \sum_{i=k+1}^K \alpha_i P_{\mathrm{UB}}^{\mathrm{C}}(R, M, \mathbf{d}) + 1} \right). \quad (38)$$

From Proposition 1, the corresponding minimum required power is given by

$$P_{\mathrm{UB}}^{\mathrm{C}}(R, M, \mathbf{d}) \triangleq \sum_{i=1}^K \left(\frac{2^{2R_{\mathbf{d},i}^{\mathrm{C}}} - 1}{h_i^2} \right) \prod_{j=1}^{i-1} 2^{2R_{\mathbf{d},j}^{\mathrm{C}}}. \quad (39)$$

Thus, the average power-memory trade-off of the proposed achievable scheme is given by $\mathbb{E}_{\mathbf{d}} \left[P_{\mathrm{UB}}^{\mathrm{C}}(R, M, \mathbf{d}) \right] = \bar{P}_{\mathrm{UB}}^{\mathrm{C}}(R, M)$, while the peak power-memory trade-off is $\hat{P}_{\mathrm{UB}}^{\mathrm{C}}(R, M) = \max_{\mathbf{d}} \left\{ P_{\mathrm{UB}}^{\mathrm{C}}(R, M, \mathbf{d}) \right\}$, as stated in Theorem 1, where the demands are distributed uniformly.

Observe that, for demand vectors with the same $\mathcal{U}_{\mathbf{d}}$ set, the required power $P_{\mathrm{UB}}^{\mathrm{C}}(R, M, \mathbf{d})$ is the same. Let $\mathcal{D}_{\mathcal{U}_{\mathbf{d}}}$ denote set of all demand vectors with the same $\mathcal{U}_{\mathbf{d}}$ set. We define $P_{\mathrm{UB}}^{\mathrm{C}}(R, M, \mathcal{D}_{\mathcal{U}_{\mathbf{d}}})$ as the required power $P_{\mathrm{UB}}^{\mathrm{C}}(R, M, \mathbf{d})$ for any demand vector $\mathbf{d} \in \mathcal{D}_{\mathcal{U}_{\mathbf{d}}}$. Thus, we have

$$\hat{P}_{\mathrm{UB}}^{\mathrm{C}}(R, M) = \max_{\mathcal{U}_{\mathbf{d}}} \left\{ P_{\mathrm{UB}}^{\mathrm{C}}(R, M, \mathcal{D}_{\mathcal{U}_{\mathbf{d}}}) \right\}. \quad (40)$$

It is shown in Appendix B that the worst-case demand combination for the proposed centralized caching scheme happens when the first $\min\{N, K\}$ users; that is, the users with the worst channel gains, request distinct files, i.e., when $\mathcal{U}_{\mathbf{d}} = [\min\{N, K\}]$, and $\hat{P}_{\mathrm{UB}}^{\mathrm{C}}(R, M)$ is given by (12). $\square$

IV. DECENTRALIZED CACHING AND DELIVERY

Here we extend our centralized caching scheme to the decentralized caching. The corresponding average and peak power-memory trade-offs are given in the following theorem.

Theorem 2. *For decentralized caching followed by delivery over a Gaussian BC, we have*

$$\bar{P}^*(R, M) \leq \bar{P}_{\mathrm{UB}}^{\mathrm{D}}(R, M) \triangleq$$
$$\frac{1}{N^K} \sum_{\mathbf{d} \in [N]^K} \left[\sum_{i=1}^K \left(\frac{2^{2R_{\mathbf{d},i}^{\mathrm{D}}} - 1}{h_i^2} \right) \prod_{j=1}^{i-1} 2^{2R_{\mathbf{d},j}^{\mathrm{D}}} \right], \quad (41a)$$

where, for $k = 1, \ldots, K$,

$$R_{\mathbf{d},k}^{\mathrm{D}} \triangleq \begin{cases} \left(1 - \frac{M}{N}\right)^k R, & \text{if } k \in \mathcal{U}_{\mathbf{d}}, \\ \left(1 - \frac{M}{N}\right)^k \left(1 - \left(1 - \frac{M}{N}\right)^{N_{\mathbf{d},k}}\right) R, & \text{otherwise,} \end{cases} \quad (41b)$$

and

$$\hat{P}^*(R, M) \leq \hat{P}_{\mathrm{UB}}^{\mathrm{D}}(R, M) \triangleq \sum_{i=1}^{\min\{N,K\}} \left(\frac{2^{2R\left(1 - \frac{M}{N}\right)^i} - 1}{h_i^2} \right)$$
$$2^{2R\left(\frac{N}{M} - 1\right)\left(1 - \left(1 - \frac{M}{N}\right)^{i-1}\right)}. \quad (42)$$

Proof. The decentralized caching scheme to achieve the average and peak power-memory trade-offs outlined in Theorem 2 is described in the following.

A. Placement Phase

We perform decentralized uncoded cache placement [17], where each user caches nMR/N random bits of each file of length nR bits independently. Since there are a total of N files, the cache capacity constraint is satisfied. The part of file i cached exclusively by the users in set $\mathcal{S} \subset [K]$ is denoted by $W_{i,\mathcal{S}}$, for $i = 1, \ldots, N$. For n large enough, the rate of $W_{i,\mathcal{S}}$ can be approximated by $\left(\frac{M}{N}\right)^{|\mathcal{S}|} \left(1 - \frac{M}{N}\right)^{K-|\mathcal{S}|} R$. The cache contents at user k is given by

$$U_k = \bigcup_{i \in [N]} \bigcup_{\mathcal{S} \subset [K]: k \in \mathcal{S}} W_{i,\mathcal{S}}. \quad (43)$$

B. Delivery phase

Consider any non-empty set of users $\mathcal{S} \subset [K]$. For a demand vector $\mathbf{d}$, by delivering the coded message $Q_{\mathcal{S}} = \bigoplus_{k \in \mathcal{S}} W_{d_k, \mathcal{S} \setminus \{k\}}$ of rate $\left(\frac{M}{N}\right)^{|\mathcal{S}|-1} \left(1 - \frac{M}{N}\right)^{K-|\mathcal{S}|+1} R$ to users in $\mathcal{S}$, each user $i \in \mathcal{S}$ can recover subfile $W_{d_i, \mathcal{S} \setminus \{i\}}$ since it has cached all the subfiles $W_{d_j, \mathcal{S} \setminus \{j\}}$, $\forall j \in \mathcal{S} \setminus \{i\}$. For each $k \in [K]$, delivering $\bigcup_{\mathcal{S} \subset [K]:k \in \mathcal{S}} Q_{\mathcal{S}}$ enables user k to recover all the subfiles $W_{d_k, \mathcal{S} \setminus \{k\}}$, $\forall \mathcal{S} \subset [K]$ and $k \in \mathcal{S}$. The demand of user k, $k \in [K]$, is fully satisfied after receiving $\bigcup_{\mathcal{S} \subset [K]:k \in \mathcal{S}} Q_{\mathcal{S}}$ along with its cache contents.

Similarly to the proposed scheme for the centralized caching scenario, given a demand vector $\mathbf{d}$, the delivery phase is designed such that only the coded messages $Q_{\mathcal{S}}$, $\forall \mathcal{S} \subset [K]$ that satisfy $\mathcal{S} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset$, are delivered, and the remaining coded messages can be recovered through (16). To achieve this, for any such set $\mathcal{S}$ with $\mathcal{S} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset$, the transmission power is adjusted such that the worst user in $\mathcal{S}$ can decode it; and so can all the other users in $\mathcal{S}$ due to the degradedness of the Gaussian BC. Therefore, the demand of every user in $\mathcal{U}_{\mathbf{d}}$ is satisfied.

Note that in the centralized scenario described in Section III-A, each coded packet $Q_{\mathcal{C}_l^{t+1}}$ is targeted for a $(t+1)$-element subset of users, where $t \in [0:K]$, for $l \in \left[\binom{K}{t+1}\right]$. While, in the decentralized scenario, coded packets are targeted for any subset of users, i.e., for $(t+1)$-element subset of users, $\forall t \in [0:K]$. By applying a similar technique as the delivery phase outlined in Algorithm 1, given a demand vector $\mathbf{d}$, the following contents are targeted for user k, $k \in [K]$:

$$\tilde{Q}_k = \bigcup_{\mathcal{S} \subset [k:K]:\mathcal{S} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, k \in \mathcal{S}} Q_{\mathcal{S}}, \tag{44}$$

which are equivalent to

$$\tilde{Q}_k = \bigcup_{\mathcal{S} \subset [k:K]:k \in \mathcal{S}} Q_{\mathcal{S}} - \bigcup_{\mathcal{S} \subset [k:K] \setminus \mathcal{U}_{\mathbf{d},k}, k \in \mathcal{S}} Q_{\mathcal{S}}. \tag{45}$$

For each user k, $k \in [K]$, there are $\binom{K-k}{i}$ different $(i+1)$-element subsets of $[k:K]$, which include k, for $i \in [0:K-k]$. Thus, if $k \in \mathcal{U}_{\mathbf{d}}$, from (45), the total rate targeted for user k is

$$R_{\mathbf{d},k}^{\mathrm{D}} = \sum_{i=0}^{K-k} \binom{K-k}{i} \left(\frac{M}{N}\right)^i \left(1 - \frac{M}{N}\right)^{K-i} R$$
$$= \left(1 - \frac{M}{N}\right)^k R. \tag{46}$$

On the other hand, for each user k, $k \in [K]$, there are $\binom{K-k-N_{\mathbf{d},k}}{i}$ different $(i+1)$-element subsets of $[k:K] \setminus \mathcal{U}_{\mathbf{d},k}$, which include k, for $i \in [0:K-k]$. Thus, if $k \notin \mathcal{U}_{\mathbf{d}}$, from (45), the total rate targeted for user $k \in [K]$ is given by

$$R_{\mathbf{d},k}^{\mathrm{D}} = \sum_{i=0}^{K-k} \binom{K-k}{i} \left(\frac{M}{N}\right)^i \left(1 - \frac{M}{N}\right)^{K-i} R$$
$$- \sum_{i=0}^{K-k-N_{\mathbf{d},k}} \binom{K-k-N_{\mathbf{d},k}}{i} \left(\frac{M}{N}\right)^i \left(1 - \frac{M}{N}\right)^{K-i} R$$
$$= \left(1 - \frac{M}{N}\right)^k \left(1 - \left(1 - \frac{M}{N}\right)^{N_{\mathbf{d},k}}\right) R. \tag{47}$$

In total, the rate of contents targeted for user k, $k \in [K]$, is given by

$$R_{\mathbf{d},k}^{\mathrm{D}} = \begin{cases} \left(1 - \frac{M}{N}\right)^k R, & \text{if } k \in \mathcal{U}_{\mathbf{d}}, \\ \left(1 - \frac{M}{N}\right)^k \left(1 - \left(1 - \frac{M}{N}\right)^{N_{\mathbf{d},k}}\right) R, & \text{otherwise.} \end{cases} \tag{48}$$

Given a demand vector $\mathbf{d}$, the transmitted codeword $x^n(\mathbf{W}, \mathbf{d})$ is generated as the linear superposition of K codewords $x_1^n(\mathbf{W}, \mathbf{d}), ..., x_K^n(\mathbf{W}, \mathbf{d})$, each chosen from an independent codebook. Codebook k consists of $2^{nR_{\mathbf{d},k}^{\mathrm{D}}}$ i.i.d. codewords $x_k^n(\mathbf{W}, \mathbf{d})$ generated according to the normal distribution $\mathcal{N}\left(0, \alpha_k P_{\mathrm{UB}}^{\mathrm{D}}(R, M, \mathbf{d})\right)$, where $\alpha_k \geq 0$ and $\sum_{i=1}^K \alpha_i = 1$, which satisfy the power constraint, for $k = 1, ..., K$.

User k, $k \in [K]$, decodes codewords $x_1^n(\mathbf{W}, \mathbf{d})$, ..., $x_k^n(\mathbf{W}, \mathbf{d})$ through successive decoding, while considering all the codewords $x_{k+1}^n(\mathbf{W}, \mathbf{d}), ..., x_K^n(\mathbf{W}, \mathbf{d})$ as noise. Thus, if user $k \in [K]$ can successfully decode all the k channel codewords intended for it, it can then recover the contents $\tilde{Q}_1, ..., \tilde{Q}_k$, from which it can obtain all the coded contents $\bigcup_{\mathcal{S} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, [k] \cap \mathcal{S} \neq \emptyset} Q_{\mathcal{S}}$. Accordingly, each user $k \in [K]$ can obtain all the coded contents targeted for it except those that are not intended for at least one user in $\mathcal{U}_{\mathbf{d}}$ (which have not been delivered), i.e., all the coded contents

$$\bigcup_{\mathcal{S} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, k \in \mathcal{S}} Q_{\mathcal{S}}. \tag{49}$$

It can further obtain all the coded contents targeted for users $[k-1]$, which are also intended for at least one user in $\mathcal{U}_{\mathbf{d}}$, i.e., all the coded contents $\bigcup_{\mathcal{S} \cap \mathcal{U}_{\mathbf{d}} \neq \emptyset, [k-1] \cap \mathcal{S} \neq \emptyset} Q_{\mathcal{S}}$. Note that, if $k \in \mathcal{U}_{\mathbf{d}}$, (49) reduces to $\bigcup_{k \in \mathcal{S}} Q_{\mathcal{S}}$, which shows that the demand of each user $k \in \mathcal{U}_{\mathbf{d}}$ is satisfied through the proposed delivery scheme. Next, we illustrate that the users in $[K] \setminus \mathcal{U}_{\mathbf{d}}$ can obtain their requested files without being delivered any extra messages. Similarly to the centralized scenario, given any set of users $\mathcal{S}$ such that $\mathcal{S} \cap \mathcal{U}_{\mathbf{d}} = \emptyset$, by setting $\mathcal{B} = \mathcal{S} \cup \mathcal{U}_{\mathbf{d}}$, from the fact that, for each user $k \in \mathcal{S}$, $k \in \mathcal{B} \setminus \mathcal{G}$ or $k' \in \mathcal{B} \setminus \mathcal{G}$, where $k' < k$, it can be illustrated that every user in $\mathcal{S}$ can decode all coded contents $Q_{\mathcal{B} \setminus \mathcal{G}}$, $\forall \mathcal{G} \in \mathcal{G}_{\mathcal{B}} \setminus \mathcal{U}_{\mathbf{d}}$. In this case, they all can also decode $Q_{\mathcal{S}}$ through (16).

C. Transmit Power Analysis

For a demand vector $\mathbf{d}$, user k can decode the channel codewords up to level k successfully, considering all the other codewords in higher levels as noise, if, for $k = 1, ..., K$,

$$R_{\mathbf{d},k}^{\mathrm{D}} \leq \frac{1}{2}\log_2\left(1 + \frac{\alpha_k h_k^2 P_{\mathrm{UB}}^{\mathrm{D}}(R, M, \mathbf{d})}{h_k^2 \sum_{i=k+1}^K \alpha_i P_{\mathrm{UB}}^{\mathrm{D}}(R, M, \mathbf{d}) + 1}\right). \tag{50}$$

From Proposition 1, the minimum required power is given by

$$P_{\mathrm{UB}}^{\mathrm{D}}(R, M, \mathbf{d}) \triangleq \sum_{i=1}^K \left(\frac{2^{2R_{\mathbf{d},i}^{\mathrm{D}}} - 1}{h_i^2}\right) \prod_{j=1}^{i-1} 2^{2R_{\mathbf{d},j}^{\mathrm{D}}}. \tag{51}$$

Thus, the average power-memory trade-off for the proposed decentralized caching and coded delivery scheme is given by $\mathrm{E}_{\mathbf{d}}\left[P_{\mathrm{UB}}^{\mathrm{D}}(R, M, \mathbf{d})\right] = \bar{P}_{\mathrm{UB}}^{\mathrm{D}}(R, M)$ stated in Theorem 2.

With the proposed decentralized caching scheme, the peak power performance $\hat{P}^{\mathrm{D}}(R, M) = \max_{\mathbf{d}}\left\{P_{\mathrm{UB}}^{\mathrm{D}}(R, M, \mathbf{d})\right\}$ can

be achieved. Observe that, for demand vectors with the same set of users $\mathcal{U}_\mathbf{d}$, the required power $P_\text{UB}^\text{D}(R, M, \mathbf{d})$ is the same. We define $P_\text{UB}^\text{D}(R, M, \mathcal{D}_{\mathcal{U}_\mathbf{d}})$ as the required power $P_\text{UB}^\text{D}(R, M, \mathbf{d})$ for any demand vector $\mathbf{d} \in \mathcal{D}_{\mathcal{U}_\mathbf{d}}$. Thus, we have

$$\hat{P}^\text{D}(R, M) = \max_{\mathcal{U}_\mathbf{d}} \left\{ P_\text{UB}^\text{D}(R, M, \mathcal{D}_{\mathcal{U}_\mathbf{d}}) \right\}. \tag{52}$$

It is shown in Appendix C that the worst-case demand combination happens when $\mathcal{U}_\mathbf{d} = [\min\{N, K\}]$, and $\hat{P}_\text{UB}^\text{D}(R, M)$ is found as in (42). $\qquad\square$

V. LOWER BOUND

If we constrain the placement phase to uncoded caching, we can lower bound $\bar{P}^*(R, M)$ and $\hat{P}^*(R, M)$ as in the following theorem. The main challenge in deriving a lower bound for the cache-aided BC studied here is the lack of degradedness due to the presence of the caches. To derive a lower bound, we assume that the files requested by users in $[k-1]$ and their cache contents are provided to the other users. We then exploit the degradedness of the resultant system to lower bound the performance of the original model.

Theorem 3. *In cache-aided content delivery over a Gaussian BC with uncoded cache placement phase, the minimum average power is lower bounded by $\bar{P}_\text{LB}(R, M)$ defined as*

$$\bar{P}_\text{LB}(R, M) \triangleq \mathrm{E}_{\mathcal{U}_\mathbf{d}} \left[\sum_{i=1}^{N_\mathbf{d}} \left(\frac{2^{2R(1-\min\{iM/N, 1\})} - 1}{h_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^2} \right) \right.$$
$$\left. \prod_{j=1}^{i-1} 2^{2R(1-\min\{jM/N, 1\})} \right], \tag{53}$$

where $\mathrm{E}_{\mathcal{U}_\mathbf{d}}[\cdot]$ takes the expectation over all possible sets $\mathcal{U}_\mathbf{d}$, and $\pi_\mathcal{S}$ is a permutation over any subset of users $\mathcal{S} \subset [K]$, such that $h_{\pi_\mathcal{S}(1)}^2 \leq h_{\pi_\mathcal{S}(2)}^2 \leq \cdots \leq h_{\pi_\mathcal{S}(|\mathcal{S}|)}^2$. The minimal required peak transmit power for the same system is lower bounded by $\hat{P}_\text{LB}(R, M)$ defined as

$$\hat{P}_\text{LB}(R, M) \triangleq \max_{\mathcal{S} \subset [\min\{N, K\}]} \left\{ \sum_{i=1}^{|\mathcal{S}|} \left(\frac{2^{2R(1-\min\{iM/N, 1\})} - 1}{h_{\pi_\mathcal{S}(i)}^2} \right) \right.$$
$$\left. \prod_{j=1}^{i-1} 2^{2R(1-\min\{jM/N, 1\})} \right\}. \tag{54}$$

Remark 3. *Considering all possible demand vectors, we have a total of 2^{K-1} different $\mathcal{U}_\mathbf{d}$ sets. This follows from the fact that $N_\mathbf{d} \leq K$ and $1 \in \mathcal{U}_\mathbf{d}$, $\forall \mathbf{d}$. For a given demand vector $\mathbf{d}$, let $\mathcal{U}_\mathbf{d} = \{u_1, u_2, ..., u_{N_\mathbf{d}}\}$, where $1 = u_1 \leq u_2 \leq \cdots \leq u_{N_\mathbf{d}}$. The number of demand vectors with the same $\mathcal{U}_\mathbf{d}$ is given by*

$$N_{\mathcal{U}_\mathbf{d}} \triangleq \binom{N}{N_\mathbf{d}} N_\mathbf{d}! \left(\prod_{j=2}^{N_\mathbf{d}} j^{u_{j+1} - u_j - 1} \right), \tag{55}$$

where we define $u_{N_\mathbf{d}+1} \triangleq K + 1$. Thus, the lower bound in (53) reduces to

$$\bar{P}_\text{LB}(R, M) = \frac{1}{N^K} \sum_{\mathcal{U}_\mathbf{d} \subset [K], 1 \in \mathcal{U}_\mathbf{d}} N_{\mathcal{U}_\mathbf{d}}$$
$$\sum_{i=1}^{N_\mathbf{d}} \left(\frac{2^{2R(1-\min\{iM/N, 1\})} - 1}{h_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^2} \right) \prod_{j=1}^{i-1} 2^{2R(1-\min\{jM/N, 1\})}. \tag{56}$$

Proof. For any given $\mathcal{U}_\mathbf{d}$, let $\mathcal{D}_{\mathcal{U}_\mathbf{d}}$ denote the set of all demand vectors with the same $\mathcal{U}_\mathbf{d}$. The union $\bigcup_{\mathcal{U}_\mathbf{d}} \mathcal{D}_{\mathcal{U}_\mathbf{d}}$ form the set of all possible demand vectors $[N]^K$. Therefore, the set of all possible demand vectors can be broken into classes $\mathcal{D}_{\mathcal{U}_\mathbf{d}}$ based on the $\mathcal{U}_\mathbf{d}$ set they correspond to.

For any given $\mathcal{U}_\mathbf{d}$, and an (n, R, M) code as defined in (1), (3) and (5) in Section II, define the error probability as follows:

$$P_{e_{\mathcal{U}_\mathbf{d}}} \triangleq \Pr \left\{ \bigcup_{\mathbf{d} \in \mathcal{D}_{\mathcal{U}_\mathbf{d}}} \bigcup_{k \in \mathcal{U}_\mathbf{d}} \left\{ \hat{W}_{d_k} \neq W_{d_k} \right\} \right\}. \tag{57}$$

Let $P_{\mathcal{U}_\mathbf{d}}(\mathbf{d})$ denote the average power of the codeword this code generates for a demand vector $\mathbf{d} \in \mathcal{D}_{\mathcal{U}_\mathbf{d}}$. We say that an $\left(R, M, \bar{P}, \hat{P}\right)$ tuple is $\mathcal{U}_\mathbf{d}$-*achievable* if for every $\varepsilon > 0$, there exists an (n, R, M) code with sufficiently large n, which satisfies $P_{e_{\mathcal{U}_\mathbf{d}}} < \varepsilon$, $\mathrm{E}_\mathbf{d}[P_{\mathcal{U}_\mathbf{d}}(\mathbf{d})] \leq \bar{P}$, and $P_{\mathcal{U}_\mathbf{d}}(\mathbf{d}) \leq \hat{P}$, $\forall \mathbf{d} \in \mathcal{D}_{\mathcal{U}_\mathbf{d}}$. We can also define $\bar{P}_{\mathcal{U}_\mathbf{d}}^*(R, M)$ and $\hat{P}_{\mathcal{U}_\mathbf{d}}^*(R, M)$ as in (7a) and (7b), respectively, by considering $\mathcal{U}_\mathbf{d}$-achievable codes.

We note from (57) that, a $\mathcal{U}_\mathbf{d}$-achievable code satisfies only the demands of the users in set $\mathcal{U}_\mathbf{d}$. Accordingly, an achievable $\left(R, M, \bar{P}, \hat{P}\right)$ tuple is also $\mathcal{U}_\mathbf{d}$-achievable, since $P_e \geq P_{e_{\mathcal{U}_\mathbf{d}}}$, for any $\mathcal{U}_\mathbf{d}$ set. Thus, lower bounds on $\bar{P}_{\mathcal{U}_\mathbf{d}}^*(R, M)$ and $\hat{P}_{\mathcal{U}_\mathbf{d}}^*(R, M)$ also serve as lower bounds on $\bar{P}^*(R, M)$ and $\hat{P}^*(R, M)$, respectively. In the following, we provide lower bounds on $\bar{P}_{\mathcal{U}_\mathbf{d}}^*(R, M)$ and $\hat{P}_{\mathcal{U}_\mathbf{d}}^*(R, M)$.

Let $\left(R, M, \bar{P}, \hat{P}\right)$ be any $\mathcal{U}_\mathbf{d}$-achievable tuple. For uniformly distributed demands, we have

$$\bar{P} \geq \mathrm{E}_\mathbf{d}[P_{\mathcal{U}_\mathbf{d}}(\mathbf{d})] = \mathrm{E}_{\mathcal{U}_\mathbf{d}} \left[\frac{1}{N_{\mathcal{U}_\mathbf{d}}} \sum_{\mathbf{d} \in \mathcal{D}_{\mathcal{U}_\mathbf{d}}} P_{\mathcal{U}_\mathbf{d}}(\mathbf{d}) \right], \tag{58}$$

where we used the fact that the probability of each demand vector in $\mathcal{D}_{\mathcal{U}_\mathbf{d}}$ is equal. We divide the set of demand vectors $\mathcal{D}_{\mathcal{U}_\mathbf{d}}$ into different subsets according to the demands of users in $\mathcal{U}_\mathbf{d}$, where each subset consists of the demand vectors for which the demands of all the users in $\mathcal{U}_\mathbf{d}$ are the same. Note that, there are $D \triangleq \binom{N}{N_\mathbf{d}} N_\mathbf{d}!$ such subsets[4], denoted by $\mathcal{D}_{\mathcal{U}_\mathbf{d}}^l$, for $l = 1, ..., D$, i.e., $\mathcal{D}_{\mathcal{U}_\mathbf{d}} = \bigcup_{l=1}^D \mathcal{D}_{\mathcal{U}_\mathbf{d}}^l$. We note that the number of demand vectors in each $\mathcal{D}_{\mathcal{U}_\mathbf{d}}^l$, denoted by $N_{\mathcal{U}_\mathbf{d}}'$, is the same, and is given by

$$N_{\mathcal{U}_\mathbf{d}}' = \prod_{j=2}^{N_\mathbf{d}} j^{u_{j+1} - u_j - 1}, \tag{59}$$

where, we remind that $\mathcal{U}_\mathbf{d} = \{u_1, u_2, ..., u_{N_\mathbf{d}}\}$, where $1 = u_1 \leq u_2 \leq \cdots \leq u_{N_\mathbf{d}}$. Thus, we have $N_{\mathcal{U}_\mathbf{d}} = D N_{\mathcal{U}_\mathbf{d}}'$, and (58) can be rewritten as follows:

$$\bar{P} \geq \mathrm{E}_{\mathcal{U}_\mathbf{d}} \left[\frac{1}{N_{\mathcal{U}_\mathbf{d}}} \sum_{\mathbf{d} \in \mathcal{D}_{\mathcal{U}_\mathbf{d}}} P_{\mathcal{U}_\mathbf{d}}(\mathbf{d}) \right]$$
$$= \mathrm{E}_{\mathcal{U}_\mathbf{d}} \left[\frac{1}{D} \sum_{l=1}^D \left(\frac{1}{N_{\mathcal{U}_\mathbf{d}}'} \sum_{\mathbf{d} \in \mathcal{D}_{\mathcal{U}_\mathbf{d}}^l} P_{\mathcal{U}_\mathbf{d}}(\mathbf{d}) \right) \right]. \tag{60}$$

For any arbitrary demand vector $\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l \in \mathcal{D}_{\mathcal{U}_\mathbf{d}}^l$, for $l \in [D]$, it is proved in [7, Lemma 14] that there exist random

---

[4]For simplicity, we drop the dependence of D on N and $N_\mathbf{d}$.

variables[5] $X\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)$, $Y_{\pi_{\mathcal{U}_\mathbf{d}}(1)}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right), \ldots, Y_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)$, and $\left\{V_1\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right), \ldots, V_{N_\mathbf{d}-1}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)\right\}$, where

$$V_1\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \to \cdots \to V_{N_\mathbf{d}-1}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \to X\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \to$$
$$Y_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \to \cdots \to Y_{\pi_{\mathcal{U}_\mathbf{d}}(1)}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \quad (61)$$

forms a Markov chain, and satisfy

$$R - \varepsilon_n \leq \frac{1}{n} I\left(W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(1)}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}\right)$$
$$+ I\left(V_{\mathcal{U}_\mathbf{d},1}; Y_{\pi_{\mathcal{U}_\mathbf{d}}(1)}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)\right), \quad (62a)$$

$$R - \varepsilon_n \leq \frac{1}{n} I\left(W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}, \ldots, U_{\pi_{\mathcal{U}_\mathbf{d}}(i)}\right.$$
$$\left.\Big| W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(1)}^l}, \ldots, W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(i-1)}^l}\right)$$
$$+ I\left(V_i\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right); Y_{\pi_{\mathcal{U}_\mathbf{d}}(i)}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \Big| V_{i-1}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)\right),$$
$$\forall i \in [2:N_\mathbf{d}-1], \quad (62b)$$

$$R - \varepsilon_n \leq \frac{1}{n} I\left(W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}, \ldots, U_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}\right.$$
$$\left.\Big| W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(1)}^l}, \ldots, W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d}-1)}^l}\right)$$
$$+ I\left(X\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right); Y_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \Big| V_{N_\mathbf{d}-1}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)\right),$$
$$\quad (62c)$$

where $d_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^l$ is the $\pi_{\mathcal{U}_\mathbf{d}}(i)$-th element of demand vector $\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l$, $i \in [N_\mathbf{d}]$, and $\varepsilon_n > 0$ tends to zeros as $n \to \infty$. We note that, due to the independence of the files and the fact that the users in $\mathcal{U}_\mathbf{d}$ demand distinct files, for any uncoded cache placement phase and any $\mathcal{U}_\mathbf{d}$ set, we have

$$I\left(W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}, \ldots, U_{\pi_{\mathcal{U}_\mathbf{d}}(i)} \Big| W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(1)}^l}, \ldots, W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(i-1)}^l}\right)$$
$$= I\left(W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}, \ldots, U_{\pi_{\mathcal{U}_\mathbf{d}}(i)}\right), \forall i \in [2:N_\mathbf{d}], l \in [D].$$
$$\quad (63)$$

Thus, for an uncoded cache placement phase, (62) is equivalent to

$$R - \varepsilon_n \leq \frac{1}{n} I\left(W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(1)}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}\right)$$
$$+ I\left(V_{\mathcal{U}_\mathbf{d},1}; Y_{\pi_{\mathcal{U}_\mathbf{d}}(1)}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)\right), \quad (64a)$$

$$R - \varepsilon_n \leq \frac{1}{n} I\left(W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}, \ldots, U_{\pi_{\mathcal{U}_\mathbf{d}}(i)}\right)$$
$$+ I\left(V_i\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right); Y_{\pi_{\mathcal{U}_\mathbf{d}}(i)}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \Big| V_{i-1}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)\right),$$
$$\forall i \in [2:N_\mathbf{d}-1], \quad (64b)$$

$$R - \varepsilon_n \leq \frac{1}{n} I\left(W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}, \ldots, U_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}\right)$$
$$+ I\left(X\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right); Y_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \Big| V_{N_\mathbf{d}-1}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)\right),$$
$$\quad (64c)$$

For the Gaussian channel (2), for $i = 1, \ldots, N_\mathbf{d}$, we have [15]

$$I\left(V_i\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right); Y_{\pi_{\mathcal{U}_\mathbf{d}}(i)}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \Big| V_{i-1}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)\right) \leq$$
$$\frac{1}{2} \log_2 \left(1 + \frac{\beta_i h_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^2 P_{\mathcal{U}_\mathbf{d}}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)}{h_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^2 \sum_{j=i+1}^{N_\mathbf{d}} \beta_j P_{\mathcal{U}_\mathbf{d}}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) + 1}\right), \quad (65)$$

for some $\beta_i \geq 0$, for $i = 1, \ldots, N_\mathbf{d}$, such that $\sum_{i=1}^{N_\mathbf{d}} \beta_i = 1$, where we set $V_0\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \triangleq 0$, and $V_{N_\mathbf{d}}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \triangleq X\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)$. From (64) and (65), for n sufficiently large, the average power $P_{\mathcal{U}_\mathbf{d}}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right)$ to satisfy any demand vector $\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l \in \mathcal{D}_{\mathcal{U}_\mathbf{d}}^l$, for $l \in [D]$, is lower bounded by

$$P_{\mathcal{U}_\mathbf{d}}\left(\mathbf{d}_{\mathcal{U}_\mathbf{d}}^l\right) \geq f\left(c_{\pi_{\mathcal{U}_\mathbf{d}}(1)}^l, \ldots, c_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}^l\right) \triangleq$$
$$\sum_{i=1}^{N_\mathbf{d}} \left(\frac{2^{2c_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^l} - 1}{h_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^2}\right) \prod_{j=1}^{i-1} 2^{2c_{\pi_{\mathcal{U}_\mathbf{d}}(j)}^l}, \quad (66a)$$

where, for $i = 1, \ldots, N_\mathbf{d}$ and $l = 1, \ldots, D$,

$$c_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^l \triangleq R - \frac{1}{n} I\left(W_{d_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}, \ldots, U_{\pi_{\mathcal{U}_\mathbf{d}}(i)}\right). \quad (66b)$$

Note that the lower bound in (66a) does not depend on any particular demand in $\mathcal{D}_{\mathcal{U}_\mathbf{d}}^l$, $l \in [D]$. Thus, from (60) and (66), we have

$$\bar{P} \geq \mathrm{E}_{\mathcal{U}_\mathbf{d}}\left[\frac{1}{D} \sum_{l=1}^{D} f\left(c_{\pi_{\mathcal{U}_\mathbf{d}}(1)}^l, \ldots, c_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}^l\right)\right]. \quad (67)$$

Lemma 1. *Given a set of users $\mathcal{U}_\mathbf{d}$ of size $N_\mathbf{d}$ with distinct demands, we have*

$$\frac{1}{D} \sum_{l=1}^{D} f\left(c_{\pi_{\mathcal{U}_d}(1)}^l, \ldots, c_{\pi_{\mathcal{U}_d}(N_d)}^l\right) \geq$$
$$\sum_{i=1}^{N_\mathbf{d}} \left(\frac{2^{2R(1-\min\{iM/N,1\})} - 1}{h_{\pi_{\mathcal{U}_\mathbf{d}}(i)}^2}\right) \prod_{j=1}^{i-1} 2^{2R(1-\min\{jM/N,1\})}. \quad (68)$$

Proof. It is proved in Appendix A that $f(\cdot)$ is a convex function of $\left(c_{\pi_{\mathcal{U}_\mathbf{d}}(1)}^l, \ldots, c_{\pi_{\mathcal{U}_\mathbf{d}}(N_\mathbf{d})}^l\right)$. Thus,

$$\frac{1}{D} \sum_{l=1}^{D} f\left(c_{\pi_{\mathcal{U}_d}(1)}^l, \ldots, c_{\pi_{\mathcal{U}_d}(N_d)}^l\right) \geq$$
$$f\left(\frac{1}{D} \sum_{l=1}^{D} c_{\pi_{\mathcal{U}_d}(1)}^l, \ldots, \frac{1}{D} \sum_{l=1}^{D} c_{\pi_{\mathcal{U}_d}(N_d)}^l\right). \quad (69)$$

From the definition, we have, for $i = 1, \ldots, N_\mathbf{d}$,

$$\frac{1}{D} \sum_{l=1}^{D} c_{\pi_{\mathcal{U}_d}(i)}^l =$$
$$R - \frac{1}{nD} \sum_{l=1}^{D} I\left(W_{d_{\pi_{\mathcal{U}_d}(i)}^l}; U_{\pi_{\mathcal{U}_\mathbf{d}}(1)}, \ldots, U_{\pi_{\mathcal{U}_\mathbf{d}}(i)}\right). \quad (70)$$

where, due to the symmetry, each file W_k, for $k = 1, \ldots, N$, appears $\binom{N-1}{N_\mathbf{d}-1}(N_\mathbf{d}-1)!$ times in the sum on the RHS of (70)

for each i value. Thus, for $i = 1, ..., N_{\mathbf{d}}$, we have

$$
\frac{1}{D} \sum_{l=1}^{D} c_{\pi_{\mathcal{U}_{\mathbf{d}}}(i)}^{l} = R - \frac{\binom{N-1}{N_{\mathbf{d}}-1}(N_{\mathbf{d}}-1)!}{nD}
$$

$$
\sum_{k=1}^{N} I\left(W_k; U_{\pi_{\mathcal{U}_{\mathbf{d}}}(1)}, \ldots, U_{\pi_{\mathcal{U}_{\mathbf{d}}}(i)}\right)
$$

$$
= R - \frac{1}{nN} \sum_{k=1}^{N} I\left(W_k; U_{\pi_{\mathcal{U}_{\mathbf{d}}}(1)}, \ldots, U_{\pi_{\mathcal{U}_{\mathbf{d}}}(i)}\right)
$$

$$
\overset{(a)}{\geq} R - \frac{1}{nN} I\left(W_1, ..., W_N; U_{\pi_{\mathcal{U}_{\mathbf{d}}}(1)}, \ldots, U_{\pi_{\mathcal{U}_{\mathbf{d}}}(i)}\right)
$$

$$
= R - \frac{R}{N} \min\{iM, N\}, \tag{71}
$$

where (a) follows from the independence of the files. From (69) and (71), and the definition of function f, we have

$$
\frac{1}{D} \sum_{l=1}^{D} f\left(c_{\pi_{\mathcal{U}_{\mathbf{d}}}(1)}^{l}, \ldots, c_{\pi_{\mathcal{U}_{\mathbf{d}}}(N_{\mathbf{d}})}^{l}\right) \geq
$$
$$
f\left(R\left(1 - \min\left\{\frac{iM}{N}, 1\right\}\right), \ldots, R\left(1 - \min\left\{\frac{iM}{N}, 1\right\}\right)\right), \tag{72}
$$

which concludes the proof of Lemma 1. $\square$

According to (67) and Lemma 1, $\bar{P}$ is lower bounded by $\bar{P}_{\mathrm{LB}}(R, M)$ defined in (53). Thus, $\bar{P}_{\mathrm{LB}}(R, M)$ is a lower bound on $\bar{P}_{\mathcal{U}_{\mathbf{d}}}^*(R, M)$ as well as $\bar{P}^*(R, M)$.

Next, we prove the lower bound on $\hat{P}^*(R, M)$ stated in Theorem 3. For any $\mathcal{U}_{\mathbf{d}}$ set, let $\left(R, M, \bar{P}, \hat{P}\right)$ tuple be $\mathcal{U}_{\mathbf{d}}$-achievable. We have $\hat{P} \geq P_{\mathcal{U}_{\mathbf{d}}}(\mathbf{d})$, $\forall \mathbf{d} \in \mathcal{D}_{\mathcal{U}_{\mathbf{d}}}$, which is equivalent to

$$
\hat{P} \geq P_{\mathcal{U}_{\mathbf{d}}}\left(\mathbf{d}_{\mathcal{U}_{\mathbf{d}}}^{l}\right), \quad \forall \mathbf{d}_{\mathcal{U}_{\mathbf{d}}}^{l} \in \mathcal{D}_{\mathcal{U}_{\mathbf{d}}}^{l}, \text{for } l = 1, ..., D. \tag{73}
$$

Averaging over all $N_{\mathcal{U}_{\mathbf{d}}} = DN_{\mathcal{U}_{\mathbf{d}}}'$ possible demands with the same $\mathcal{U}_{\mathbf{d}}$ set, we have

$$
\hat{P} \geq \frac{1}{D} \sum_{l=1}^{D} \left(\frac{1}{N_{\mathcal{U}_{\mathbf{d}}}'} \sum_{\mathbf{d}_{\mathcal{U}_{\mathbf{d}}}^{l} \in \mathcal{D}_{\mathcal{U}_{\mathbf{d}}}^{l}} P_{\mathcal{U}_{\mathbf{d}}}\left(\mathbf{d}_{\mathcal{U}_{\mathbf{d}}}^{l}\right)\right). \tag{74}
$$

According to (66) and Lemma 1, $\hat{P}$ is lower bounded as follows:

$$
\hat{P} \geq \sum_{i=1}^{N_{\mathbf{d}}} \left(\frac{2^{2R(1-\min\{iM/N, 1\})} - 1}{h_{\pi_{\mathcal{U}_{\mathbf{d}}}(i)}^2}\right) \prod_{j=1}^{i-1} 2^{2R(1-\min\{jM/N, 1\})},
$$
$$
\forall \mathcal{U}_{\mathbf{d}}. \tag{75}
$$

Thus, we have

$$
\hat{P} \geq \max_{\mathcal{U}_{\mathbf{d}}} \left\{ \sum_{i=1}^{N_{\mathbf{d}}} \left(\frac{2^{2R(1-\min\{iM/N, 1\})} - 1}{h_{\pi_{\mathcal{U}_{\mathbf{d}}}(i)}^2}\right) \right.
$$
$$
\left. \prod_{j=1}^{i-1} 2^{2R(1-\min\{jM/N, 1\})} \right\}. \tag{76}
$$

We note that the term on the RHS of the inequality in (76) is equivalent to $\hat{P}_{\mathrm{LB}}(R, M)$ defined in (54). Thus, $\hat{P}_{\mathrm{LB}}(R, M)$ is a lower bound on $\hat{P}_{\mathcal{U}_{\mathbf{d}}}^*(R, M)$ as well as $\hat{P}^*(R, M)$. $\square$

VI. NUMERICAL RESULTS

For the numerical results, we assume that the rate of the files in the library is fixed to $R = 1$, and the channel gains

are $1/h_k^2 = 2 - 0.2(k-1)$, $k \in [K]$.

The average power-memory trade-off $\bar{P}^*(R, M)$ and peak power-memory trade-off $\hat{P}^*(R, M)$ are shown in Fig. 1a and Fig. 1b, respectively, for $K = 5$ users, and $N = 8$ files in the library. The gap between centralized and decentralized caching, which measures the power required to compensate for the decentralization of the cache placement phase, is relatively small, particularly for small and large values of the cache capacity. We observe that the minimum average and peak powers drop very quickly even with a small cache capacity available at the users. The lower bound is generally tight for both average and peak power values with respect to the upper bound for the centralized scenario for the whole range of cache capacities[6]. The peak power values exhibit similar behaviour to the average power, with significantly higher values. This shows that adapting the transmit power to the demand combination can reduce the average energy consumption significantly, while the transmitter spends much higher energy for some demand combinations.

We define the following multiplicative gaps between the two bounds on the average and peak power values for the centralized and decentralized caching scenarios:

$$
\bar{G}_{\mathrm{C}} \triangleq \frac{\bar{P}_{\mathrm{UB}}^{\mathrm{C}}(R, M)}{\bar{P}_{\mathrm{LB}}(R, M)}, \tag{77a}
$$

$$
\bar{G}_{\mathrm{D}} \triangleq \frac{\bar{P}_{\mathrm{UB}}^{\mathrm{D}}(R, M)}{\bar{P}_{\mathrm{LB}}(R, M)}, \tag{77b}
$$

$$
\hat{G}_{\mathrm{C}} \triangleq \frac{\hat{P}_{\mathrm{UB}}^{\mathrm{C}}(R, M)}{\hat{P}_{\mathrm{LB}}(R, M)}, \tag{77c}
$$

$$
\hat{G}_{\mathrm{D}} \triangleq \frac{\hat{P}_{\mathrm{UB}}^{\mathrm{D}}(R, M)}{\hat{P}_{\mathrm{LB}}(R, M)}. \tag{77d}
$$

Fig. 2 illustrates the multiplicative gaps for $K = 5$ and $N = 8$. The multiplicative gap is significantly smaller for centralized caching for both the average and peak power values. Observe that the multiplicative gap is relatively small for the entire range of cache capacities, and it is higher for the intermediate values of the cache capacity.

Fig. 3 illustrates the effect of the number of files N on the upper bound. In Fig. 3a and Fig. 3b the average power values and the peak power values are considered, respectively, for $K = 5$ users, and different number of files $N \in \{10, 20, 40, 100\}$. The gap between the centralized and decentralized caching scenarios increases N. Another observation to be noted is the increase in the average power with N, even though the number of files requested by the users remains the same. This stems from two reasons: first, the effect of the local caches diminishes as the library size increases; and second, the users are less likely to request the same files from a larger library (and hence the increase in the average power values for low M values). We also observe that the peak power increases with N, but only for non-zero M values. This is because the increase in the peak power is only due to the diminished utility of the cache capacity, whereas the peak power depends only on the worst-case demand combination;
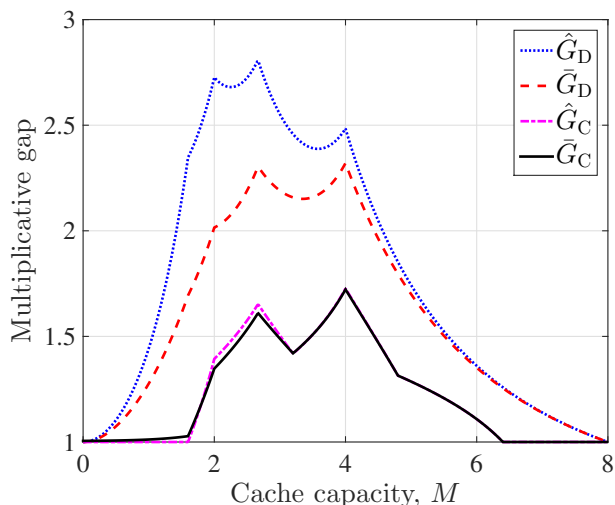
[6]Note that the figure does not include the cache capacity range $4 \leq M \leq 8$, in which case the three curves coincide in both figures.
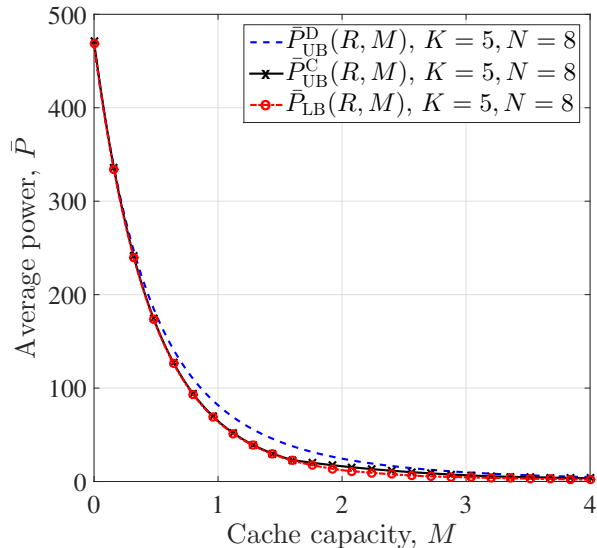
(a) Average power-memory trade-off
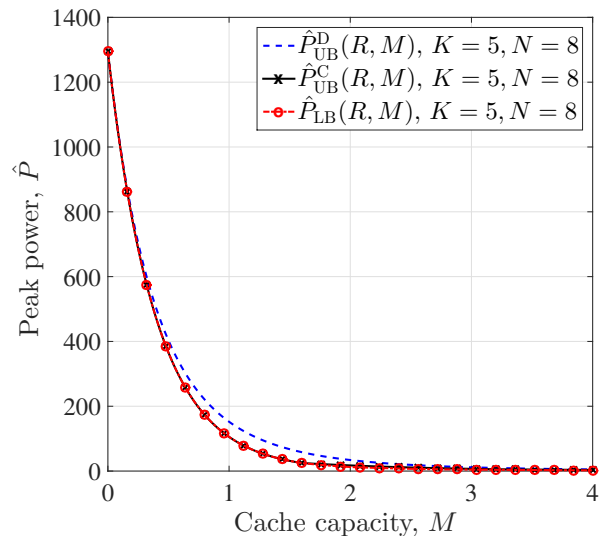
(b) Peak power-memory trade-off

Fig. 1: Power-memory trade-off for a Gaussian BC with $K = 5$ users, and $N = 8$ files.

Fig. 2: Multiplicative gap between the upper and lower bounds for $K = 5$ users and $N = 8$ files.

and thus, does not depend on the likelihood of common requests. We note that, for a fixed number of users K, the gap between the average power and the peak power reduces by increasing N, and for $N \gg K$, the gap diminishes, since, with high probability, the users demand distinct files.

Fig. 4 illustrates the multiplicative gap $\bar{G}_C$, i.e., for the upper and lower bounds on the average transmit power, with respect to M, for $K = 5$ users and $N \in \{10, 20, 40, 100\}$. Observe that the bounds are relatively close, and the multiplicative gap follows a similar pattern for different N values, expanded horizontally with increasing N.

The effect of the number of users K on the average and peak power values for both the centralized and decentralized scenarios is shown in Fig. 5. Fig. 5a and Fig. 5b demonstrate

the average and peak power values as a function of M, respectively, for $N = 10$ and $K \in \{3, 4, 5\}$ users. The gap between centralized and decentralized caching increases with K in both figures. The average and peak power values also increase with K, as expected. For a fixed N value, the increase in the peak power is higher than the one in the average power. This is due to the fact that the likelihood of common demands increases with K, and so does the gap between the average and peak power values.

Finally, we consider the multiplicative gap $\bar{G}_C$ for $K \in \{3, 4, 5\}$ and $N = 10$ in Fig. 6. It can be seen that the gap increases with K.

VII. CONCLUSIONS

We have considered cache-aided content delivery over a Gaussian BC. Considering same rate contents in the library, we have studied both the *minimum peak transmission power*, which is the minimum transmit power that can satisfy all user demand combinations, and the *minimum average transmit power*, averaged across all demand combinations, assuming uniform demand distributions. We have proposed a centralized caching and coded delivery scheme assuming that the channel conditions in the delivery phase are not known beforehand. Coded contents are transmitted in the delivery phase to their intended receivers using superposition coding and power allocation. We have then extended the achievable scheme to the decentralized caching scenario. We have also provided a lower bound on the required peak and average transmission power values assuming uncoded cache placement. Our results indicate that even a small cache capacity at the receivers can provide a significant reduction in the required transmission power level highlighting the benefits of caching in improving the energy efficiency of wireless networks.

(a) Average power-memory trade-off
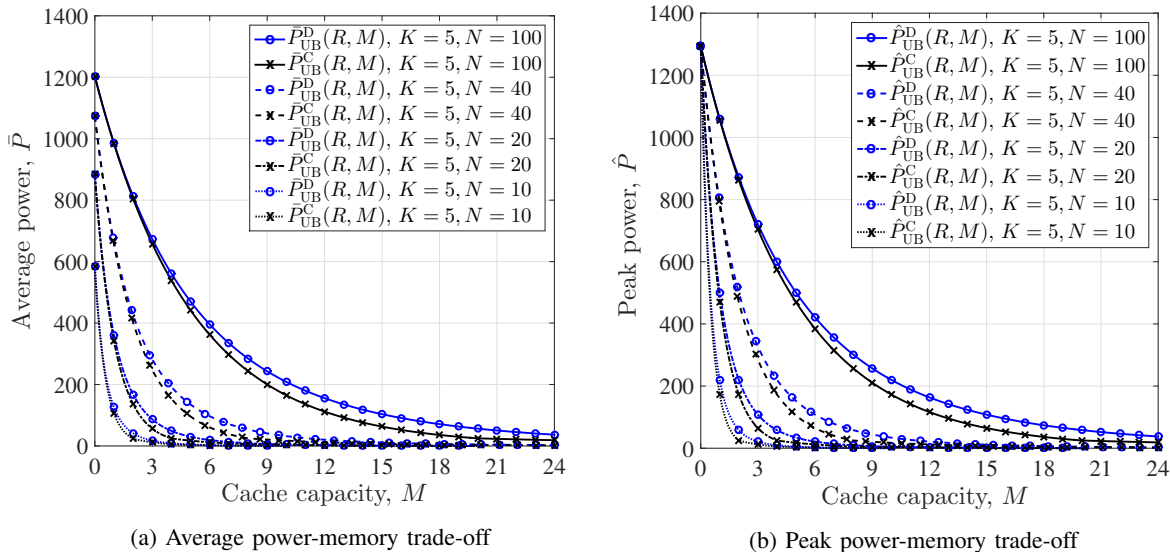

(b) Peak power-memory trade-off

Fig. 3: Power-memory trade-off for a Gaussian BC with $K = 5$ users, and various number of files $N \in \{10, 20, 40, 100\}$ in the library.
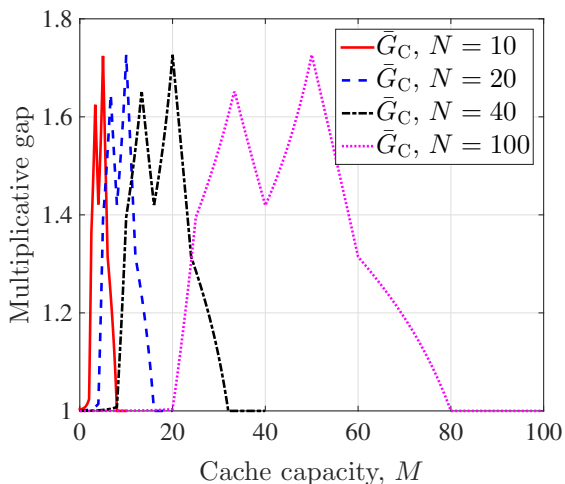

Fig. 4: The multiplicative gap between the upper and lower bounds on the average power for the centralized caching scenario for $K = 5$ users and $N \in \{10, 20, 40, 100\}$ files.

APPENDIX A
PROOF OF CONVEXITY OF FUNCTION $f(\cdot)$

We show that, for $1 \leq N_{\mathbf{d}} \leq K$, function $f : \mathbb{R}^{N_{\mathbf{d}}} \to \mathbb{R}$ is a convex function of $(s_1, ..., s_{N_{\mathbf{d}}})$:

$$f(s_1, \ldots, s_{N_{\mathbf{d}}}) = \sum_{i=1}^{N_{\mathbf{d}}} \left(\frac{2^{2s_k} - 1}{h^2_{\pi_{\mathcal{U}_{\mathbf{d}}}(i)}} \right) \prod_{j=1}^{i-1} 2^{2s_j}. \quad (78)$$

After mathematical manipulation, one can obtain

$$f(s_1, \ldots, s_{N_{\mathbf{d}}}) =$$
$$\sum_{i=1}^{N_{\mathbf{d}}} \left(\left(\frac{1}{h^2_{\pi_{\mathcal{U}_{\mathbf{d}}}(i)}} - \frac{1}{h^2_{\pi_{\mathcal{U}_{\mathbf{d}}}(i+1)}} \right) 2^{2\mathbf{a}_i^T \mathbf{s}} \right) - \frac{1}{h^2_{\pi_{\mathcal{U}_{\mathbf{d}}}(1)}}, \quad (79)$$

where $h^2_{\pi_{\mathcal{U}_{\mathbf{d}}}(N_{\mathbf{d}}+1)} \triangleq \infty$, $\mathbf{s} \triangleq [s_1 \ s_2 \ \cdots \ s_{N_{\mathbf{d}}}]^T$, and

$$\mathbf{a}_i \triangleq \left[\underbrace{1\,1\cdots1}_{1 \times i} \underbrace{0\,0\cdots0}_{1 \times (N_{\mathbf{d}}-i)} \right]^T, \quad \text{for } i = 1, ..., N_{\mathbf{d}}. \quad (80)$$

We note that $h^2_{\pi_{\mathcal{U}_{\mathbf{d}}}(i)} \leq h^2_{\pi_{\mathcal{U}_{\mathbf{d}}}(i+1)}$, $\forall i \in [N_{\mathbf{d}}]$, and function 2^{2s}, $s \in \mathbb{R}$, is a convex function of s. Thus, all the functions $2^{2\mathbf{a}_i^T \mathbf{s}}$, $\forall i \in [N_{\mathbf{d}}]$, are convex since the affine substitution of the arguments preserves convexity. Hence, function f is convex with respect to $(s_1, ..., s_{N_{\mathbf{d}}})$ since any linear combination of convex functions with non-negative coefficients is convex.

APPENDIX B
PROOF OF WORST-CASE DEMAND COMBINATION FOR CENTRALIZED CACHING

We first illustrate that, among all the demand combinations with a fixed $N_{\mathbf{d}}$, the worst-case is the one when $\mathcal{U}_{\mathbf{d}} = [N_{\mathbf{d}}]$. Let, for a fixed $N_{\mathbf{d}}$, $\mathcal{D}^*_{N_{\mathbf{d}}}$ denotes the set of demand vectors with $\mathcal{U}_{\mathbf{d}} = [N_{\mathbf{d}}]$. Note that, the required power, given in (39), is the same for any demand vector $\mathbf{d} \in \mathcal{D}^*_{N_{\mathbf{d}}}$. Assume that there is a demand combination $\tilde{\mathbf{d}} \notin \mathcal{D}^*_{N_{\mathbf{d}}}$, with required transmitted power $\tilde{P}^C \triangleq P^C_{\text{UB}}\left(R, M, \tilde{\mathbf{d}}\right)$ determined through (39), that has the highest transmitted power among all the demand combinations with the same $N_{\mathbf{d}}$. For such demand vector $\tilde{\mathbf{d}} \notin \mathcal{D}^*_{N_{\mathbf{d}}}$, there is at least one user $k \in \mathcal{U}_{\tilde{\mathbf{d}}}$ such that $k - 1 \notin \mathcal{U}_{\tilde{\mathbf{d}}}$, where we have $N_{\tilde{\mathbf{d}},k-1} = N_{\tilde{\mathbf{d}},k} + 1$. Now, consider another demand vector $\hat{\mathbf{d}}$ with the same entries as demand vector $\tilde{\mathbf{d}}$ except that $k \notin \mathcal{U}_{\hat{\mathbf{d}}}$ and $k - 1 \in \mathcal{U}_{\hat{\mathbf{d}}}$. An example of such demand vector $\hat{\mathbf{d}}$ is as follows:
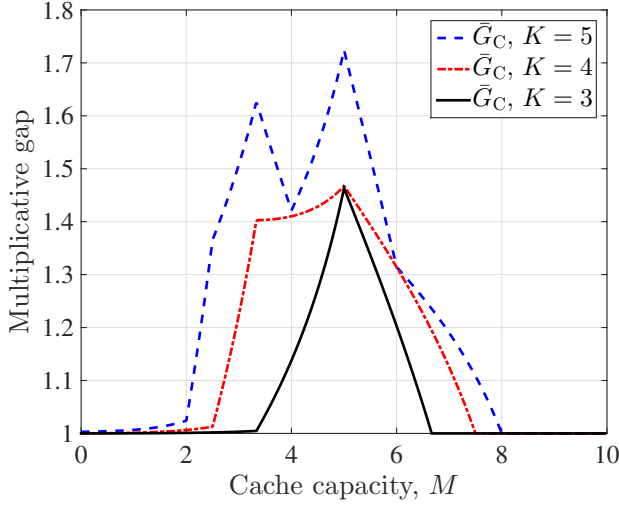
$$\hat{\mathbf{d}}_i = \begin{cases} \tilde{\mathbf{d}}_{i+1}, & \text{if } i = k - 1, \\ \tilde{\mathbf{d}}_{i-1}, & \text{if } i = k, \\ \tilde{\mathbf{d}}_i, & \text{otherwise.} \end{cases} \quad (81)$$
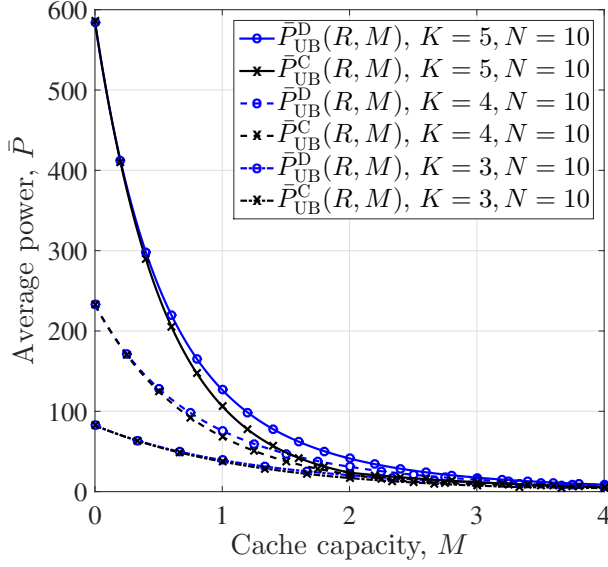
(a) Average power-memory trade-off
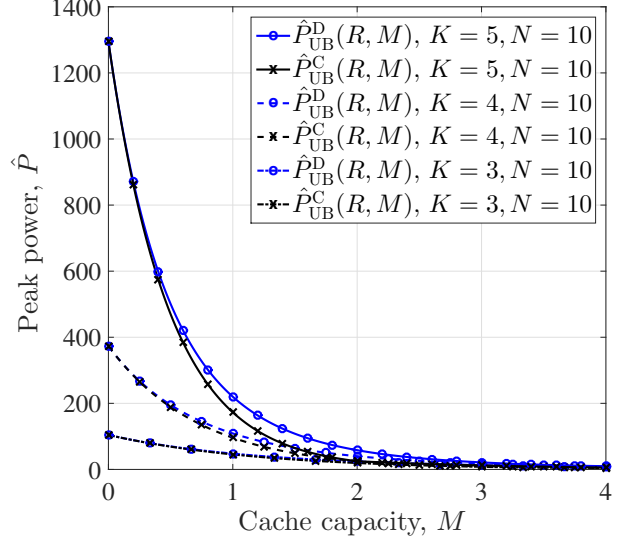
(b) Peak power-memory trade-off

Fig. 5: Power-memory trade-off for a Gaussian BC with various number of users $K \in \{3, 4, 5\}$, and $N = 10$ in the library.

Fig. 6: The multiplicative gap between the upper and lower bounds on the average power for the centralized caching scenario for $K \in \{3, 4, 5\}$ users and $N = 10$ files.

We denote the required transmitted power to satisfy demand vector $\hat{\mathbf{d}}$ by $\hat{P}^{\mathrm{C}}$, where $\hat{P}^{\mathrm{C}} \triangleq P_{\mathrm{UB}}^{\mathrm{C}}\left(R, M, \hat{\mathbf{d}}\right)$. Note that, for demand vector $\hat{\mathbf{d}}$, we have $N_{\hat{\mathbf{d}},k-1} = N_{\hat{\mathbf{d}},k} = N_{\tilde{\mathbf{d}},k}$, and from (37),

$$R_{\hat{\mathbf{d}},i}^{\mathrm{C}} = R_{\tilde{\mathbf{d}},i}^{\mathrm{C}}, \quad \forall i \in [K] \setminus \{k-1, k\}. \tag{82}$$

Having defined

$$R_{\mathrm{L},k} \triangleq \frac{\binom{K-k}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}}\left(\lfloor t \rfloor + 1 - t\right) R + \frac{\binom{K-k}{\lfloor t \rfloor+1}}{\binom{K}{\lfloor t \rfloor+1}}\left(t - \lfloor t \rfloor\right) R, \tag{83a}$$

$$R_{\mathrm{NL},k}\left(N_{\mathbf{d},k}\right) \triangleq \frac{\binom{K-k}{\lfloor t \rfloor} - \binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}}\left(\lfloor t \rfloor + 1 - t\right) R + \frac{\binom{K-k}{\lfloor t \rfloor+1} - \binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor+1}}{\binom{K}{\lfloor t \rfloor+1}}\left(t - \lfloor t \rfloor\right) R, \tag{83b}$$

according to (39), we have

$$\hat{P}^{\mathrm{C}} - \tilde{P}^{\mathrm{C}} = \frac{1}{h_{k-1}^2}\left(2^{2R_{\mathrm{L},k-1}} - 1\right) \prod_{j=1}^{k-2} 2^{2R_{\tilde{\mathbf{d}},j}^{\mathrm{C}}}$$
$$+ \frac{1}{h_k^2}\left(2^{2R_{\mathrm{NL},k}\left(N_{\tilde{\mathbf{d}},k}\right)} - 1\right)\left(2^{2R_{\mathrm{L},k-1}}\right) \prod_{j=1}^{k-2} 2^{2R_{\tilde{\mathbf{d}},j}^{\mathrm{C}}}$$
$$+ \sum_{i=k+1}^{K} \frac{1}{h_i^2}\left(2^{2R_{\tilde{\mathbf{d}},i}^{\mathrm{C}}} - 1\right)$$
$$\prod_{\substack{j=1 \\ j \neq k-1,k}}^{i-1}\left(2^{2R_{\tilde{\mathbf{d}},j}^{\mathrm{C}}}\right)\left(2^{2R_{\mathrm{L},k-1}}\right)\left(2^{2R_{\mathrm{NL},k}\left(N_{\tilde{\mathbf{d}},k}\right)}\right)$$
$$- \frac{1}{h_{k-1}^2}\left(2^{2R_{\mathrm{NL},k-1}\left(N_{\tilde{\mathbf{d}},k}+1\right)} - 1\right) \prod_{j=1}^{k-2} 2^{2R_{\tilde{\mathbf{d}},j}^{\mathrm{C}}}$$
$$- \frac{1}{h_k^2}\left(2^{2R_{\mathrm{L},k}} - 1\right)\left(2^{2R_{\mathrm{NL},k-1}\left(N_{\tilde{\mathbf{d}},k}+1\right)}\right) \prod_{j=1}^{k-2} 2^{2R_{\tilde{\mathbf{d}},j}^{\mathrm{C}}}$$
$$- \sum_{i=k+1}^{K} \frac{1}{h_i^2}\left(2^{2R_{\tilde{\mathbf{d}},i}^{\mathrm{C}}} - 1\right)$$
$$\prod_{\substack{j=1 \\ j \neq k-1,k}}^{i-1}\left(2^{2R_{\tilde{\mathbf{d}},j}^{\mathrm{C}}}\right)\left(2^{2R_{\mathrm{NL},k-1}\left(N_{\tilde{\mathbf{d}},k}+1\right)}\right)\left(2^{2R_{\mathrm{L},k}}\right), \tag{84}$$

where we have used $N_{\tilde{\mathbf{d}},k-1} = N_{\tilde{\mathbf{d}},k}+1$, and $N_{\hat{\mathbf{d}},k-1} = N_{\hat{\mathbf{d}},k} = N_{\tilde{\mathbf{d}},k}$. From (83), we have

$$R_{\mathrm{NL},k-1}\left(N_{\tilde{\mathbf{d}},k}+1\right) = R_{\mathrm{NL},k}\left(N_{\tilde{\mathbf{d}},k}\right) + R_{\mathrm{L},k-1} - R_{\mathrm{L},k}. \tag{85}$$

By substituting (85) into (84), we obtain

$$\hat{P}^{\mathrm{C}} - \tilde{P}^{\mathrm{C}} = \left(\prod_{j=1}^{k-2} 2^{2R_{\hat{\mathbf{d}},j}^{\mathrm{C}}} \right) \left(2^{2R_{\mathrm{L},k-1}} \right)$$
$$\left(1 - 2^{2R_{\mathrm{NL},k}\left(N_{\hat{\mathbf{d}},k}\right)-2R_{\mathrm{L},k}} \right) \left(\frac{1}{h_{k-1}^2} - \frac{1}{h_k^2} \right). \quad (86)$$

Note that,

$$R_{\mathrm{NL},k}\left(N_{\mathbf{d},k}\right) - R_{\mathrm{L},k} = -\frac{\binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}} \left(\lfloor t \rfloor + 1 - t \right) R$$
$$- \frac{\binom{K-k-N_{\mathbf{d},k}}{\lfloor t \rfloor + 1}}{\binom{K}{\lfloor t \rfloor + 1}} \left(t - \lfloor t \rfloor \right) R \le 0, \quad (87)$$

and since $h_{k-1}^2 \le h_k^2$, from (86), we have $\hat{P}^{\mathrm{C}} \ge \tilde{P}^{\mathrm{C}}$, which contradicts the assumption that demand vector $\tilde{\mathbf{d}} \notin \mathcal{D}_{N_{\mathbf{d}}}^*$ has the highest required transmit power among all the demand combinations with the same $N_{\mathbf{d}}$. Therefore, for a fixed $N_{\mathbf{d}}$, the demand vectors in $\mathcal{D}_{N_{\mathbf{d}}}^*$, i.e., the demand vectors with $\mathcal{U}_{\mathbf{d}} = [N_{\mathbf{d}}]$, require the maximum transmit power with the proposed scheme. From (37), for all the demand vectors in $\mathcal{D}_{N_{\mathbf{d}}}^*$, user $k \in [K]$ is delivered a message of rate

$$R_{\mathbf{d},k}^{*\mathrm{C}}\left(N_{\mathbf{d}}\right) \triangleq \begin{cases} \frac{\binom{K-k}{\lfloor t \rfloor}}{\binom{K}{\lfloor t \rfloor}} \left(\lfloor t \rfloor + 1 - t \right) R \\ \quad + \frac{\binom{K-k}{\lfloor t \rfloor + 1}}{\binom{K}{\lfloor t \rfloor + 1}} \left(t - \lfloor t \rfloor \right) R, & \text{if } k \in [N_{\mathbf{d}}], \quad (88) \\ 0, & \text{otherwise.} \end{cases}$$

Thus, according to (39), for a fixed $N_{\mathbf{d}}$, the required power to satisfy any demand combination $\mathbf{d} \in \mathcal{D}_{N_{\mathbf{d}}}^*$ is given by

$$P_{\mathrm{UB}}^{\mathrm{C}}\left(R, M, \mathbf{d}\right) = \sum_{i=1}^{N_{\mathbf{d}}} \left(\frac{2^{2R_{\mathbf{d},i}^{*\mathrm{C}}(N_{\mathbf{d}})} - 1}{h_i^2} \right) \prod_{j=1}^{i-1} 2^{2R_{\mathbf{d},j}^{*\mathrm{C}}(N_{\mathbf{d}})}. \quad (89)$$

Observe that, (89) is an increasing function of $N_{\mathbf{d}}$, and the power is maximized for $N_{\mathbf{d}} = \min\{N, K\}$. Thus, the worst-case demand combination for the proposed centralized caching and coded delivery scheme happens when $\mathcal{U}_{\mathbf{d}} = [\min\{N, K\}]$.

APPENDIX C
PROOF OF WORST-CASE DEMAND COMBINATION FOR DECENTRALIZED CACHING

Similarly to the centralized caching scenario, assume that, for a fixed $N_{\mathbf{d}}$, there is a demand combination $\tilde{\mathbf{d}} \notin \mathcal{D}_{N_{\mathbf{d}}}^*$, with required transmitted power $\tilde{P}^{\mathrm{D}} \triangleq P_{\mathrm{UB}}^{\mathrm{D}}\left(R, M, \tilde{\mathbf{d}}\right)$ determined through (51), that has the highest transmitted power among all the demand combinations with the same $N_{\mathbf{d}}$. For such demand vector $\tilde{\mathbf{d}} \notin \mathcal{D}_{N_{\mathbf{d}}}^*$, there is at least one user $k \in \mathcal{U}_{\tilde{\mathbf{d}}}$ such that $k - 1 \notin \mathcal{U}_{\tilde{\mathbf{d}}}$, where we have $N_{\tilde{\mathbf{d}},k-1} = N_{\tilde{\mathbf{d}},k} + 1$. Similarly to (81), consider another demand vector $\hat{\mathbf{d}}$ with the same entries as demand vector $\tilde{\mathbf{d}}$ except that $k \notin \mathcal{U}_{\hat{\mathbf{d}}}$ and $k - 1 \in \mathcal{U}_{\hat{\mathbf{d}}}$. We denote the required transmitted power for demand vector $\hat{\mathbf{d}}$ by $\hat{P}^{\mathrm{D}}$, defined as $\hat{P}^{\mathrm{D}} \triangleq P_{\mathrm{UB}}^{\mathrm{D}}\left(R, M, \hat{\mathbf{d}}\right)$. Note that, for demand vector $\hat{\mathbf{d}}$, we have $N_{\hat{\mathbf{d}},k-1} = N_{\hat{\mathbf{d}},k} = N_{\tilde{\mathbf{d}},k}$, and from (48),

$$R_{\hat{\mathbf{d}},i}^{\mathrm{D}} = R_{\tilde{\mathbf{d}},i}^{\mathrm{D}}, \quad \forall i \in [K] \backslash \{k-1, k\}. \quad (90)$$

According to (51), we have

$$\hat{P}^{\mathrm{D}} - \tilde{P}^{\mathrm{D}} = \frac{1}{h_{k-1}^2} \left(2^{2(1-M/N)^{k-1}R} - 1 \right) \prod_{j=1}^{k-2} 2^{2R_{\hat{\mathbf{d}},j}^{\mathrm{D}}}$$
$$+ \frac{1}{h_k^2} \left(2^{2(1-M/N)^k\left(1-(1-M/N)^{N_{\hat{\mathbf{d}},k}}\right)R} - 1 \right)$$
$$\left(2^{2(1-M/N)^{k-1}R} \right) \prod_{j=1}^{k-2} 2^{2R_{\hat{\mathbf{d}},j}^{\mathrm{D}}}$$
$$+ \sum_{i=k+1}^{K} \frac{1}{h_i^2} \left(2^{2R_{\hat{\mathbf{d}},i}^{\mathrm{D}}} - 1 \right) \prod_{\substack{j=1 \\ j \ne k-1,k}}^{i-1} \left(2^{2R_{\hat{\mathbf{d}},j}^{\mathrm{D}}} \right)$$
$$\left(2^{2(1-M/N)^{k-1}R} \right) \left(2^{2(1-M/N)^k\left(1-(1-M/N)^{N_{\hat{\mathbf{d}},k}}\right)R} \right)$$
$$- \frac{1}{h_{k-1}^2} \left(2^{2(1-M/N)^{k-1}\left(1-(1-M/N)^{N_{\tilde{\mathbf{d}},k}+1}\right)R} - 1 \right) \prod_{j=1}^{k-2} 2^{2R_{\tilde{\mathbf{d}},j}^{\mathrm{D}}}$$
$$- \frac{1}{h_k^2} \left(2^{2(1-M/N)^k R} - 1 \right)$$
$$\left(2^{2(1-M/N)^{k-1}\left(1-(1-M/N)^{N_{\tilde{\mathbf{d}},k}+1}\right)R} \right) \prod_{j=1}^{k-2} 2^{2R_{\tilde{\mathbf{d}},j}^{\mathrm{D}}}$$
$$- \sum_{i=k+1}^{K} \frac{1}{h_i^2} \left(2^{2R_{\tilde{\mathbf{d}},i}^{\mathrm{D}}} - 1 \right) \prod_{\substack{j=1 \\ j \ne k-1,k}}^{i-1} \left(2^{2R_{\tilde{\mathbf{d}},j}^{\mathrm{D}}} \right) \left(2^{2(1-M/N)^k R} \right)$$
$$\left(2^{2(1-M/N)^{k-1}\left(1-(1-M/N)^{N_{\tilde{\mathbf{d}},k}+1}\right)R} \right), \quad (91)$$

which is equal to

$$\hat{P}^{\mathrm{D}} - \tilde{P}^{\mathrm{D}} = \left(\prod_{j=1}^{k-2} 2^{2R_{\hat{\mathbf{d}},j}^{\mathrm{D}}} \right) \left(2^{2(1-M/N)^{k-1}R} \right.$$
$$\left. - 2^{2(1-M/N)^{k-1}\left(1-(1-M/N)^{N_{\tilde{\mathbf{d}},k}+1}\right)R} \right) \left(\frac{1}{h_{k-1}^2} - \frac{1}{h_k^2} \right). \quad (92)$$

Since $h_{k-1}^2 \le h_k^2$, and $0 \le (1 - M/N)^{N_{\tilde{\mathbf{d}},k}+1} \le 1$, from (92), we have $\hat{P}^{\mathrm{D}} \ge \tilde{P}^{\mathrm{D}}$, which contradicts the assumption that demand vector $\tilde{\mathbf{d}} \notin \mathcal{D}_{N_{\mathbf{d}}}^*$ has the highest required power among all the demand combinations with the same $N_{\mathbf{d}}$. Therefore, for a fixed $N_{\mathbf{d}}$, the demand vectors in $\mathcal{D}_{N_{\mathbf{d}}}^*$, i.e., the demand vectors with $\mathcal{U}_{\mathbf{d}} = [N_{\mathbf{d}}]$, require the maximum transmitted power with the proposed scheme. From (48), for all the demand vectors in $\mathcal{D}_{N_{\mathbf{d}}}^*$, we have

$$R_{\mathbf{d},k}^{*\mathrm{D}}\left(N_{\mathbf{d}}\right) \triangleq \begin{cases} \left(1 - \frac{M}{N}\right)^k R, & \text{if } k \in [N_{\mathbf{d}}], \\ 0, & \text{otherwise.} \end{cases} \quad (93)$$

Thus, according to (51), the required power to satisfy any demand combination $\mathbf{d} \in \mathcal{D}_{N_{\mathbf{d}}}^*$ is

$$P_{\mathrm{UB}}^{\mathrm{D}}\left(R, M, \mathbf{d}\right) = \sum_{i=1}^{N_{\mathbf{d}}} \left(\frac{2^{2(1-M/N)^i R} - 1}{h_i^2} \right) \prod_{j=1}^{i-1} 2^{2(1-M/N)^j R}. \quad (94)$$

Observe that, (94) is an increasing function of $N_{\mathbf{d}}$, and the required power is maximized for the highest value of $N_{\mathbf{d}}$, which is $\min\{N, K\}$. Thus, the worst-case demand combination for the proposed scheme happens when $\mathcal{U}_{\mathbf{d}} = [\min\{N, K\}]$.

REFERENCES

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

[2] W. Huang, S. Wang, L. Ding, F. Yang, and W. Zhang, "The performance analysis of coded cache in wireless fading channel," *arXiv:1504.01452v1 [cs.IT]*, Apr. 2015.

[3] N. Naderializadeh, M. Maddah-Ali, and A. S. Avestimehr, "Fundamental limits of cache-aided interference management," *IEEE Trans. Inform. Theory*, vol. 63, no. 5, pp. 3092–3107, May 2017.

[4] S. Saeedi Bidokhti, R. Timo, and M. Wigger, "Noisy broadcast networks with receiver caching," *arXiv:1605.02317v1 [cs.IT]*, May 2016.

[5] M. Mohammadi Amiri and D. Gündüz, "Cache-aided content delivery over erasure broadcast channels," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 370–381, Jan. 2018.

[6] J. Zhang and P. Elia, "Wireless coded caching: A topological perspective," *arXiv:1606.08253v1 [cs.IT]*, Jun. 2016.

[7] S. Saeedi Bidokhti, M. Wigger, and A. Yener, "Benefits of cache assignment on degraded broadcast channels," *arXiv:1702.08044v1 [cs.IT]*, Feb. 2017.

[8] L. Zheng, Q. Yan, Q. Chen, and X. Tang, "On the coded caching delivery design over wireless networks," *arXiv:1611.04853 [cs.IT]*, Nov. 2016.

[9] E. Piovano, H. Joudeh, and B. Clerckx, "On coded caching in the overloaded miso broadcast channel," in *Proc. IEEE Int'l Symp. on Inform. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2795–2799.

[10] K.-H. Ngo, S. Yang, and M. Kobayashi, "Scalable content delivery with coded caching in multi-antenna fading channels," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 548–562, Jan. 2018.

[11] S. P. Shariatpanahi, G. Caire, and B. H. Khalaj, "Physical-layer schemes for wireless coded caching," *arXiv:1711.05969 [cs.IT]*, Nov. 2017.

[12] A. C. Güngör and D. Gündüz, "Proactive wireless caching at mobile user devices for energy efficiency," in *Proc. IEEE Int'l Symp. on Wireless Comm. Systems (ISWCS)*, Brussels, Belgium, Aug. 2015.

[13] M. Gregori, J. Gomez-Vilardebo, J. Matamoros, and D. Gündüz, "Wireless content caching for small cell and D2D networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1222–1234, May 2016.

[14] M. Mohammadi Amiri and D. Gündüz, "Decentralized caching and coded delivery over Gaussian broadcast channels," in *Proc. IEEE Int'l Symp. on Inform. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2785–2789.

[15] P. Bergmans, "Coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, no. 2, pp. 197–207, Mar. 1973.

[16] Q. Yu, M. A. Maddah-Ali, and A. S. Avestimehr, "The exact rate-memory tradeoff for caching with uncoded prefetching," *IEEE Trans. Inform. Theory*, vol. 64, no. 2, pp. 1281–1296, Feb. 2018.

[17] M. A. Maddah-Ali and U. Niesen, "Decentralized caching attains order optimal memory-rate tradeoff," *IEEE/ACM Trans. Netw.*, vol. 23, no. 4, pp. 1029–1040, Apr. 2014.

Deniz Gündüz (S'03-M'08-SM'13) received his M.S. and Ph.D. degrees in electrical engineering from NYU Polytechnic School of Engineering (formerly Polytechnic University) in 2004 and 2007, respectively. After his PhD, he served as a postdoctoral research associate at Princeton University, as a consulting assistant professor at Stanford University, and as a research associate at CTTC. In September 2012, he joined the Electrical and Electronic Engineering Department of Imperial College London, UK, where he is currently a Reader in information theory and communications, and leads the Information Processing and Communications Lab. His research interests lie in the areas of communications and information theory, machine learning, and security and privacy in cyber-physical systems. Dr. Gündüz is an Editor of the IEEE Transactions on Communications, and the IEEE Transactions on Green Communications and Networking. He is the recipient of the IEEE Communications Society - Communication Theory Technical Committee (CTTC) Early Achievement Award in 2017 and a Starting Grant of the European Research Council (ERC) in 2016. He coauthored papers that received a Best Paper Award at the 2016 IEEE WCNC, and Best Student Paper Awards at 2007 IEEE ISIT and 2018 IEEE WCNC.

Mohammad Mohammadi Amiri (S'16) received the B.Sc. degree (Hons.) in Electrical Engineering from the Iran University of Science and Technology in 2011, and the M.Sc. degree (Hons.) in Electrical Engineering from the University of Tehran in 2014. He is currently pursuing the Ph.D. degree with the Imperial College London. His research interests include information and coding theory, wireless communications, MIMO systems, cooperative networks, cognitive radio, and signal processing.