Fundamental Limits of Latency in a Cache-Aided 4×4 Interference Channel

Joan S. Pujol Roig
Imperial College London
jp5215@imperial.ac.uk

Seyed Abolfazl Motahari
Sharif University of Technology
motahari@sharif.ir

Filippo Tosato
Toshiba Research Europe
filippo.tosato@toshiba-trel.com

Deniz Gündüz
Imperial College London
d.gunduz@imperial.ac.uk

***Abstract*— Fundamental limits of communication is studied in a 4×4 interference network, in which the transmitters are equipped with cache memories. Each of the receivers requests one file from a library of N equal-size files. The caches at the transmitters are filled without the knowledge of the user demands, such that all possible demand combinations can be satisfied reliably over the interference channel. The achievable *normalized delivery time* (NDT) is studied under *centralized* cache placement. By combining the interference alignment (IA) and zero-forcing (ZF) techniques, a novel caching and transmission scheme is presented, and is shown to be optimal for all possible cache sizes; fully characterizing the NDT for the 4×4 interference network with caches at the transmitter side.**

***Index terms*—** cache-aided networks, centralized caching, interference alignment, zero forcing, normalized delivery time.

I. Introduction

Wireless network traffic is experiencing a substantial transformation as it is increasingly becoming dominated by video content. To exploit the features of the video traffic and to cope with the exponential growth of future network traffic, *cache-aided networks* have emerged as a promising approach. In their pioneering work [1], Maddah-Ali and Niesen proposed a novel centralized coded caching scheme, which creates and exploits multicasting opportunities across users, significantly reducing the required delivery rate compared to uncoded caching. While the model in [1] assumes the delivery of demands over a noiseless multicast link, further studies extended this model to noisy delivery over broadcast channels [2], [3], [4], broadcast channels with feedback [5], [6], and the Wyner network model in [7]. While these models are still limited to a single server delivering all the demands to users equipped with caches, a 3×3 interference channel is studied in [8] considering cache memories at the transmitter side. In this model, contents are cached in advance at the caches of the transmitters, and are delivered in a distributed manner over an interference channel. Since the contents are distributed across multiple transmitters, the delivery requires exploiting both the interference alignment (IA) and zero-forcing (ZF) techniques. In [9], authors address interference management in a cache-aided network model with an arbitrary number of cache-enabled transmitters and receivers, employing ZF as well as interference cancellation (IC) exploiting the side information at the receivers caches. A constant-factor approximation of the *sum degrees-of-freedom* (sDoF) is provided in [10] for a general $K_T \times K_R$ cache-aided interference network with storage capabilities at both ends, when leveraging IA and IC. In [11], a general interference management scheme for both centralized and decentralized placement schemes is presented, considering caches at both ends and exploiting IA, IC and ZF techniques. In [12], the normalized delivery time (NDT), introduced in [5] and [13], is studied for a centralized cache-aided Gaussian interference network with an arbitrary number of transmitters and receivers and caches at both ends, and exploiting a combination of IA, IC and ZF techniques.

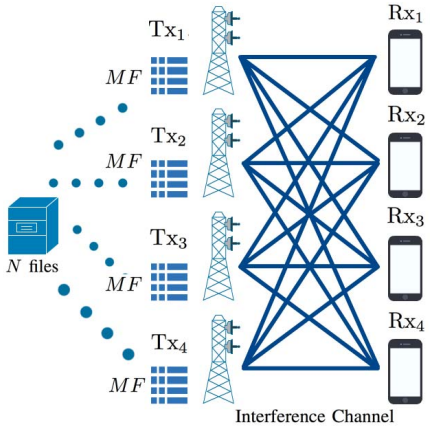

Fig. 1: The 4×4 cache-aided wireless interference network with transmitter caches.

The interference channel model is extended in [14] to a cloud-aided fog radio access network (F-RAN), in which the transmitters can fetch contents from the cloud through finite-capacity links. The authors study the achievable NDT for a centralized placement phase and propose a delivery phase that leverages transmitter caches as well as the cloud links. They also provide a lower-bound on the minimum achievable NDT for a general cloud and cache aided network. Decentralized caching for interference networks is studied in [11] and [15].

In this paper, we consider a wireless interference network consisting of single antenna terminals. We address the 4×4 network scenario with cache capabilities only at the trans-

This work was partially funded by the European Research Council through Starting Grant BEACON (agreement No. 677854). The first author acknowledges the generous support from Toshiba Research Europe for carrying out his PhD studies.

mitters. We propose an achievable scheme which aims to minimize the NDT, capturing the worst-case latency under centralized placement, that is, the cache contents can be centrally coordinated. The proposed delivery scheme jointly exploits IA and ZF techniques. Our results indicate a general improvement of the achieved NDT compared to the results of [9]–[12] for the 4×4 network with caches at the transmitter side. Moreover, the scheme is proved to be optimal in terms of the NDT performance as it meets the lower bound presented in [14] for any cache capacity. The gain emerges from the combination of ZF and IA compared to [9], [10], [14], and from the increased number of aligned interference signals in comparison with [12] and [11]. The results presented here fully characterize the NDT for a 4×4 network with cache capabilities at the transmitter side.

II. System Model

We consider a wireless network with 4 transmitters $\{\mathrm{Tx}_1, \mathrm{Tx}_2, \mathrm{Tx}_3, \mathrm{Tx}_4\}$ and 4 receivers $\{\mathrm{Rx}_1, \mathrm{Rx}_2, \mathrm{Rx}_3, \mathrm{Rx}_4\}$ (see Figure 1), and a library of N files $\{W_1, W_2, \ldots, W_N\}$, each consisting of F bits. Each transmitter is equipped with a cache memory of size MF bits. We impose the condition $4M \geq N$, so that the transmitters' caches are sufficient to collaboratively cache all the files in the library. We refer to the global normalized cache size at the transmitters as $t \triangleq 4M/N$, where $t \in [1, 4]$, and limit our attention to integer t values, while the extension to non-integer values will follow from memory-sharing [8].

The system operates in two phases; in the ***placement phase***, all the caches in the network are filled without the knowledge of user demands. The cache content of Tx_i at the end of the placement phase is denoted by a binary sequence P_i of length MF, $\forall i \in [4]\}$, where we denote the set $\{1, \ldots, N\}$ by $[N]$ for any positive integer N. The cache placement function that maps the library to the transmitters caches in a centralized manner is known by all the nodes, while each transmitter knows only the contents of its own cache.

The receivers reveal their requests after the placement phase, and the ***delivery phase*** follows. Let W_{d_j} denote the file requested by Rx_j, $\forall j \in [4]$, and $\mathbf{d} \triangleq [d_1, \ldots, d_4] \in [N]^4$ denote the vector of receivers' demands. The delivery phase takes place over an independent and identically distributed (i.i.d.) additive white Gaussian noise interference channel. The signal received at Rx_j at time t is:

$$Y_j(t) = \sum_{i=1}^{4} h_{ji} X_i(t) + Z_j(t), \tag{1}$$

where $X_i(t) \in \mathbb{C}$ represents the signal transmitted by Tx_i, $h_{ji} \in \mathbb{C}$ represents the channel coefficient between Rx_j and Tx_i, and $Z_j(t)$ is the additive Gaussian noise term at Rx_j. We assume that the channel coefficients $\mathbf{H} \triangleq \{h_{i,j}\}_{i\in[4],j\in[4]}$, and the demand vector $\mathbf{d}$ are known by all the transmitters and receivers.

Each transmitter Tx_i, $\forall i \in [4]$, maps the demand vector $\mathbf{d}$, the channel matrix $\mathbf{H}$ and its own cache contents P_i to a channel input vector of length L, $\mathbf{X_i} = [X_i(1), \ldots, X_i(L)]$, where L is a function of file size F. We impose an average power constraint P on each transmitted codeword, i.e., $\frac{1}{L}\|\mathbf{X}_i\|^2 \leq P$.

Receiver Rx_j, $\forall j \in [4]$, decodes its desired file W_{d_j} using $\mathbf{d}, \mathbf{H}$ and the corresponding channel output $\mathbf{Y}_j = [Y_j(1), \ldots, Y_j(L)]$. Let $\hat{W}_j$ denote its estimate of W_{d_j}.The error probability is defined as:

$$P_e = \max_{\mathbf{d}\in[N]^4} \max_{j\in[4]} \Pr\left(\hat{W}_j \neq W_{d_j}\right). \tag{2}$$

We now introduce the proposed delivery metric, NDT, first used in [5], [14], which accounts for the worst-case latency in the delivery phase.

Definition 1. *Delivery time per bit* $\Delta(t, P)$ is *achievable*, if there exists a sequence of codes, indexed by file size F, such that $P_e \to 0$ as $F \to \infty$, and

$$\Delta(t, P) = \liminf_{F\to\infty} \frac{L}{F}, \tag{3}$$

Definition 2. For a given family of codes achieving a delivery time per bit of $\Delta(t, P)$, the *normalized delivery time* (NDT) of the family of codes in the high signal-to-noise ratio (SNR) regime is defined as:

$$\delta(t) \triangleq \lim_{P\to\infty} \frac{\Delta(t, P)}{1/\log P}. \tag{4}$$

The optimal NDT is defined as:

$$\delta^*(t) \triangleq \inf\{\delta(t) : \delta(t) \text{ is } \textit{achievable}\}.$$

Our goal in this paper is to characterize the optimal NDT, $\delta^*(t)$, for the 4×4 network with caches at the transmitter side.

III. Main result

Our main result is the following theorem, which presents the optimal NDT, $\delta^*(t)$, for a 4×4 network for any cache capacity M at the transmitter side, or equivalently, for any value of $t \in [1, 4]$. The proof of the theorem is presented in Section IV.

Theorem 1. *For a centralized cache-aided 4×4 interference network with a library of N files, and transmitter cache capacity of MF bits each, the following NDT is achievable for a global cache size $t = 4M/N$:*

$$\delta^*(t) = \begin{cases} -7/12t + 7/3 & \text{if } 1 \leq t \leq 2 \\ -1/12t + 4/3 & \text{if } 2 < t \leq 4 \end{cases}. \tag{5}$$

The placement and delivery algorithms for the scheme that achieves the above NDT performance is outlined in Section IV. The optimality of this scheme follows from the comparison of the achievable NDT with the lower bound provided in [14] when the fronthaul links are ignored.

IV. Achievable scheme

In this section, we develop a coding scheme that exploits ZF and IA jointly in a 4×4 centralized cache-aided interference network with cache memories at the transmitter side. We present our scheme for integer t values, that is, for $t = 1, 2$ and 4. The achievable NDT for any other t value follows from memory-sharing between the integer points [8].

A. Placement Phase

We use the same algorithm as in [9] for the placement phase, which divides each file into $\binom{4}{t}$ equal-size non-overlapping subfiles, which are indexed as $W_{i,\mathcal{V}}$, for $i \in [N]$ and $\mathcal{V} \subset [4]$, $|\mathcal{V}| = t$. Subfile $W_{i,\mathcal{V}}$ is stored at all transmitters $k \in \mathcal{V}$. Each transmitters stores $\binom{3}{t-1}$ disjoint partitions of each file, fulfilling the cache size constraint.

Consider, for example, $M = 2$ and $N = 4$, i.e., $t = 2$. According to the placement phase explained above, file W_1 is divided into 6 equal-size subfiles as follows:

$$W_{1,12},\ W_{1,13},\ W_{1,14},\ W_{1,23},\ W_{1,24},\ \text{and}\ W_{1,34},$$

where Subfile $W_{1,23}$ denotes the subfile of file W_1 stored at Tx_2 and Tx_3. Same partition applies to files W_2, W_3 and W_4.

B. Delivery Algorithm

We present the delivery scheme separately for $t = 1, 2$, and 4, as different t values require exploiting different interference management schemes. We do not present the scheme for $t = 3$ as it turns out that the optimal performance for $t = 3$ can be obtained through memory-sharing between $t = 2$ and $t = 4$. Without loss of optimality, we assume that each receiver requests a different file from the database, which corresponds to the worst case demand.

1) $t = 1$: In this case, the transmitters collectively have just enough cache capacity to store the whole database; and therefore, they all cache distinct contents. According to the placement phase explained above, each file of the database is split into four different subfiles: $W_n = (W_{n,1}, W_{n,2}, W_{n,3}, W_{n,4}),\ \forall n \in [N]$. When each receiver requests a distinct file from the library, each of the transmitters has one subfile intended for each of the receivers. This corresponds to a 4×4 X-Channel. Thus, the placement phase has transformed the interference channel into an X-Channel. For this type of channels, using the IA technique of [16], it can be shown that a NDT of $\delta(1) = 7/4$ can be achieved.

2) t=4: This point corresponds to the trivial scenario, in which all the database is available at each of the transmitter caches ($M = N$). Therefore, in the delivery phase, the transmitters can fully cooperate and act as a multiple-antenna transmitter with 4 antennas. Hence, they can leverage zero-forcing to cancel the interference at the unintended receivers, resulting in the NDT of $\delta(4) = 1$.

3) t=2: For this point, each of the transmitters is able to store half of the database, allowing us to use a combination of IA and ZF schemes. We explain the proposed transmission scheme for the delivery phase on an example. Consider the following network configuration: $M = 2$ and $N = 4$ with the file library denoted by $\{A, B, C, D\}$. During the placement phase, file A is divided into $\binom{4}{2} = 6$ subfiles as follows:

$$A_{12},\ A_{13}, A_{23}, A_{14},\ A_{24},\ \text{and}\ A_{34},$$

where subfile A_{ij} is cached by transmitters Tx_i, Tx_j. The same file division and placement scheme is applied to files B, C and D as well. Note that each transmitter caches 3 subfiles, each of size $F/6$, from every file, fully utilizing their cache capacities.

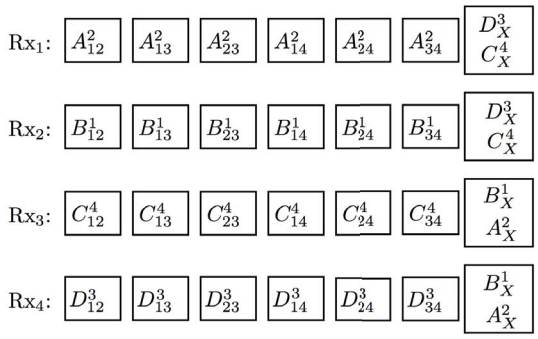

Fig. 2: Alignment of the subfiles in (6) at different receivers.

In the delivery phase, we assume, without loss of generality, that receivers 1, 2, 3 and 4 request files A, B, C and D, respectively (i.e., the worst case demand combination). Let A_{ij}^k denote the subfiles of file A that have been cached by transmitters Tx_i, Tx_j, while k indicates the receiver at which the transmitters Tx_i and Tx_j zero-force this subfile. Using this notation, the whole database can be represented with the following subfiles:

$$\begin{aligned} &A_{12}^2,\ A_{13}^2,\ A_{23}^2,\ A_{14}^2,\ A_{24}^2,\ A_{34}^2,\\ &B_{12}^1,\ B_{13}^1,\ B_{23}^1,\ B_{14}^1,\ B_{24}^1,\ B_{34}^1,\\ &C_{12}^4,\ C_{13}^4,\ C_{23}^4,\ C_{14}^4,\ C_{24}^4,\ C_{34}^4,\\ &D_{12}^3,\ D_{13}^3,\ D_{23}^3,\ D_{14}^3,\ D_{24}^3,\ D_{34}^3. \end{aligned} \tag{6}$$

Note that, each of these subfiles is zero-forced at a receiver that has not requested it. For example, each subfile of file A in (6) is zero-forced at Rx_2, while it is requested by Rx_1. In the same way, subfiles of file B, requested by Rx_2, are zero-forced at Rx_1.

Next, we focus on the transmission of subfile A_{12}^2. This subfile is intended to be zero-forced at Rx_2; hence, Tx_1 and Tx_2 precode it with scaling factors h_{22} and $-h_{21}$, respectively. It will be received at receivers Rx_1, Rx_2, Rx_3 and Rx_4 with the equivalent channel gains: $h_{11}h_{22} - h_{12}h_{21}$, $h_{21}h_{22} - h_{22}h_{21}$, $h_{31}h_{22} - h_{32}h_{21}$ and $h_{41}h_{22} - h_{42}h_{21}$, respectively. These channel gains can be restated in terms of the channel matrix minors, thus, receivers Rx_1, Rx_2, Rx_3 and Rx_4 receive subfile A_{12}^2 with scaling factors $M_{34,34},\ 0,\ -M_{14,34}$ and -$M_{13,34}$, respectively[1].

We now annalyze the interference caused by the transmission of subfile A_{12}^2. Rx_1 requests file A, thus subfile A_{12}^2 is intended for this receiver. Due to ZF pre-coding, this subfile does not cause interference at Rx_2. However, signals corresponding to A_{12}^2 do cause interference at Rx_3 and Rx_4. Based on the choice of ZF receivers for the subfiles in (6),

[1] $M_{i,j}$ is the minor of the channel matrix $\mathbf{H}$, which is defined as the determinant of the submatrix of $\mathbf{H}$ formed by deleting the i-th row and j-th column of $\mathbf{H}$.

the transmission of A subfiles cause interference only at Rx_3 and Rx_4. Following a similar reasoning, the subfiles of file B interfere only at Rx_3 and Rx_4, while both C and D act as interference at Rx_1 and Rx_2. All these subfiles are transmitted simultaneously, and at this point, the potential benefits of IA in this setting are conspicuous. If we accomplish to align all the interfering signals into the same subspace at each receiver (see Figure 2), we would achieve a NDT of $\delta(2) = 7/6$.

Next, we present the construction of the IA scheme. To guarantee the decodability of files, we precode files C and D with the scaling factor β. The purpose of this scaling factor is explained later. As mentioned above, the interference at Rx_1 and Rx_2 is caused by the subfiles of files C and D. C and D subfiles arrive at Rx_1 with equivalent channel coefficients $\beta M_{23,X}$ and $\beta M_{24,X}$, $\forall X \subset [4]$ with $|X| = 2$. While these same files arrive at Rx_2 with equivalent channel coefficients $\beta M_{13,X}$ and $\beta M_{14,X}$, $\forall X \subset [4]$ with $|X| = 2$. The goal is to align all these interfering signals in the same subspace at each interfered receiver. To do so, the real IA technique presented in [16] will be used. Defining the interference vector as the absolute value of the equivalent channel coefficients with which, the interference signals arrive at both receivers, we can construct:

$$
\begin{aligned}
&u^{(1)} \triangleq \\
&[\beta M_{23,34},\ \beta M_{23,14},\ \beta M_{23,23},\ \beta M_{23,24},\ \beta M_{23,13},\ \beta M_{23,12} \\
&\beta M_{24,34},\ \beta M_{24,14},\ \beta M_{24,23},\ \beta M_{24,24},\ \beta M_{24,13},\ \beta M_{24,12} \\
&\beta M_{13,34},\ \beta M_{13,14},\ \beta M_{13,23},\ \beta M_{13,24},\ \beta M_{13,13},\ \beta M_{13,12} \\
&\beta M_{14,34},\ \beta M_{14,14},\ \beta M_{14,23},\ \beta M_{14,24},\ \beta M_{14,13},\ \beta M_{14,12}].
\end{aligned} \tag{7}
$$

Vector $u^{(1)}$ is the input to the function $\mathcal{G}(u)$ in [16, Section IV]. We define $\mathcal{G}_N(u)$ as follows: let u be a vector consisting of K components; then we have:

$$\mathcal{G}_N(u) \triangleq (u_1^{n_1}, \ldots, u_K^{n_K}, 1 \leq n_1, \ldots, n_K \leq N).$$

The function $\mathcal{G}_N(u)$ provides the monomials that are used as beamforming directions for the transmission of subfiles. These subfiles are transmitted using the integer constellation consisting of points $\mathbb{Z}_Q$ [16, Section V]. As a result, subfiles of file A and B are streams that consist of points from the constellation:

$$\sum_{v \in \mathcal{G}_N(u^{(1)})} v\mathbb{Z}_Q.$$

Similarly, the interference at Rx_3 and Rx_4 is caused by subfiles A_X^2 and B_X^1, $\forall X \subset [4]$, with $|X| = t$. These subfiles are received with equivalent channel coefficients $M_{14,X}$, $M_{24,X}$, $M_{13,X}$ and $M_{23,X}$. The resulting interference vector is:

$$
\begin{aligned}
&u^{(2)} \triangleq \\
&[M_{14,34},\ M_{14,24},\ M_{14,14},\ M_{14,23},\ M_{14,13},\ M_{14,12} \\
&M_{24,34},\ M_{24,24},\ M_{24,14},\ M_{24,23},\ M_{24,13},\ M_{24,12} \\
&M_{13,34},\ M_{13,24},\ M_{13,14},\ M_{2134,23},\ M_{13,13},\ M_{13,12} \\
&M_{23,34},\ M_{23,24},\ M_{23,14},\ M_{23,23},\ M_{23,13},\ M_{23,12}].
\end{aligned} \tag{8}
$$

Again, we can construct a constellation that is scaled using the transmission directions provided by the monomials $\mathcal{G}_N(u^{(2)})$ created using the interference vector in (8), resulting in the following signal constellation:

$$\sum_{v \in \mathcal{G}_N(u^{(2)})} v\mathbb{Z}_Q.$$

Focusing now on the receivers, we want to assess whether the interfering signals have been aligned, and if the requested subfiles arrive with independent channel coefficients, so that their decodability is guaranteed. Starting with Rx_1, we see that the desired files A_{12}^2, A_{13}^2, A_{23}^2, A_{14}^2, A_{24}^2 and A_{34}^2 are received with equivalent channel coefficients $M_{34,34}$, $M_{34,24}$, $M_{34,14}$, $M_{34,23}$, $M_{34,13}$ and $M_{34,12}$, respectively. So the signal constellation at Rx_1 is:

$$
\begin{aligned}
C_D \triangleq\ & M_{34,34} \sum_{v \in \mathcal{T}_L(u^{(2)})} v\mathbb{Z}_Q + M_{34,24} \sum_{v \in \mathcal{G}_N(u^{(2)})} v\mathbb{Z}_Q \\
&+ M_{34,14} \sum_{v \in \mathcal{G}_N(u^{(2)})} v\mathbb{Z}_Q + M_{34,23} \sum_{v \in \mathcal{G}_N(u^{(2)})} v\mathbb{Z}_Q \\
&+ M_{34,13} \sum_{v \in \mathcal{G}_N(u^{(2)})} v\mathbb{Z}_Q + M_{34,12} \sum_{v \in \mathcal{G}_N(u^{(2)})} v\mathbb{Z}_Q.
\end{aligned} \tag{9}
$$

Regarding the interference, the subfiles C_{12}^4, C_{13}^4, C_{23}^4, $C_{14,3}^4$, C_{24}^4, C_{34}^4, D_{12}^4, D_{13}^4, D_{23}^4, D_{14}^4, D_{24}^4 and D_{34}^4 are received with the equivalent channel coefficients $\beta M_{14,34}$, $\beta M_{14,24}$, $\beta M_{14,14}$, $\beta M_{14,23}$, $\beta M_{14,13}$, $\beta M_{14,12}$ $\beta M_{24,34}$, $\beta M_{24,24}$, $\beta M_{24,14}$, $\beta M_{24,23}$, $\beta M_{24,13}$ and $\beta M_{24,12}$, respectively. Remember that these subfiles are transmitted using beforming vectors v that are obtained from the function of monomials $\mathcal{G}_N(u^{(1)})$, which has been constructed using the same equivalent channel coefficients. Hence, for these subfiles the received interference constellation is given by:

$$
\begin{aligned}
C_I \triangleq\ & \beta M_{14,34} \sum_{v \in \mathcal{G}_N(u^{(1)})} v\mathbb{Z}_Q + \beta M_{14,24} \sum_{v \in \mathcal{G}_N(u^{(1)})} v\mathbb{Z}_Q \\
&+ \beta M_{14,14} \sum_{v \in \mathcal{G}_N(u^{(1)})} v\mathbb{Z}_Q + \beta M_{14,23} \sum_{v \in \mathcal{G}_N(u^{(1)})} v\mathbb{Z}_Q + \\
&+ \beta M_{14,13} \sum_{v \in \mathcal{G}_N(u^{(1)})} v\mathbb{Z}_Q + \beta M_{14,12} \sum_{v \in \mathcal{T}_(u^{(1)})} v\mathbb{Z}_Q \\
&+ \beta M_{24,34} \sum_{v \in \mathcal{G}_N(u^{(1)})} v\mathbb{Z}_Q + \beta M_{24,24} \sum_{v \in \mathcal{T}_(u^{(1)})} v\mathbb{Z}_Q \\
&+ \beta M_{24,14} \sum_{v \in \mathcal{G}_N(u^{(1)})} v\mathbb{Z}_Q + \beta M_{24,23} \sum_{v \in \mathcal{T}_(u^{(1)})} v\mathbb{Z}_Q \\
&+ \beta M_{24,13} \sum_{v \in \mathcal{G}_N(u^{(1)})} v\mathbb{Z}_Q + \beta M_{24,12} \sum_{v \in \mathcal{T}_(u^{(1)})} v\mathbb{Z}_Q \\
&\subset \sum_{v \in \mathcal{G}_{N+1}(u^{(1)})} v\mathbb{Z}_Q.
\end{aligned} \tag{10}
$$

Equation (10) proves that all the interfering signals have collapsed into the same constellation space. Thus, proving that all the received directions are independent is the only

remaining condition to satisfy the constrains of [16, Theorem 6]. The proof that the intended messages are received along independent directions is straightforward, as the equivalent channel scaling factors $M_{34,34}$, $M_{34,24}$, $M_{34,14}$, $M_{34,23}$, $M_{34,13}$ and $M_{34,12}$ do not have any contribution to $\mathcal{G}(u^{(2)})$, at the same time, these channel factors can be shown to be linerly independent. Regarding C_I, the received directions represent a subspace of the set of directions $\mathcal{G}_{N+1}(u^{(1)})$. Finally, the only condition that remains to be proved is that the directions of the intended messages C_D and the directions of the interference signals C_I do not overlap. This is where we use the scaling factor β. This scaling factor is chosen to be independent of the channel minors. As C_D is received along directions that are independent from β, we can assert that these two sets of directions do not overlap, ensuring linear independence; and therefore, separability. As a result, all interfering subfiles have been aligned in a subspace of dimension $1/7$.

Due to the symmetry of the problem, this result can be extended to the interfering signals at Rx_2, Rx_3 and Rx_4, proving that IA is feasible at all the receivers. Note that 6 subfiles are transmitted, and all the interfering signals collapse into the same subspace (see Figure 2), leading to the achieved NDT of $\delta(2) = 7/6$.

The proposed joint IA-ZF scheme implemented in this paper achieves the lowest NDT in the literature for the above 4×4 scenario with caches at the transmitters of normalized global size $t = 2$. It also matches the lower-bound on the NDT presented in [14]; and hence, the optimal NDT is characterized for all cache memory sizes for this particular setting.

A comparison of the optimal NDT performance for the 4×4 cache-aided interference network, characterized in this paper, with schemes proposed in other papers is presented in Fig. 3 for different transmitter global cache sizes t. The scheme presented in [14] is referred to as STN, and is plotted here for zero fronthaul link rate. The scheme proposed in [12] is referred to as XTL in the figure. We also include NDT lower bound of [14], which matches our scheme. The STS scheme corresponds to memory-sharing between the points $t = 1$ and $t = 4$, and does not exploit IA and ZF techniques simultaneously. The XTL scheme also exploits IA and ZF jointly for $t = 3$, at which point it meets the lower bound; however, the figure shows that the same performance can also be achieved memory-sharing between our proposed scheme for $t = 2$ and ZF for $t = 4$.

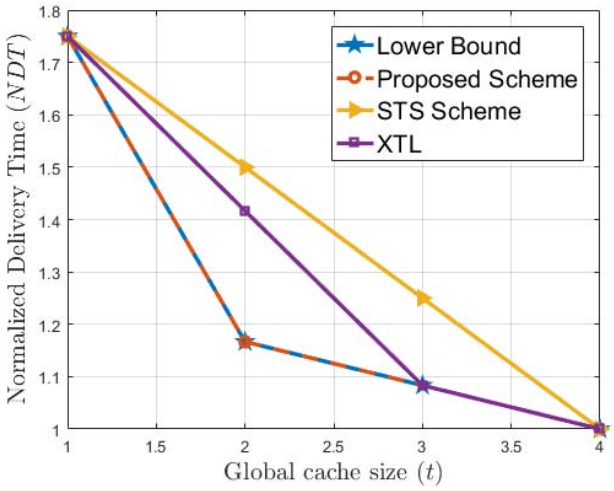

Fig. 3: NDT vs global cache size t performance comparison for different schemes in the literature.

V. Conclusions

A novel caching and transmission scheme has been proposed for a 4×4 centralized cache-aided wireless interference network with caches at the transmitters. The proposed scheme uses the ZF and IA techniques simultaneously to fully exploit the available cache contents for interference management. It is shown that the proposed caching and delivery scheme achieves the optimal NDT for this setting at all values of the global cache size t, fully characterizing the NDT for 4×4 interference networks. We highlight that the extension of this scheme to larger networks is non-trivial, and is the focus of our ongoing research efforts.

References

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

[2] S. S. Bidokhti, M. Wigger, and R. Timo, "Noisy broadcast networks with receiver caching," *arXiv:1605.02317*, 2016.

[3] M. M. Amiri and D. Gündüz, "Cache-aided data delivery over erasure broadcast channels," *arXiv:1702.05454*, 2017.

[4] ——, "Decentralized caching and coded delivery over gaussian broadcast channels," in *IEEE Int'l Symp. on Inform. Theory*, Aachen, Germany, Jul. 2017.

[5] J. Zhang and P. Elia, "Fundamental limits of cache-aided wireless BC: Interplay of coded-caching and CSIT feedback," *IEEE Trans. on Inf. Theory*, 2017.

[6] A. Ghorbel, M. Kobayashi, and S. Yang, "Content delivery in erasure broadcast channels with cache and feedback," *IEEE Trans. Inf. Theor.*, vol. 62, no. 11, pp. 6407–6422, Nov. 2016.

[7] M. Wigger, R. Timo, and S. Shamai, "Complete interference mitigation through receiver-caching in Wyners networks," in *IEEE Inf. Theory Workshop*, Cambridge, UK, Sep. 2016.

[8] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in *IEEE Int. Symp. Inf. Theory (ISIT)*, 2015, pp. 809–813.

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

[10] J. Hachem, U. Niesen, and S. Diggavi, "Degrees of freedom of cache-aided wireless interference networks," *arXiv:1606.03175*, 2016.

[11] J. Pujol, F. Tosato, and D. Gündüz, "Interference networks with caches at both ends," in *IEEE Int'l Conf. on Commun.*, Paris, France, 2017.

[12] F. Xu, M. Tao, and K. Liu, "Fundamental tradeoff between storage and latency in cache-aided wireless interference networks," *arXiv:1605.00203*, 2016.

[13] A. Sengupta, R. Tandon, and O. Simeone, "Cache aided wireless networks: Tradeoffs between storage and latency," in *Conf. on Inf. Science and Systems (CISS)*, 2016, pp. 320–325.

[14] ——, "Cloud and cache-aided wireless networks: Fundamental latency trade-offs," *arXiv:1605.01690*, 2016.

[15] A. Girgis, O. Ercetin, M. Nafie, and T. ElBatt, "Decentralized coded caching in wireless networks: Trade-off between storage and latency," *arXiv:1701.06673*, 2017.

[16] A. S. Motahari, S. Oveis-Gharan, M.-A. Maddah-Ali, and A. K. Khandani, "Real interference alignment: Exploiting the potential of single antenna systems," *IEEE Trans. Inform. Theory*, 2014.