Extended Placement Delivery Arrays for Multi-Antenna Coded Caching Scheme

K. K. Krishnan Namboodiri, Elizabath Peter and B. Sundar Rajan
Department of Electrical Communication Engineering, *Indian Institute of Science*, Bengaluru, India
Email: {krishnank,elizabathp, bsrajan}@iisc.ac.in

2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834422

***Abstract*—The multi-antenna coded caching problem, where the server having L transmit antennas communicating to K users through a wireless broadcast link, is addressed. In the problem setting, the server has a library of N files, and each user is equipped with a dedicated cache of capacity M. The idea of extended placement delivery array (EPDA), an array which consists of a special symbol $\star$ and integers in a set $\{1, 2, \ldots, S\}$, is proposed to obtain a novel solution for the aforementioned multi-antenna coded caching problem. From a (K, L, F, Z, S) EPDA, a multi-antenna coded caching scheme with K users, and the server with L transmit antennas, can be obtained in which the normalized memory $\frac{M}{N} = \frac{Z}{F}$, and the delivery time $T = \frac{S}{F}$. The placement delivery array (for single-antenna coded caching scheme) is a special class of EPDAs with $L = 1$. For the multi-antenna coded caching schemes constructed from EPDAs, it is shown that the maximum possible Degree of Freedom (DoF) that can be achieved is $t+L$, where $t = \frac{KM}{N}$ is an integer. Furthermore, two constructions of EPDAs are proposed: a) $K = t + L$, and b) $K = nt + (n-1)L$, $L \geq t$, where $n \geq 2$ is an integer. In the resulting multi-antenna schemes from those EPDAs achieve the full DoF, while requiring a subpacketization number $\frac{K}{\gcd(K,t,L)}$. This subpacketization number is less than that required by previously known schemes in the literature.**

I. Introduction

Coded caching, introduced in [1], is a promising technique to reduce the peak hour network traffic in content delivery networks by exploiting the caches at the user end. This technique involves two phases, namely the placement phase and the delivery phase. During the placement phase, the central server (having a library of N files) populates the caches at the user end. There are K users, each having a cache of size M units, where $0 \leq M \leq N$. In the delivery phase, each user requests a single file from the server library. After knowing those requests, the server transmits a message created by encoding across the requested file contents. The setting in [1] considered an error-free broadcast link between the server (with a single transmit antenna) and the users. The normalized capacity of the link was assumed to be one file per unit of time. The scheme in [1] can meet all the user demands in a normalized delivery time of $T = \frac{K-t}{t+1}$, where $t = \frac{KM}{N}$ is an integer. The term $t + 1$ in the denominator shows the multiplicative reduction obtained in the delivery time by the scheme in [1] compared to the conventional uncoded caching. The term $t + 1$ is termed as the global caching gain or the degree of freedom (DoF) achieved by the scheme. The DoF achieved by a scheme is defined as the number of users simultaneously served in unit time. In the further discussions, we refer to the scheme proposed in [1] as the MaN scheme. The MaN scheme achieves the DoF $t + 1$ by splitting the finite-length files into $\binom{K}{t}$ subfiles. The exponentially increasing subpacketization issue was addressed in [2] by constructing coded caching schemes from placement delivery arrays (PDA). But the reduction in subpacketization number was achieved at the cost of the DoF. Furthermore, in [2], the authors showed that the MaN scheme could also be obtained from a class of PDAs. We refer to that class of PDAs as the MaN PDA.

The coded caching problem with multiple transmit antennas (with the server) were explored in [3], [4], and proposed a scheme that achieves a DoF of $t + L$, where $t = \frac{KM}{N}$ is an integer and L is the number of transmit antennas. Under the assumption of uncoded cache placement and one-shot data delivery, $t + L$ (where $t + L \leq K$) is proven to be the optimal DoF [5]. For the schemes in [3], [4], to achieve this theoretical DoF, a file has to be split into, $\binom{K}{t}\binom{K-t-1}{L-1}$ subfiles. Even though those schemes enabled to achieve the optimal DoF, the subpacketization requirement was far more than the MaN scheme. The multi-antenna coded caching scheme proposed in [6] tackled the subpacketization bottleneck without paying in the DoF. The presented scheme required only a subpacketization number of $\binom{K/L}{t/L}$ while achieving the full DoF of $t + L$. But, the scheme is valid only when $L|K$ (L divides K) and $L|t$. The schemes in [7]–[9] address the multi-antenna coded caching problem under the subpacketization constraint. The scheme proposed in [8] achieves the DoF $t + L$ with a linear subpacketization number of $K(t + L)$, when $L \geq t$. In [9], the authors proposed a scheme that achieves the DoF $t + \alpha$, where $t \leq \alpha \leq L$. The subpacketization requirement is $\frac{K(t+L)}{\gamma^2}$, where $\gamma = \gcd(K, t, L)$, which is even less compared to the subpacketization requirement in [8]. Other interesting works in multi-antenna coded caching literature include [10], [11] that consider optimized precoder design, [12], [13] that reduce the complexity by limiting the number of messages received by the users in every time slot.

A. Contributions

In this work, we study the coded caching problem where the server has L transmit antennas. We provide an alternate solution for the multi-antenna coded caching problem by designing a special array termed as extended placement delivery array (EPDA). The PDAs proposed in [2] is a special class of EPDAs with $L = 1$. The technical contributions of this paper are summarized:

- A novel solution for the multi-antenna coded caching problem via EPDAs is provided. It is shown that for every EPDA there exists a corresponding multi-antenna coded caching scheme (Section III: Theorem 1).
- In [5], it is shown that for a multi-antenna coded caching scheme with L transmit antennas, the optimal DoF is $t+L$ (under the assumption of uncoded cache placement and

one-shot delivery). An alternate proof for the same is given in the context of EPDAs. That is, for the multi-antenna coded caching schemes constructed from regular EPDAs (an EPDA is said to be regular if all the integers in the array appear exactly the same number of times), it is shown that the maximum possible DoF that can be achieved is $t+L$ (Section III: Lemma 1).

- A class of EPDAs corresponding to the coded caching scheme proposed in [6] is identified (Section IV: Remark 2).
- A class of EPDAs corresponding to the multi-antenna coded caching scheme proposed in [6] is identified (Section IV: Remark 2).
- Two constructions of EPDAs are proposed. The resulting multi-antenna coded caching schemes are applicable
  1) when $K = t + L$, (Section IV: Construction I)
  2) when $K = nt + (n-1)L,\ L \geq t$, where $n \geq 2$ is an integer (Section IV: Construction II).
- The multi-antenna coded caching schemes resulting from Construction I and Construction II achieve full DoF $t + L$, while requiring a subpacketization number $\frac{K}{\gcd(K,t,L)}$ (Section IV: Theorem 2, Section IV: Theorem 3). The subpacketization number for those schemes is less than that required by previously known schemes in the literature.

Our primary focus is on the construction of multi-antenna coded caching schemes from EPDAs. We are mainly interested in DoF analysis (i.e., at higher SNR) rather than the design of sophisticated beamformers.

Due to space constraints, the proofs of some of the results stated in this paper are omitted. They are presented in detail and available online in [14].

B. Notations

For a positive integer n, $[n]$ denotes the set $\{1, 2, \ldots, n\}$. For two positive integers a, b such that $a \leq b$, $[a : b] = \{a, a+1, \ldots, b\}$. For integers $a, b \leq K$,

$$[a:b]_K = \begin{cases} \{a, a+1, \ldots, b\} & \text{if } a \leq b. \\ \{a, a+1, \ldots, K, 1, \ldots, b\} & \text{if } a > b. \end{cases}$$

For any two integers, i and K,

$$< i >_K = \begin{cases} i \ (mod\ K) & \text{if } i \ (mod\ K) \neq 0. \\ K & \text{if } i \ (mod\ K) = 0. \end{cases}$$

For two vectors $\mathbf{u}$ and $\mathbf{v}$, $\mathbf{u} \perp \mathbf{v}$ means that $\mathbf{v}^T\mathbf{u} = 0$, and $\mathbf{u} \not\perp \mathbf{v}$ means that $\mathbf{v}^T\mathbf{u} \neq 0$. All the vectors are assumed to be column vectors by default. Finally, the symbol $\mathbb{C}$ represents a complex number.

II. System Model and Problem Formulation

The system model consists of a central server having a library of N files, $W_{[1:N]} \triangleq \{W_n : n \in [N]\}$ each of size 1 unit. We consider a multiple-input, single-output (MISO) broadcast channel in which the server with L transmit antennas communicates to K users, each having a single receive antenna. The wireless shared link is assumed to be of capacity 1 file per unit of time. Furthermore, each node in the system (the server and K users) has perfect channel state information (CSI). Each user is equipped with a dedicated cache of capacity M units, where $0 \leq M \leq N$. The system operates in two phases: the placement phase and the delivery phase. In the placement phase, the server stores some of the file contents in the caches of the users without knowing the future demands. In general, the placement can be coded or uncoded. In this work, we concentrate only on schemes with uncoded placement. The contents stored in cache k is denoted as $\mathcal{Z}_k$. In the delivery phase, each user requests a single file from the server. Let $\mathbf{d} = (d_1, d_2, \ldots, d_K)$ be the demand vector, i.e., the k^{th} user requests for the file W_{d_k}, for every $k \in [K]$. After knowing the demand vector, the server makes transmission vectors $\{\mathbf{x}(\tau)\}_{\tau=1}^{T}$. That is, the server transmission is for T time slots. During the time slot τ, the server transmits $\mathbf{x}(\tau)$, where $\mathbf{x}(\tau) \in \mathbb{C}^L$. At time slot τ, the k^{th} user receives,

$$y_k(\tau) = \mathbf{h}_k^T \mathbf{x}(\tau) + w_k(\tau) \tag{1}$$

where $\mathbf{h}_k^T \in \mathbb{C}^L$ is the channel vector and $w_k(\tau) \sim \mathbb{C}\mathcal{N}(0,1)$ is the additive noise (complex normal distributed with zero mean and unit variance) observed at user k at time slot τ. Define the channel matrix $\mathbf{H} := [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K]$. We assume that the received signal to noise ratio (SNR) is high as in [3], [6]–[8], and neglect the additive noise component during the analysis. The correctness of the scheme implies that, using the local cache content $\mathcal{Z}_k$ and the received coded files $\{y_k(\tau)\}_{\tau=1}^{T}$, user k should be able to decode the demanded file W_{d_k}. The coded caching system under the aforementioned setting is called the (K, L, M, N) multi-antenna coded caching system.

The number of time slots T taken by the server to meet the user demands is termed the delivery time. It is already proven in the literature [5] that under the assumption of uncoded placement and one-shot delivery, the optimal delivery time is $T^* = \frac{K-t}{t+L}$, where $t = \frac{KM}{N}$ is an integer. For any other value of M (when $\frac{KM}{N}$ is not an integer), the delivery time is the lower convex envelope of adjacent corner points (connect the delivery times at the adjacent integer t values with a straight line). The term $1 - \frac{t}{K} = 1 - \frac{M}{N}$ denotes the local caching gain, which is achieved simply from the caching. The extra multiplicative reduction factor $t+L$ is termed as the Degree of Freedom (DoF) achieved. The ultimate goal of the coded caching problem is to design the placement and the delivery phases jointly such that the DoF is maximized (maximizing the DoF is the same as minimizing the delivery time).

While designing coded caching schemes, one other important parameter to be considered is the subpacketization number. The subpacketization number is defined as the number of subfiles into which a file is divided during the coded caching scheme. It is always desired to have a low subpacketization number.

III. Extended Placement Delivery Array (EPDA)

In this section, we propose the idea of extended placement delivery array to obtain multi-antenna coded caching schemes.

Definition 1. *Let $K, L(\leq K), F, Z, S$ be positive integers. An array $\mathbf{A} = [a_{j,k}]$, $j \in [F]$, $k \in [K]$ consisting of the symbol $\star$ and positive integers in $[S]$ is called a (K, L, F, Z, S) extended placement delivery array (EPDA) if it satisfies the following conditions:*
C1. The symbol $\star$ appears Z times in each column.
C2. Every integer in the set $[S]$ occurs at least once in $\mathbf{A}$.
C3. No integer appears more than once in any column.
C4. Consider the sub-array $\mathbf{A}^{(s)}$ of $\mathbf{A}$ obtained by deleting all

the rows and columns of $\mathbf{A}$ *that do not contain the integer* s. *Then for any* $s \in [S]$, *no row of* $\mathbf{A}^{(s)}$ *contains more than* L *integers.*

We present an example of an EPDA before discussing the construction of multi-antenna coded caching scheme from EPDAs. A $(K = 3, L = 2, F = 3, Z = 1, S = 2)$ EPDA is given:

$$\mathbf{A} = \begin{bmatrix} \star & 1 & 1 \\ 1 & \star & 2 \\ 2 & 2 & \star \end{bmatrix}$$

It is easy to verify that $\mathbf{A}$ satisfies the conditions *C1, C2* and *C3*. To verify *C4*, consider the sub-arrays $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$.

$$\mathbf{A}^{(1)} = \begin{bmatrix} \star & 1 & 1 \\ 1 & \star & 2 \end{bmatrix}, \quad \mathbf{A}^{(2)} = \begin{bmatrix} 1 & \star & 2 \\ 2 & 2 & \star \end{bmatrix}$$

In both $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, all the rows contain two integers (which is equal to L). Therefore, $\mathbf{A}$ satisfies the condition *C4* as well. Note that, in the sub-array corresponding to the integer 1, integer 2 is also present. In general, a sub-array $\mathbf{A}^{(s)}$ of a (K, L, F, Z, S) EPDA may contain the integers in $[S]\backslash\{s\}$ as well.

From a (K, L, F, Z, S) EPDA, $\mathbf{A} = [a_{j,k}]$, $j \in [F]$ and $k \in [K]$, a multi-antenna coded caching scheme with the server having L transmit antennas and the normalized cache size $\frac{M}{N} = \frac{Z}{F}$, can be obtained as follows:

1. **Placement phase:** The server divides each file into F subfiles of equal size. Thus, for every $n \in [N]$, we have, $W_n = \{W_{n,j} | j \in [F]\}$. The k^{th} user's cache is populated as follows:
$$\mathcal{Z}_k = \{W_{n,j} \mid a_{j,k} = \star, \ \forall n \in [N]\}. \tag{2}$$
For any $k \in [K]$ if $a_{j,k} = \star$, then it means that user k has access to the j^{th} subfile of all the files.
2. **Delivery phase:** Let $\mathbf{d} = (d_1, d_2, \ldots, d_K)$ be the demand vector. Assume that the integer s appears g_s times in $\mathbf{A}$. Let $a_{j_1,k_1} = a_{j_2,k_2} = \cdots = a_{j_{g_s},k_{g_s}} = s$. Then the server transmits $\mathbf{V}^s \cdot (W_{d_{k_1},j_1}, W_{d_{k_2},j_2}, \ldots, W_{d_{k_{g_s}},j_{g_s}})^T$, where $\mathbf{V}^s = (\mathbf{v}_1^s, \mathbf{v}_2^s, \ldots, \mathbf{v}_{g_s}^s)$ is a precoding matrix of size $L \times g_s$. For every $i \in [g_s]$, define $\mathcal{B}_i := \{\beta : a_{j_i,\beta} \neq \star, \ \beta \in \{k_1, \ldots, k_{i-1}, k_{i+1}, \ldots, k_{g_s}\}\}$. Then the i^{th} column of $\mathbf{V}^s$ is $\mathbf{v}_i^s \perp \mathbf{h}_\alpha$ for all $\alpha \in \mathcal{B}_i$, and $\mathbf{v}_i^s \not\perp \mathbf{h}_{k_i}$.

For the above placement and delivery strategies, we have the following theorem.

Theorem 1. *Corresponding to any* (K, L, F, Z, S) *EPDA, there exists a* (K, L, M, N) *multi-antenna coded caching scheme with* $\frac{M}{N} = \frac{Z}{F}$ *and subpacketization number* F. *Furthermore, the server can meet any user demand* $\mathbf{d}$ *with a delivery time* $T = \frac{S}{F}$.

Proof: The server stores Z subfiles of every file in the k^{th} user's cache (Eq:(2)) during the placement phase. Therefore, the normalized memory of the cache is $\frac{M}{N} = \frac{Z}{F}$. In the delivery phase, the server makes a transmission corresponding to every integer present in the EPDA considered. Thus there are S transmissions, each of size $\left(\frac{1}{F}\right)^{th}$ of a file. So, the normalized delivery time is $T = \frac{S}{F}$.

Now, it remains to show the decodability of the demanded files by the users. Consider user k which does not have the subfile $W_{d_k,j}$ from the placement phase. Assume that $a_{j,k} = s$ for some $s \in [S]$ (if $a_{j,k} = \star$, then $W_{n,j}$ would have been available for user k from the placement phase itself). Then the claim is that user k will receive the subfile $W_{d_k,j}$ from the transmission corresponding to the integer s. Consider the sub-array $\mathbf{A}^{(s)}$, and assume that $a_{j,k} = a_{j_2,k_2} = \cdots = a_{j_{g_s},k_{g_s}} = s$. Then the server transmission is $\mathbf{V}^s.(W_{d_k,j}, W_{d_{k_2},j_2}, \ldots, W_{d_{k_{g_s}},j_{g_s}})^T$. But, user k receives (neglecting the additive noise by the high-SNR assumption)

$$\begin{aligned} Y_k &= \mathbf{h}_k^T \mathbf{V}^s \cdot (W_{d_k,j}, W_{d_{k_2},j_2}, \ldots, W_{d_{k_{g_s}},j_{g_s}})^T \\ &= \mathbf{h}_k^T (\mathbf{v}_1^s, \mathbf{v}_2^s, \ldots, \mathbf{v}_{g_s}^s) \cdot (W_{d_k,j}, W_{d_{k_2},j_2}, \ldots, W_{d_{k_{g_s}},j_{g_s}})^T \\ &= \mathbf{h}_k^T \mathbf{v}_1^s W_{d_k,j} \\ &\quad + \sum_{\substack{i=2 \\ a_{j_i,k} \neq \star}}^{g_s} \left(\mathbf{h}_k^T \mathbf{v}_i^s\right) W_{d_{k_i},j_i} + \sum_{\substack{i=2 \\ a_{j_i,k} = \star}}^{g_s} \left(\mathbf{h}_k^T \mathbf{v}_i^s\right) W_{d_{k_i},j_i}. \end{aligned}$$

From the design of the precoding vectors, we have, $h_k^T \mathbf{v}_i^s = 0$ for all $i \in [2 : g_s]$ such that $a_{j_i,k} \neq \star$. In other words, the precoding vectors are designed such that whichever subfiles (subfiles involved in the transmission corresponding to the integer s) are not cached in the k^{th} cache will be nulled out from Y_k (except the subfile required for user k). Therefore, Y_k will be a linear combination of required subfile (for user k) and some other subfiles that are available in cache k. All the users know all the channel coefficients $\mathbf{H}$ completely, hence users k can decode $W_{d_k,j}$. Since k and j are arbitrary, all the users can decode their demanded files.

This completes the proof of Theorem 1. ∎

Example 1. *Consider the* $(K = 4, L = 2, F = 4, Z = 1, S = 4)$ *EPDA given below:*

$$\boldsymbol{A} = \begin{bmatrix} \star & 1 & 1 & 4 \\ 1 & \star & 2 & 2 \\ 3 & 2 & \star & 3 \\ 4 & 4 & 3 & \star \end{bmatrix}$$

Consider a coded caching scheme with $K = 4$ *users and the server having* $L = 2$ *transmit antennas. The server has* N *files* $W_n, n \in [N]$. *The placement and delivery is in accordance with* $\boldsymbol{A}$. *In the placement phase, the server divides each file into 4 subfiles,* $W_n = \{W_{n,1}, W_{n,2}, W_{n,3}, W_{n,4}\}$ *for all* $n \in [N]$. *The contents stored in each user's cache are:*

$$\mathcal{Z}_1 = \{W_{n,1} : n \in [N]\}, \quad \mathcal{Z}_2 = \{W_{n,2} : n \in [N]\},$$
$$\mathcal{Z}_3 = \{W_{n,3} : n \in [N]\}, \quad \mathcal{Z}_4 = \{W_{n,4} : n \in [N]\}.$$

Let $\mathbf{d} = (1, 2, 3, 4)$ *be the demand vector. Then the server makes the transmissions summarized in TABLE I.*

TABLE I: Transmissions in the delivery Phase: Example 1

Time Slot	Transmission
1	$\mathbf{V}^1 \cdot (W_{1,2}, W_{2,1}, W_{3,1})^T$
2	$\mathbf{V}^2 \cdot (W_{2,3}, W_{3,2}, W_{4,2})^T$
3	$\mathbf{V}^3 \cdot (W_{1,3}, W_{3,4}, W_{4,3})^T$
4	$\mathbf{V}^4 \cdot (W_{1,4}, W_{2,4}, W_{4,1})^T$

To design the precoding matrix $\mathbf{V}^1 = (\mathbf{v}_1^1, \mathbf{v}_2^1, \mathbf{v}_3^1)$, *consider the sub-array corresponding to the integer 1,*

$$\boldsymbol{A}^{(1)} = \begin{bmatrix} \star & 1 & 1 \\ 1 & \star & 2 \end{bmatrix}.$$

To design the precoding vector $\mathbf{v}_1^1$, consider the row in which the integer 1 is present in the first column of $\boldsymbol{A}^{(1)}$, that is the second row of $\boldsymbol{A}^{(1)}$. Find out all the columns (excluding the first column) in which an integer is present in the second row of $\boldsymbol{A}^{(1)}$. The integer 2 is present in the third column. Therefore, $\boldsymbol{v}_1^1 \perp \mathbf{h}_3$. Similarly, we can see that $\boldsymbol{v}_2^1 \perp \mathbf{h}_3$ and $\boldsymbol{v}_3^1 \perp \mathbf{h}_2$. By considering $\boldsymbol{A}^{(2)}$, $\boldsymbol{A}^{(3)}$ and $\boldsymbol{A}^{(4)}$, we can design the rest of the precoding matrices.

User 1, user 2 and user 3 benefit from the transmission corresponding to integer 1, since integer 1 is present in column 1, column 2 and column 3 of $\mathbf{A}$. Let us see how user 2 is benefiting from that transmission. User 2 receives,

$$\begin{aligned} Y_2 &= \mathbf{h}_2^T\mathbf{V}^1 \cdot (W_{1,2}, W_{2,1}, W_{3,1})^T \\ &= (\mathbf{h}_2^T\mathbf{v}_1^1, \mathbf{h}_2^T\mathbf{v}_2^1, \mathbf{h}_2^T\mathbf{v}_3^1) \cdot (W_{1,2}, W_{2,1}, W_{3,1})^T \\ &= (\mathbf{h}_2^T\mathbf{v}_1^1, \mathbf{h}_2^T\mathbf{v}_2^1, 0) \cdot (W_{1,2}, W_{2,1}, W_{3,1})^T \\ &= \mathbf{h}_2^T\mathbf{v}_1^1 W_{1,2} + \mathbf{h}_2^T\mathbf{v}_2^1 W_{2,1}. \end{aligned}$$

Since, user 2 has access to the subfile $W_{1,2}$, the user can decode the desired subfile $W_{2,1}$. More details are given in [14].

Remark 1. *If $L = 1$, no row of $\boldsymbol{A}^{(s)}$ should contain more than one integer. That is, every row in $\boldsymbol{A}^{(s)}$ will be consisting of one integer and rest all $\star$'s. That one integer will be s, since $\boldsymbol{A}^{(s)}$ is obtained by deleting all the rows and columns that do not contain the integer s. Therefore, $\boldsymbol{A}^{(s)}$ will have the form,*

$$\begin{bmatrix} s & \star & \dots & \star \\ \star & s & \dots & \star \\ \vdots & \vdots & \ddots & \vdots \\ \star & \star & \dots & s \end{bmatrix}$$

up to row and column permutation. That means, a (K, L, F, Z, S) EPDA becomes a (K, F, Z, S) PDA introduced in [2], when $L = 1$.

In a (K, L, F, Z, S) EPDA, if $g_1 = g_2 = \dots, g_S = g$, then the EPDA is said to be g-regular.

Definition 2. *An array $\mathbf{A}$ is said to be a g-regular (K, L, F, Z, S) EPDA if $\mathbf{A}$ satisfies the condition C2', in addition to C1, C3 and C4.*
C2': Each integer in $[S]$ should appear exactly g times in A.

In a multi-antenna coded caching scheme obtained from a g-regular (K, L, F, Z, S) EPDA, the positive integer g represents the number of users benefited from a transmission in any given time slot. In other words, g is the DoF achieved by the scheme. By *C2*, the regularity of an EPDA g is upper bounded by the number of columns K in the EPDA.

Lemma 1. *For a multi-antenna coded caching scheme obtained from a g-regular (K, L, F, Z, S) EPDA, the delivery time,*

$$T = \frac{K}{g}\left(1 - \frac{Z}{F}\right). \tag{3}$$

Furthermore, we have,

$$g \leq L + \frac{KZ}{F}. \tag{4}$$

Proof: Lemma 1, Eq: (3) can be proved by counting the number of integers in the array in different ways. Similarly, Eq: 4 can be proved by counting the number of stars in the array column-wise and row-wise. When $\frac{Z}{F} = \frac{M}{N} = \frac{t}{K}$ for some integer t, we have, $g \leq L + t$. The detailed proof is given in the extended version [14]. ■

IV. New Constructions

In this section, we introduce two constructions of EPDAs for certain values of K, L, F, Z, S. Using those EPDAs, we obtain the corresponding multi-antenna coded caching schemes. Before dealing with the constructions, we present the definition of u-row concatenation of arrays.

Definition 3. *The process of obtaining an array $\boldsymbol{A}_{F\times uK}$ by concatenating another array $\hat{\boldsymbol{A}}_{F\times K}$, row-wise, u times is referred to as u-row concatenation of $\hat{\mathbf{A}}$. This can be expressed as,*

$$\mathbf{A} = \underbrace{[\hat{\boldsymbol{A}}|\hat{\boldsymbol{A}}|\dots|\hat{\boldsymbol{A}}]}_{u\ times}.$$

Lemma 2. *The u-row concatenation of a (K, L, F, Z, S) EPDA results in a (uK, uL, F, Z, S) EPDA.*

Proof: Consider a (K, L, F, Z, S) EPDA $\mathbf{A}_1$. Let $\mathbf{A}$ be the array obtained by u-row concatenation of $\mathbf{A}_1$. It is easy to see that the array $\mathbf{A}$ satisfies *C1, C2* and *C3*, since $\mathbf{A}_1$ satisfies *C1, C2* and *C3*. Now to show $\mathbf{A}$ satisfies *C4*, consider an integer $s \in [S]$ present in $\mathbf{A}_1$. Then, the sub-array $\mathbf{A}^{(s)}$ is obtained by the u-row concatenation of $\mathbf{A}_1^{(s)}$. In any row of $\mathbf{A}_1^{(s)}$, there will be L integers at the maximum. It means that, no row of $\mathbf{A}^{(s)}$ will be containing more than uL integers. This is true for every $s \in [S]$. Therefore, $\mathbf{A}$ is a (uK, uL, F, Z, S) EPDA.

This completes the proof of Lemma 2. ■

Corollary 1. *The u-row concatenation of a (K, F, Z, S) PDA results in a (uK, u, F, Z, S) EPDA.*

Proof: The result follows from Lemma 2 since a (K, F, Z, S) PDA is a $(K, L = 1, F, Z, S)$ EPDA. ■

Remark 2. *The multi-antenna coded caching scheme presented in [6] can also be obtained from EPDA. The proposed scheme works with subpacketization number $\binom{K/L}{t/L}$ if $L|K$ and $L|t$, where t is a positive integer such that $\frac{M}{N} = \frac{t}{K}$. To see the corresponding EPDA representation, first consider the $(K', F', Z'S')$ MaN PDA (given in [2]) with $K' = K/L, F' = \binom{K/L}{t/L}, Z' = \binom{(K/L)-1}{(t/L)-1}$ and $S' = \binom{K/L}{(t/L)+1}$. The L-row concatenation of this (K', F', Z', S') PDA will give our desired $(K, L, F', Z'S')$ EPDA (follows from Corollary 1). Note that, in the resulting (K, L, M, N) multi-antenna coded caching scheme, the subpacketization number is $\binom{K/L}{t/L}$, and the delivery time is $\frac{S'}{F'} = \frac{K-t}{t+L}$ (DoF achieved is $t + L$).*

Now, we see two constructions of EPDAs with the maximum possible regularity $L + \frac{KZ}{F}$.

a) **Construction I**: A $(K, K - Z, K, Z, K - Z)$ EPDA.

In this construction, $F = K$, $L = K - Z$ and $S = K - Z$. We denote the EPDA with $\mathbf{A} = [a_{j,k}]$. Then, the $\star$'s appear in $\mathbf{A}$ as follows,

$$a_{j,k} = \star, \quad \forall (j, k) \in [K] \times [K] : (j - k)\ (mod\ K) < Z. \tag{5}$$

That is, in the k^{th} column of $\mathbf{A}$, $a_{k,k} = a_{<k+1>_K,k} = \dots = a_{<k+Z-1>_K,k} = \star$. Now, consider an integer $s \in [K - Z]$. Then s occurs in $\mathbf{A}$ as,

$$a_{j,k} = s, \text{ such that } j = (Z + s + k - 1)\,(mod\,K). \tag{6}$$

The array $\mathbf{A}$ obtained by the above construction is, in fact, a K-regular $(K, K-Z, K, Z, K-Z)$ EPDA (the proof that $\mathbf{A}$ is an EPDA is provided in [14]).

Theorem 2. *For a (K, L, M, N) multi-antenna coded caching scheme with $\frac{M}{N} = \frac{t}{K}$ and $L = K - t$, where t is an integer, the delivery time $T^* = \frac{K-t}{t+L} = \frac{L}{K}$ is achievable with a subpacketization number $\frac{K}{gcd(K,t,L)}$.*

Proof: Let $t = \frac{KM}{N}$ be an integer, and let $L = K - t$. Define $\gamma \triangleq \gcd(K, t, L)$. Let $\tilde{K} = \frac{K}{\gamma}, \tilde{L} = \frac{L}{\gamma}$ and $\tilde{t} = \frac{t}{\gamma}$. Since $K = t+L$, we have, $\tilde{K} = \tilde{t}+\tilde{L}$. Now construct a $(\tilde{K}, \tilde{L}, \tilde{K}, \tilde{t}, \tilde{L})$ EPDA $\mathbf{A}_1$ using Construction I. The γ-row concatenation of $\mathbf{A}_1$ results in a $(K, L, \tilde{K}, \tilde{t}, \tilde{L})$ EPDA $\mathbf{A}$. Now, in the multi-antenna coded caching scheme corresponding to $\mathbf{A}$, $\frac{M}{N} = \frac{\tilde{t}}{\tilde{K}} = \frac{t}{K}$. Using Theorem 1, in the resulting scheme, we have, the delivery time, $T = \frac{\tilde{L}}{\tilde{K}} = \frac{K-t}{t+L}$, and the subpacketization number is $\frac{K}{\gamma}$. This completes the proof of Theorem 2. ■

b) **Construction II**: A $\left(K, \frac{K-nZ}{n-1}, K, Z, (n-1)K\right)$ EPDA with $K \geq (2n-1)Z$, where n is an integer greater than 1.

We construct an EPDA $\mathbf{B} = [b_{j,k}]$ with parameters $F = K, L = \frac{K-nZ}{n-1}$, $S = (n-1)K$ and $K \geq (2n-1)Z$, where n is an integer greater than 1. The $\star$'s appear in $\mathbf{B}$ as follows,

$$b_{j,k} = \star, \quad \forall (j,k) \in [K] \times [K] : (j-k)\ (mod\ K) < Z. \quad (7)$$

That is, in the k^{th} column of $\mathbf{B}$, $b_{k,k} = b_{<k+1>_K,k} = \cdots = b_{<k+Z-1>_K,k} = \star$. Now, consider an integer $s \in [(n-1)K]$. Let $s = pK + q$, where $p \in [0 : n-2], q \in [1 : K]$. If p is even, then s occurs in $\mathbf{B}$ as follows:

$$b_{q,<\frac{p}{2}(Z+L)+q+i>_K} = s \quad \forall i \in [L], \quad (8)$$

$$b_{<\frac{p}{2}(Z+L)+Z+q>_K,<q-Z+i>_K} = s \quad \forall i \in [Z]. \quad (9)$$

If p is odd, then s occurs in $\mathbf{B}$ as follows:

$$b_{q,<(\frac{p-1}{2})(Z+L)+L+q+i>_K} = s \quad \forall i \in [Z], \quad (10)$$

$$b_{<(\frac{p+1}{2})(Z+L)+q>_K,<q-Z+i>_K} = s \quad \forall i \in [L]. \quad (11)$$

The array $\mathbf{B}$ obtained by the above construction is a $(Z+L)$-regular $(K, L, K, Z, (n-1)K)$ EPDA with $L = \frac{K-nZ}{n-1}$ if $K \geq (2n-1)Z$ (the proof that $\mathbf{B}$ is an EPDA is provided in [14]).

Theorem 3. *For a (K, L, M, N) multi-antenna coded caching scheme with $\frac{M}{N} = \frac{t}{K}$ and $K = nt + (n-1)L;\ L \geq t$, where t and $n \geq 2$ are integers, the delivery time $T^* = \frac{K-t}{t+L} = n-1$ is achievable with a subpacketization number $\frac{K}{gcd(K,t,L)}$.*

Proof: Let $t = \frac{KM}{N}$ be an integer, and let $n \geq 2$ be an integer such that $K = nt + (n-1)L$. Define $\gamma \triangleq \gcd(K, t, L)$. Let $\tilde{K} = \frac{K}{\gamma}, \tilde{L} = \frac{L}{\gamma}$ and $\tilde{t} = \frac{t}{\gamma}$. Since $K = nt+(n-1)L$, we have, $\tilde{K} = n\tilde{t}+(n-1)\tilde{L}$. Now construct a $(\tilde{K}, \frac{\tilde{K}-n\tilde{t}}{n-1}, \tilde{K}, \tilde{t}, (n-1)\tilde{K})$ EPDA $\mathbf{B}_1$ using Construction II. The γ-row concatenation of $\mathbf{B}_1$ results in a $(K, \frac{K-nt}{n-1}, \tilde{K}, \tilde{t}, \tilde{L})$ EPDA $\mathbf{B}$. Now, in the multi-antenna coded caching scheme corresponding to $\mathbf{B}$, $\frac{M}{N} = \frac{\tilde{t}}{\tilde{K}} = \frac{t}{K}$. Using Theorem 1, the delivery time obtained in the resulting scheme, $T = \frac{(n-1)\tilde{K}}{\tilde{K}} = n-1 = \frac{K-t}{t+L}$, and the subpacketization number is $\frac{K}{\gamma}$. This completes the proof of Theorem 3. ■

Example 2. *The array $\mathbf{A}$ is a 4-regular $(4, 3, 4, 1, 3)$ EPDA obtained using Construction I. The array $\mathbf{B}$ is a 3-regular $(4, 2, 4, 1, 4)$ EPDA obtained using Construction II, where $K = nZ + (n-1)L$ with $n = 2$.*

$$\boldsymbol{A} = \begin{bmatrix} \star & 3 & 2 & 1 \\ 1 & \star & 3 & 2 \\ 2 & 1 & \star & 3 \\ 3 & 2 & 1 & \star \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \star & 1 & 1 & 4 \\ 1 & \star & 2 & 2 \\ 3 & 2 & \star & 3 \\ 4 & 4 & 3 & \star \end{bmatrix}$$

V. Comparison

In this section, we compare the multi-antenna coded caching schemes resulting from Construction I and Construction II to different multi-antenna coded caching schemes in the literature [3], [6], [8], [9]. All these schemes achieve the optimal DoF of $t+L$. Therefore, the comparison is in terms of subpacketization number.

Consider the multi-antenna coded caching scheme obtained from Construction I. The scheme requires a file to be split into $\frac{K}{\gcd(K,t,L)}$ subfiles. When $K = t + L$, the multi-antenna scheme proposed in [3] has a subpacketization number $\binom{K}{t}\binom{K-t-1}{L-1} = \binom{K}{t}$. For any $t \in [2 : K-2]$, the subpacketization number for our proposed scheme is strictly less than that for the scheme proposed in [3]. The scheme in [6] is applicable only when $L|K$ and $t|K$. When both those conditions are valid, along with $K = t + L$, then $t = iL$ and $K = (i + 1)L$ for some positive integer i. In that case, the subpacketization number required for the scheme from Construction I and for the scheme in [6] is $\frac{K}{L}$ (because $\gcd(K, t, L) = L$). The multi-antenna coded caching schemes in [8] and [9] are applicable when $L \geq t$. If $K = t+L$, then the subpacketization number required for the scheme in [8] is, $K(t + L) = K^2$, and that required for the scheme in [9] is $\left(\frac{K}{\gamma}\right)^2$, where $\gamma = \gcd(K, t, L)$. Both these subpacketization requirements are high compared to our proposed scheme.

Now, we compare the multi-antenna coded caching scheme obtained from Construction II with the schemes in [3], [6], [8], [9] in terms of subpacketization number. The comparison is for the case $K = nt + (n-1)L;\ L \geq t$, where $n \geq 2$ is an integer. The scheme from Construction II has a subpacketization number $\frac{K}{\gcd(K,t,L)}$. When $K = nt + (n-1)L$, the multi-antenna scheme proposed in [3] has a subpacketization number of $\binom{K}{t}\binom{K-t-1}{L-1} = \binom{K}{t}\binom{(n-1)L-1}{L-1}$. The subpacketization number is increasing exponentially with K in the case of [3]. Meaningful comparison with the scheme in [6] is when $L = t$, since our proposed scheme works if $L \geq t$, and the scheme in [6] works if $L|t$. When $t = L$, and $K = nt + (n-1)L$, our proposed scheme and the scheme in [6] require a subpacketization number $\frac{K}{L}$ (because $\gcd(K, t, L) = L$). The multi-antenna coded caching scheme in [8] requires a subpacketization number of $K(t + L)$ which is strictly greater than the subpacketization number required for our proposed scheme. Similarly, the scheme in [9] requires a subpacketization number $\frac{K(t+L)}{\gamma^2}$ which is also strictly greater than $\frac{K}{\gamma}$, since $t+L > \gamma$, where $\gamma = \gcd(K, t, L)$.

Acknowledgment

This work was supported partly by the Science and Engineering Research Board (SERB) of Department of Science and Technology (DST), Government of India, through J.C. Bose National Fellowship to B. Sundar Rajan, and by the Ministry of Human Resource Development (MHRD), Government of India, through Prime Minister's Research Fellowship (PMRF) to K. K. Krishnan Namboodiri and Elizabath Peter.

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, May 2014.

[2] Q. Yan, M. Cheng, X. Tang and Q. Chen, "On the Placement Delivery Array Design for Centralized Coded Caching Scheme," *IEEE Transactions on Information Theory*, vol. 63, no. 9, pp. 5821–5833, Sept. 2017.

[3] S. P. Shariatpanahi, S. A. Motahari, and B. H. Khalaj, "Multi-server coded caching," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 7253–7271, 2016.

[4] S. P. Shariatpanahi, G. Caire, and B. H. Khalaj, "Physical-Layer schemes for Wireless Coded Caching," *IEEE Transactions on Information Theory*, vol. 65, no. 5, pp. 2792–2807, 2019.

[5] E. Lampiris, A. Bazco-Nogueras and P. Elia, "Resolving a feedback bottleneck of multi-antenna coded caching," *arXiv: 1811.03935v2 [cs.IT]*, Apr 2020.

[6] E. Lampiris and P. Elia, "Adding Transmitters Dramatically Boosts Coded-Caching Gains for Finite File Sizes," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1176–1188, June 2018.

[7] M. Salehi, A. Tölli, S. P. Shariatpanahi, and J. Kaleva, "Subpacketization-rate trade-off in multi-antenna coded caching," in *IEEE Global Communications Conference, GLOBECOM* 2019, pp. 1–6.

[8] M. Salehi, A. Tölli and S. P. Shariatpanahi, "A Multi-Antenna Coded Caching Scheme with Linear Subpacketization," in *IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[9] M. Salehi, E. Parrinello, S. P. Shariatpanahi, P. Elia and A. Tölli, "Low-Complexity High-Performance Cyclic Caching for Large MISO Systems," *arXiv: 2009.12231v2 [cs.IT]*, Oct 2021.

[10] A. Tölli, S. P. Shariatpanahi, J. Kaleva, and B. Khalaj, "Multicast Beamformer Design for Coded Caching," in *IEEE International Symposium on Information Theory*, 2018, pp. 1914–1918.

[11] A. Tölli, S. P. Shariatpanahi, J. Kaleva, and B. H. Khalaj, "Multi-antenna interference management for coded caching," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2091–2106, 2020.

[12] J. Zhao, M. M. Amiri and D. Gündüz, "Multi-Antenna Coded Content Delivery With Caching: A Low-Complexity Solution," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7484–7497, Nov. 2020.

[13] E. Lampiris, J. Zhang, O. Simeone, and P. Elia, "Fundamental Limits of Wireless Caching under Uneven-Capacity Channels," *arXiv:1908.04036v2 [cs.IT]*, Jan. 2020.

[14] K. K. K. Namboodiri, E. Peter and B. S. Rajan, "Extended Placement Delivery Arrays for Multi-Antenna Coded Caching Scheme," *arXiv:2201.10585v1 [cs.IT]*, Jan 2022.