IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 23, 2022, accepted March 8, 2022, date of publication March 16, 2022, date of current version April 5, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3160187

The Tensor Multi-Linear Channel and Its Shannon Capacity

DIVYANSHU PANDEY, (Graduate Student Member, IEEE), AND HARRY LEIB, (Life Senior Member, IEEE)
Department of Electrical and Computer Engineering, McGill University, Montreal, QC H3A 0E9, Canada

Corresponding author: Divyanshu Pandey (divyanshu.pandey@mail.mcgill.ca)

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) titled "Tensor modulation for space-time-frequency communication systems" under Grant RGPIN-2016-03647.

ABSTRACT Tensors are multi-way arrays which can be used to model systems spanning many domains. This work proposes to use tensors for characterizing, analyzing, and designing multi-domain communication systems. Most modern day communication systems make use of coding and modulation across different domains such as space, frequency, time. Hence a unified mathematical framework characterizing such a multiple-domain system in an intuitive manner is well needed. In this paper, we present such a unified framework that characterizes a communication system with N input domains, M output domains and an $M + N$ domains multi-linear tensor channel. The proposed framework is generic where the physical interpretation of the domains is system specific. We illustrate a few examples from multi-antenna multi-carrier and multi-user systems that fit the proposed framework. Assuming a fixed tensor channel, we provide an information theoretic analysis by deriving its Shannon capacity and input power allocation under a variety of power constraints. In this paper we show how the tensor framework's suitability to mathematically describe a family of power constraints can be used to design and analyze various multiple domain communication systems. The tensor based approach extends water-filling from a matrix setting to tensors, encapsulating the effects of multiple domains thereby allowing joint multi-domain precoding. We show that the capacity pre-log for a tensor channel increases exponentially in the number of domains, indicating the potential of tensor based multi-domain communication systems to provide the large information transmission rates envisaged for 5G and beyond systems. We also show the application of the tensor framework in characterizing the capacity and rate regions of multi-user MIMO channels. Both multiple access and interference channels are considered where the tensor based approach leads to a coordinated users transmission scheme. Such a scheme ensures higher achievable sum rates as compared to independent user transmissions.

INDEX TERMS Tensors, MIMO channels, multi-user MIMO, Shannon capacity, tensor SVD and EVD.

I. INTRODUCTION

A tensor is a multi-domain array that can be seen as an Nth order generalization of a vector or a matrix, where a vector is a tensor of order one and a matrix is a tensor of order two [1]. Tensors were introduced in the early nineteenth century with applications in Physics where such mathematical structures can be seen as a mapping from a linear space to another whose coordinates transform multi-linearly under a change of bases [2]. Later, tensors found applications in psychometrics in the sixties with the work of Tucker [3] as an extension of two-way data analysis to higher-order datasets, and in chemometrics in the eighties [4], [5]. In the last few decades, tensors as an extension of matrices have found extensive applications in various engineering disciplines including computer vision [6], [7], data mining [8], [9], machine learning [10], neuroscience [11], signal processing [12]–[14] and multi-linear system theory [15], [16]. Tensors provide a unified and intuitive framework to represent processes with dependencies on more than two indices. Through a tensor based approach, we can develop models which capture interactions between various parameters

The associate editor coordinating the review of this manuscript and approving it for publication was Filbert Juwono.

enhancing the understanding of their mutual effects. A detailed summary of tensor algebra results can be found in many publications such as [1], [9], [10], [17]–[20].

In the field of communication systems, so far tensors have been used primarily from a signal processing point of view. Most modern communication systems use multi-domain modulation and coding schemes for effectively exploiting resources in transmission and reception. A common example in wireless communications is that of using the space domain, i.e. multiple input multiple output (MIMO) systems based on multiple antennas, to achieve a significant increase in channel capacity [21]. Systems including MIMO in conjunction with multi-carrier techniques such as Orthogonal Frequency Division Multiplexing (OFDM) [22], [23], Generalized Frequency Division Multiplexing (GFDM) [24], [25], Filter Bank Multi-carrier (FBMC) [26], etc., have been widely researched over past few years. Techniques have been developed to improve link reliability through space-time, space-frequency, and space-time-frequency coding methods [27] that exploit diversity in all spatial, temporal and frequency domains. Hence the received and transmitted signals have an inherent multi-domain structure which can be represented by tensors. Such a multi-domain approach using tensors has been considered in [28], [29], where the focus has been around developing blind detection techniques at the receiver. A tensor based space-time-frequency coding structure is proposed in [30] for a MIMO OFDM-Code Division Multiple Access (CDMA) system where signals are represented using order-5 tensors. Applications of tensors are also being considered for modelling various 5G and 6G communication technologies such as Intelligent Reflecting Surfaces [31], massive MIMO [32], [33], millimeter wave [34], [35], MIMO relay systems [36]. Recently, a tensor based framework for a general multi-domain communication system using the tensor contracted product has been developed in [37]. In addition, [37] also considers tensor based joint-domain equalization at the receiver to combat inter-domain interferences. Hence the tensor approach of handling a multi-domain communication system can lead to design of new and improved transmission and reception schemes. This paper considers information theoretic aspects of discrete version of the tensor-modelled multi-domain communication systems of [37].

The domains in a communication system can represent space, time, frequency, users, propagation delay, spreading sequence etc. This necessitates a generic unified mathematical framework with which we can model any multi-domain communication system, and hence tensors naturally come into play. Having such a framework would not only provide a mathematical basis for existing schemes, but would also act as a stepping stone for developing new and improved systems spanning multiple domains such as G.hn networks [38], 5G and beyond [39], as well as 6G [40]. The main contribution of this paper is the generic tensor based system model where an order $M + N$ channel tensor is used to connect an order N input tensor with an order M output tensor using the Einstein product. We find the Shannon capacity of such a deterministic tensor channel with various input constraints, revealing its exponential increase with the number of domains. A tensor framework endowed with the Einstein product has been used in [15], [16] to extend basic linear system concepts to the multi-linear realm. Our work introduces the multi-linear tensor channel and extends basic information theoretic concepts to such models.

The initial motivation to use tensors for our purpose stems from their unique suitability to retain the distinction between multiple domains in the system model, thereby allowing a convenient representation of a variety of power constraints across domains. Even though tensor entities occur naturally in multi-domain communication systems, it should be noted that in principle a tensor can be represented using a matrix or a vector. For instance, the slices of a third order tensor can be stacked together to form a bigger matrix. Such matrix representations are sometimes used in order to leverage the well established linear algebra concepts for analysis. However, representing a naturally occurring higher order tensor using a lower order array such as a matrix or vector collapses the distinct multiple indices which are used to identify the domains. Thus in order to restore the identifiability of domains, it becomes imperative to use tensors [41]. In addition, the tensor based approach leads to joint domain precoding and power allocation operations which perform much better than matrix approaches using per domain processing. Also, the proposed framework in this paper can be used to find the capacity of multi-user MIMO channels. The tensor model allows to keep users and antennas as distinct domains in a multi-user system. Hence the problem of finding the channel capacity under per user power constraints for any number of users can be approached using the tensor framework.

Capacity of the MIMO matrix channel was analyzed in [21] by converting the MIMO channel into parallel non-interfering scalar channels through a singular value decomposition of the channel matrix. Since the late 1990's and early 2000's a significant research effort has been invested in extending the work in [21] under various assumptions. A detailed summary can be found in [42]. Capacity behaviour of OFDM based MIMO systems has been explored in [43] by concatenating the transmit vectors over different antennas and different sub-carriers into a single vector and using a block-diagonal channel matrix. There has not been an attempt so far to quantify the capacity for a channel spanning more than two domains in terms of the number of domains itself and also to find the multi-domain power allocation required to achieve capacity, which calls for a tensor based approach. One impediment en-route such application of tensors in Information Theory is the inability to completely diagonalize a tensor and thereby convert a higher order channel into scalar channels. Complete diagonalization of any tensor with tensor decomposition techniques such as Parallel Factorization (PARAFAC) or Tucker decomposition is not achievable in general [1]. However, recent work in [18], [44] proposes

a tensor singular value decomposition (SVD) and eigenvalue decomposition (EVD) approach as special cases of Tucker decomposition for solving multi-linear systems of equations. A generalization of such a tensor SVD and EVD is presented in this paper which is used to compute the channel capacity.

The capacity of a MIMO GFDM channel modelled as a sixth order tensor, under sum power constraint is considered in [45]. Further, [46] presents the notion of tensor partial response signalling as a means to generate multi-domain signals with desired spectral and cross-spectral properties. The trade-off between domains of a communication system as revealed through the tensor approach is also studied in [46]. In [47], the capacity of tensor channels under a family of power constraints, which includes per antenna or sum power constraints as its specific cases, is considered. Going beyond [47], in this paper we consider the channel characterization as a tensor and its Shannon capacity under different power constraints in detail with new results and applications. In particular, we consider the application of the tensor framework for MIMO multiple access and interference channels. We use our proposed solution to characterize rate regions and quantify achievable sum rates in a multi-user system under user cooperation with per user power constraints. A comparison of these results with other methods in literature which assume independent user transmissions is also included. We also present examples of several multi-domain systems such as multi-user MIMO OFDM where the channel can be characterized as higher order tensor. For any such multi-domain system, we present an algorithmic approach of approximating the optimal input covariance under a variety of input power constraints along with a discussion on its computational complexity. For MIMO GFDM systems, we use our proposed solution to prescribe a precoding scheme under per antenna power constraints and analyze its BER performance. We also present results on multiplexing gain achieved by the tensor channel. Through several numerical examples, we present the capacity behaviour of tensor channels for various channel sizes, with different domain power constraints, and channel with correlated entries.

This paper is organized as follows: A brief review of tensor algebra is presented in section II. Section III introduces the discrete system model for a multi-domain communication system using tensors. The notion of a channel as a higher order tensor is proposed with an example. Section IV presents the required information theoretic notions for tensors. Section V considers the capacity of tensor channels under a family of power constraints, an algorithm for finding the input covariance along with its complexity, and multiplexing gain provided by tensor channels. Section VI presents numerical and simulation results showing an application of our work to MIMO GFDM systems. Section VII considers multi-user MIMO Multiple Access Channels (MAC) and Interference Channels (IC) using the tensor framework. The paper is concluded in Section VIII.

II. ELEMENTS OF TENSOR ALGEBRA

In this section we present essential tensor algebra results needed for this paper. A detailed treatment of tensor algebra can be found in [1], [18], [37] and references within.

A. NOTATIONS

Throughout this paper, deterministic vectors are represented using lower-case underline fonts, e.g. $\underline{\mathrm{x}}$, matrices using upper-case fonts, e.g. X and tensors using upper-case calligraphic fonts, e.g. $\mathcal{X}$. Their corresponding random quantities are denoted by bold fonts, e.g. $\underline{\mathbf{x}}$, $\mathbf{X}$ and $\boldsymbol{\mathcal{X}}$ for random vectors, matrices and tensors respectively. The individual entries of a tensor are denoted by indices in subscript, e.g. the (i, j, k)th element of a third order tensor $\mathcal{X}$ is denoted by $\mathcal{X}_{i,j,k}$. A colon in subscript is used to indicate all elements of a mode, e.g. $\mathcal{X}_{:,j,k}$ represents all the elements of first mode corresponding to the jth second and the kth third mode. The nth element in a sequence is denoted by a superscript in parentheses, e.g. $\mathcal{A}^{(n)}$ denotes the nth tensor in a sequence of tensors. Expectation is denoted by $\mathbb{E}[\cdot]$, entropy by $\mathcal{H}(\cdot)$ and mutual information by $\mathcal{I}(\cdot;\cdot)$. The set of complex numbers is denoted by $\mathbb{C}$. An all-zero tensor is represented by $0_{\mathcal{T}}$.

B. BASIC TENSOR OPERATIONS AND DEFINITIONS

Tensors are multi-way arrays with components indexed by N indices also known as *modes*. The number of modes, N is called the *order* of the tensor. The set of all tensors of size $I_1 \times \cdots \times I_K$ over $\mathbb{C}$ forms a linear space, denoted as $\mathbb{T}_{I_1,\ldots,I_K}(\mathbb{C})$.

Definition 1 (Matricization Transformation): Let us denote the linear space of $P \times Q$ matrices over $\mathbb{C}$ as $\mathbb{M}_{P,Q}(\mathbb{C})$. For an order $K = N + M$ tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\cdots\times I_N\times J_1\times\cdots\times J_M}$, the transformation $f_{I_1,\ldots,I_N|J_1,\ldots,J_M} : \mathbb{T}_{I_1,\ldots,I_N,J_1,\ldots,J_M}(\mathbb{C}) \Rightarrow \mathbb{M}_{I_1\cdot I_2\cdots I_N, J_1\cdot J_2\cdots J_M}(\mathbb{C})$ with $f_{I_1,\ldots,I_N|J_1,\ldots,J_M}(\mathcal{A}) = \mathrm{A}$ is defined component-wise as [18]:

$$\mathcal{A}_{i_1,i_2,\ldots,i_N,j_1,j_2,\ldots,j_M} \xrightarrow{f_{I_1,\ldots,I_N|J_1,\ldots,J_M}} \mathrm{A}_{i_1+\sum_{k=2}^{N}(i_k-1)\prod_{l=1}^{k-1} I_l,\, j_1+\sum_{k=2}^{M}(j_k-1)\prod_{l=1}^{k-1} J_l} \tag{1}$$

This transformation is referred to as matricization, or matrix unfolding by partitioning the indices into two disjoint subsets [17]. The vectorization operation as defined in [48] can be seen as a specific case of (1) by using $J_1 = \cdots = J_M = 1$. The bar in subscript of $f_{I_1,\ldots,I_N|J_1,\ldots,J_M}$ represents the partitioning after N modes of an $N + M$ order tensor where first N modes correspond to the rows of the representing matrix, and the last M modes correspond to the columns of the representing matrix. This mapping is bijective [49], and it preserves addition and scalar multiplication operations. Hence the linear spaces $\mathbb{T}_{I_1,\ldots,I_N,J_1,\ldots,J_M}(\mathbb{C})$ and $\mathbb{M}_{I_1\cdot I_2\cdots I_{N-1}\cdot I_N, J_1\cdot J_2\cdots J_{M-1}\cdot J_M}(\mathbb{C})$ are isomorphic and the transformation $f_{I_1,\ldots,I_N|J_1,\ldots,J_M}$ is an isomorphism between the linear spaces.

Definition 2 (Tensor Contracted product [17]): Consider two tensors $\mathcal{X} \in \mathbb{C}^{I_1\times\cdots\times I_M\times J_1\times\cdots\times J_N}$ and $\mathcal{Y} \in \mathbb{C}^{I_1\times\cdots\times I_M\times K_1\times\cdots\times K_P}$. A contraction between the two

tensors along their M common modes denoted by $\langle \mathcal{X}, \mathcal{Y} \rangle_{\{1,\ldots,M;1,\ldots,M\}}$ leads to a resulting tensor, $\mathcal{Z} \in \mathbb{C}^{J_1 \times \cdots \times J_N \times K_1 \times \cdots \times K_P}$ given by:

$$\mathcal{Z}_{j_1,\ldots,j_N,k_1,\ldots,k_P} = \sum_{i_1,\ldots,i_M} \mathcal{X}_{i_1,\ldots,i_M,j_1,\ldots,j_N} \mathcal{Y}_{i_1,\ldots,i_M,k_1,\ldots,k_P} \tag{2}$$

In general for the contracted product, the modes to be contracted need not be consecutive and can be in any location. However, the dimensions of the corresponding modes must be equal. For instance, tensors $\mathcal{X} \in \mathbb{C}^{I \times J \times K \times L}$ and $\mathcal{Y} \in \mathbb{C}^{J \times P \times L \times Q}$ can be contracted as $\mathcal{Z} = \langle \mathcal{X}, \mathcal{Y} \rangle_{\{2,4;1,3\}}$ to generate $\mathcal{Z} \in \mathbb{C}^{I \times K \times P \times Q}$ with elements $\mathcal{Z}_{i,k,p,q} = \sum_{j,l} \mathcal{X}_{i,j,k,l} \mathcal{Y}_{j,p,l,q}$. Other tensor products, such as the Einstein product or the mode-n product of tensor with matrices can be seen as special cases of the tensor contracted product.

Definition 3 (Einstein product [18]): For any positive integer N, the Einstein product is defined using the operation $*_N$ by

$$(\mathcal{A} *_N \mathcal{B})_{i_1,\ldots,i_P,j_1,\ldots,j_M} = \sum_{k_1,\ldots,k_N} \mathcal{A}_{i_1,i_2,\ldots,i_P,k_1,\ldots,k_N} \mathcal{B}_{k_1,\ldots k_N,j_1,j_2,\ldots,j_M} \tag{3}$$

where $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_P \times K_1 \times \ldots \times K_N}$ and $\mathcal{B} \in \mathbb{C}^{K_1 \times \cdots \times K_N \times J_1 \times \cdots \times J_M}$.

Einstein product is a tensor contracted product where contraction is over the last N consecutive modes of $\mathcal{A}$ and first N consecutive modes of $\mathcal{B}$. The outer and inner products can be seen as special cases of Einstein product. For tensors $\mathcal{X}, \mathcal{Y} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ and $\mathcal{Z} \in \mathbb{C}^{J_1 \times \cdots \times J_M}$, we have:

$$\begin{aligned} &\textit{Inner Product} : \langle \mathcal{X}, \mathcal{Y} \rangle = \mathcal{X} *_N \mathcal{Y} = \mathcal{Y} *_N \mathcal{X} \\ &= \sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \mathcal{X}_{i_1,\ldots,i_N} \mathcal{Y}_{i_1,\ldots,i_N} \end{aligned} \tag{4}$$

$$\begin{aligned} &\textit{Outer Product} : (\mathcal{X} \circ \mathcal{Z})_{i_1,i_2,\ldots,i_N,j_1,j_2,\ldots,j_M} \\ &= \mathcal{X}_{i_1,i_2,\ldots,i_N} \mathcal{Z}_{j_1,j_2,\ldots,j_M} = \mathcal{X} *_0 \mathcal{Z} \end{aligned} \tag{5}$$

where $\langle \mathcal{X}, \mathcal{Y} \rangle$ is a scalar and $(\mathcal{X} \circ \mathcal{Z}) \in \mathbb{C}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$.

Definition 4 (mode-n product [50]): The mode-n product of a tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots I_N}$ with a matrix $\mathrm{U} \in \mathbb{C}^{J \times I_n}$ is denoted by $\mathcal{X} \times_n \mathrm{U}$ and is defined as

$$(\mathcal{X} \times_n \mathrm{U})_{i_1,i_2,\ldots,i_{n-1},j,i_{n+1},\ldots,i_N} = \sum_{i_n=1}^{I_n} \mathcal{X}_{i_1,i_2,\ldots,i_N} \mathrm{U}_{j,i_n} \tag{6}$$

where $(\mathcal{X} \times_n \mathrm{U}) \in \mathbb{C}^{I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N}$.

Definition 5 (Square tensors [49]): A tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$ is called a square tensor if $N = M$ and $I_k = J_k$ for $k = 1, \ldots, N$.

For square tensors $\mathcal{A}, \mathcal{B}$ of size $I \times J \times I \times J$, it was shown in [18] that $f_{I,J|I,J}(\mathcal{A} *_2 \mathcal{B}) = f_{I,J|I,J}(\mathcal{A}) \cdot f_{I,J|I,J}(\mathcal{B})$ where $\cdot$ refers to the usual matrix multiplication. This property can be easily generalized to square or rectangular non-square tensors of any order in the form of the following Lemma [51]:

Lemma 1: For tensors $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$ and $\mathcal{B} \in \mathbb{C}^{J_1 \times \cdots \times J_M \times K_1 \times \cdots \times K_P}$ under the transformation from (1)*, the following holds*:

$$\begin{aligned} &f_{I_1,\ldots,I_N|K_1,\ldots,K_P}(\mathcal{A} *_M \mathcal{B}) \\ &= f_{I_1,\ldots,I_N|J_1,\ldots,J_M}(\mathcal{A}) \cdot f_{J_1,\ldots,J_M|K_1,\ldots,K_P}(\mathcal{B}). \end{aligned} \tag{7}$$

Definition 6: (Pseudo-diagonal Tensors): Any tensor $\mathcal{D} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$ of order $N+M$ is called pseudo-diagonal if its transformation to a matrix, $\mathrm{D} = f_{I_1,\ldots,I_N|J_1,\ldots,J_M}(\mathcal{D}) \in \mathbb{C}^{I_1 \cdots I_N \times J_1 \cdots J_M}$ results into a diagonal matrix such that $\mathrm{D}_{i,j}$ is non-zero only when $i = j$. In case of a rectangular D, such a matrix is often called main diagonal or principal diagonal matrix [52].

A square tensor $\mathcal{D} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ is pseudo-diagonal if all its entries $\mathcal{D}_{i_1,\ldots,i_N,j_1,\ldots,j_N}$ are zero except when $i_1 = j_1, i_2 = j_2, \ldots, i_N = j_N$. In [18], [53], such a tensor is termed as diagonal. However, we tend to call it pseudo-diagonal for our purpose of discussion, so as to make a clear distinction from the diagonal tensor definition more widely found in literature. A diagonal tensor is one whose entries $\mathcal{D}_{i_1,\ldots,i_N}$ are zero except when $i_1 = i_2 = \cdots = i_N$ [1]. An illustration of the difference between diagonal and pseudo-diagonal tensor is presented in [37]. Note that pseudo-diagonality for a non-square tensor as in Definition 6 is defined with respect to partition after N modes. For instance, if we refer to a third order tensor as pseudo-diagonal, it is important to specify whether it is pseudo-diagonal with respect to partition after the first mode or the second mode. Hence to avoid overload of notation in this paper whenever we write a tensor explicitly as order $N+M$ or order $2N$ and call it pseudo-diagonal, then it is with respect to a partition after N modes.

Using the Einstein product and Lemma 1, we can extend many linear algebra concepts to a multi-linear setting [18], [49]. For instance, a square pseudo-diagonal tensor of order $2N$ denoted as $\mathcal{I}_N \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ is called an *identity* tensor if $(\mathcal{I}_N)_{i_1,\ldots,i_N,i_1\ldots,i_N} = 1$ for all $i_1, \ldots, i_N$. The tensor $\mathcal{X}^{-1} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ is an *inverse* of a square tensor of same size, $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ if $\mathcal{X} *_N \mathcal{X}^{-1} = \mathcal{X}^{-1} *_N \mathcal{X} = \mathcal{I}_N$. The *Hermitian* of a tensor $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$ is a tensor $\mathcal{B} \in \mathbb{C}^{J_1 \times \cdots \times J_M \times I_1 \times \cdots \times I_N}$ with entries $\mathcal{B}^*_{j_1,j_2,\ldots,j_M,i_1,i_2,\ldots,i_N} = \mathcal{A}_{i_1,i_2,\ldots,i_N,j_1,j_2,\ldots,j_M}$ and it is denoted as $\mathcal{A}^H$. Also, the *transpose* is denoted as $\mathcal{A}^T$. A square tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ is called *Hermitian* if $\mathcal{X} = \mathcal{X}^H$ and is called unitary if $\mathcal{X}^H *_N \mathcal{X} = \mathcal{X} *_N \mathcal{X}^H = \mathcal{I}_N$.

Tensor SVD: For $\mathcal{A} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$, the *tensor Singular Value Decomposition (SVD)* of $\mathcal{A}$ has the form $\mathcal{A} = \mathcal{U} *_N \mathcal{D} *_M \mathcal{V}^H$ where $\mathcal{U} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ and $\mathcal{V} \in \mathbb{C}^{J_1 \times \cdots \times J_M \times J_1 \times \cdots \times J_M}$ are unitary tensors and $\mathcal{D} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times J_1 \times \cdots \times J_M}$ is a pseudo-diagonal tensor whose non-zero values are the singular values of $\mathcal{A}$. The existence of such a tensor SVD can be shown by using Lemma 1 and the SVD of matrix $f_{I_1,\ldots,I_N|J_1,\ldots,J_M}(\mathcal{A})$ [18]. Note that Tucker decomposition of a tensor can also be seen as a higher order SVD of tensors [50] and has found many applications particularly in extracting low rank structures in higher dimensional

data [54]. The relation between Tucker decomposition and the tensor SVD presented here is explained in [18].

Tensor EVD: Let $\mathcal{A} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$, $\mathcal{X} \in \mathbb{C}^{I_1\times\cdots\times I_N}$, $\lambda \in \mathbb{C}$, where $\mathcal{X}$ and λ satisfy $\mathcal{A} *_N \mathcal{X} = \lambda\mathcal{X}$, then we call $\mathcal{X}$ and λ as *eigentensor* and *eigenvalue* of $\mathcal{A}$ respectively. The *tensor Eigenvalue Decomposition (EVD)* of a Hermitian tensor $\mathcal{A} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ is given as $\mathcal{A} = \mathcal{U} *_N \mathcal{D} *_N \mathcal{U}^H$ where $\mathcal{U} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ is a unitary tensor and $\mathcal{D} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ is a square pseudo-diagonal tensor with its non-zero values being the eigenvalues of $\mathcal{A}$ and $\mathcal{U}$ containing the eigentensors of $\mathcal{A}$. As a generalization of matrix eigenvalues, several other definitions exist in literature for tensor eigenvalues [55], [56]. But most of these definitions apply to super-symmetric tensors which are defined as a class of tensors that are invariant under any permutation of their indices [56]. Such an approach has applications in Physics and Mechanics [56]. But there is no single generalization to the tensor case that preserves all the properties of matrix eigenvalues [57]. For our purposes, we have presented a particular generalization from [44] which can be used for any square tensor, irrespective of symmetry in its elements.

The eignevalues of a Hermitian tensor are real. Subsequently, a Hermitian tensor $\mathcal{X} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$ is defined as *positive semi-definite*, denoted by $\mathcal{X} \succeq 0$, if all its eigenvalues are non-negative, and as *positive definite*, denoted by $\mathcal{X} \succ 0$, if all its eigenvalues are strictly positive. *Determinant* of $\mathcal{X}$, denoted by $\det(\mathcal{X})$, is defined as the product of its eigenvalues. *Trace* of a tensor $\mathcal{A}$, denoted as $\mathrm{tr}(\mathcal{A})$, is defined as the sum of its pseudo-diagonal entries. Also, the following properties can be established using Lemma 1:

(a) *Associativity*: For tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\cdots\times I_P\times J_1\times\cdots\times J_N}$, $\mathcal{B} \in \mathbb{C}^{J_1\times\cdots\times J_N\times K_1\times\cdots\times K_M}$ and $\mathcal{C} \in \mathbb{C}^{K_1\times\cdots\times K_M\times T_1\times\cdots\times T_Q}$, we have

$$\begin{aligned}(\mathcal{A} *_N \mathcal{B}) *_M \mathcal{C} &= \mathcal{A} *_N (\mathcal{B} *_M \mathcal{C})\\ (\mathcal{A} *_N \mathcal{B}) \circ \mathcal{C} &= \mathcal{A} *_N (\mathcal{B} \circ \mathcal{C})\end{aligned} \tag{8}$$

(b) *Commutativity*: Einstein product is not commutative in general. However for the specific case where the contraction is over all the N modes of one of the tensors, say for tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\cdots\times I_P\times J_1\times\cdots\times J_N}$ and $\mathcal{B} \in \mathbb{C}^{J_1\times\cdots\times J_N}$, we have

$$\mathcal{A} *_N \mathcal{B} = \mathcal{B} *_N \mathcal{A}^T \tag{9}$$

(c) For invertible tensors $\mathcal{A}$ and $\mathcal{B} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$, we have

$$(\mathcal{A} *_N \mathcal{B})^{-1} = \mathcal{B}^{-1} *_N \mathcal{A}^{-1} \tag{10}$$

$$\det(\mathcal{A} *_N \mathcal{B}) = \det(\mathcal{B} *_N \mathcal{A}) = \det(\mathcal{A}) \cdot \det(\mathcal{B}) \tag{11}$$

(d) For $\mathcal{A}, \mathcal{B} \in \mathbb{C}^{I_1\times\cdots\times I_N}$ of same size and order N,

$$\mathcal{A} *_N \mathcal{B} = \mathcal{B} *_N \mathcal{A} = \mathrm{tr}(\mathcal{A} \circ \mathcal{B}) = \mathrm{tr}(\mathcal{B} \circ \mathcal{A}) \tag{12}$$

(e) For tensors $\mathcal{A} \in \mathbb{C}^{I_1\times\cdots\times I_N\times J_1\times\cdots\times J_M}$ and $\mathcal{B} \in \mathbb{C}^{J_1\times\cdots\times J_M\times I_1\times\cdots\times I_N}$, we have:

$$\det(\mathcal{I}_N + \mathcal{A} *_M \mathcal{B}) = \det(\mathcal{I}_M + \mathcal{B} *_N \mathcal{A}) \tag{13}$$

To prove (13), we can use Lemma 1 and Sylvester's matrix determinant identity [58].

Tensor Gradient: Consider a real-valued scalar function $g(\mathrm{X})$ of a complex matrix X. The corresponding complex gradient matrix is defined as $\nabla_{\mathrm{X}} g \triangleq \partial g/\partial \mathrm{X}^*$, where $[\nabla_{\mathrm{X}} g]_{i,j} = \dfrac{\partial g}{\partial \mathrm{X}^*_{i,j}} \triangleq \dfrac{1}{2}\Big(\dfrac{\partial g}{\partial \mathrm{Re}\{\mathrm{X}_{i,j}\}} + j\dfrac{\partial g}{\partial \mathrm{Im}\{\mathrm{X}_{i,j}\}}\Big)$ [59]. We similarly define the complex gradient of a scalar function $f(\mathcal{X})$ of a tensor $\mathcal{X} \in \mathbb{C}^{I_1\times\cdots\times I_N}$ as $\nabla_{\mathcal{X}} f \triangleq \partial f/\partial \mathcal{X}^*$ where the gradient is a tensor of the same size as $\mathcal{X}$ whose individual components are the derivatives with respect to the components of $\mathcal{X}^*$, i.e. $[\nabla_{\mathcal{X}} f]_{i_1,\ldots,i_N} = \partial f/\partial \mathcal{X}^*_{i_1,\ldots,i_N}$. Using this definition, we can extend several results on matrix complex gradients from [59], [60], [61, Appendix A.7] to a tensor setting. For instance, given $\mathcal{A} \in \mathbb{C}^{J_1\times\cdots\times J_M\times I_1\times\cdots\times I_N}$ and Hermitian positive semi-definite tensors $\mathcal{B}, \mathcal{C} \in \mathbb{C}^{I_1\times\cdots\times I_N\times I_1\times\cdots\times I_N}$, using results from [59], [60] and Lemma 1, we can write:

$$\begin{aligned}&\nabla_{\mathcal{C}} \log[\det(\mathcal{I}_M + \mathcal{A} *_N \mathcal{C} *_N \mathcal{A}^H)]\\ &\quad = \mathcal{A}^H *_M (\mathcal{A} *_N \mathcal{C} *_N \mathcal{A}^H + \mathcal{I}_M)^{-1} *_M \mathcal{A},\end{aligned} \tag{14}$$

$$\nabla_{\mathcal{C}} \mathrm{tr}(\mathcal{B} *_N \mathcal{C}) = \mathcal{B}. \tag{15}$$

III. DISCRETE TENSOR SYSTEM MODEL

A. THE TENSOR CHANNEL

We define the input and output in a multi-domain communication system as tensor symbols of order N and M respectively. Let $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{I_1\times\cdots\times I_N}$ represent the input (transmitted) tensor symbol with $I_1, I_2, \ldots, I_N$ as the dimensions of its N domains where each component $\boldsymbol{\mathcal{X}}_{i_1,\ldots,i_N}$ is a discrete complex symbol. Similarly, we represent the output (received) tensor symbol by $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{J_1\times\cdots\times J_M}$ with $J_1, J_2, \ldots, J_M$ as the dimensions of its M domains. We define a multi-linear channel between the transmit and the receive side as a tensor of order $M + N$ represented by $\mathcal{H} \in \mathbb{C}^{J_1\times\cdots\times J_M\times I_1\times\cdots\times I_N}$. With additive noise, the system model can be specified by using the Einstein product of the channel tensor with the input tensor where we contract along all the modes of the input tensor:

$$\boldsymbol{\mathcal{Y}} = \mathcal{H} *_N \boldsymbol{\mathcal{X}} + \boldsymbol{\mathcal{N}} \tag{16}$$

where

$$\begin{aligned}\boldsymbol{\mathcal{Y}}_{j_1,\ldots,j_M} = &\sum_{i_1=1}^{I_1}\sum_{i_2=1}^{I_2}\cdots\sum_{i_N=1}^{I_N} \mathcal{H}_{j_1,\ldots,j_M,i_1,\ldots,i_N}\boldsymbol{\mathcal{X}}_{i_1,\ldots,i_N}\\ &+ \boldsymbol{\mathcal{N}}_{j_1,\ldots,j_M}\end{aligned} \tag{17}$$

with $\boldsymbol{\mathcal{N}}$ representing the received noise tensor of same size as $\boldsymbol{\mathcal{Y}}$. It is straightforward to see that the narrowband discrete time MIMO matrix channel model is a special case of this tensor model where input and output are order-1 tensors (vectors), the channel is an order-2 tensor (matrix) and the Einstein product reduces to regular matrix multiplication, $\underline{\mathbf{y}} = \mathrm{H}\underline{\mathbf{x}} + \underline{\mathbf{n}}$.

In case of a matrix representation of a discrete MIMO channel which characterizes only the space domain, each component of the channel matrix $\mathrm{H}_{i,j}$ represents the complex

gain (i.e. amounting for both phase change and amplitude gain) of different paths between transmit and receive antennas. In (17), each component of the tensor channel is a complex gain that couples the elements between the order N input and order M output tensors. The proposed system model is generic and the number of modes along with the physical interpretation of the individual modes is system-specific. Such modes can represent space, time, frequency, propagation delay, users, spreading sequence, etc., depending on the system.

For a MIMO communication system with N_R receive and N_T transmit antennas, the continuous time input-output relation in a linear time varying channel is given as [62]:

$$\underline{\mathbf{y}}(t) = \int \mathrm{H}(t,\tau)\underline{\mathbf{x}}(t-\tau)d\tau + \underline{\mathbf{n}}(t) \tag{18}$$

where $\underline{\mathbf{x}}(t)$ is the $N_T \times 1$ continuous time input vector, $\underline{\mathbf{n}}(t)$ and $\underline{\mathbf{y}}(t)$ are the $N_R \times 1$ noise and received vectors respectively. The $N_R \times N_T$ matrix $\mathrm{H}(t,\tau)$ has components $\mathrm{H}_{n_r,n_t}(t,\tau)$ that represent the channel impulse response between transmit antenna n_t and receive antenna n_r at time instant t and delay τ. If the channel is assumed time-invariant with a maximum delay τ_{max}, the discretization of (18) at a sampling frequency f_s gives the input/output relation at an instant k as [62]:

$$\underline{\mathbf{y}}[k] = \sum_{d=1}^{D} \mathrm{H}[d]\underline{\mathbf{x}}[k-(d-1)] + \underline{\mathbf{n}}[k], \quad k = 0, 1, \ldots, (N-1) \tag{19}$$

where $\underline{\mathbf{x}}[k]$ is the $N_T \times 1$ channel input at time index k, $\underline{\mathbf{n}}[k]$ and $\underline{\mathbf{y}}[k]$ are the $N_R \times 1$ noise and received vectors respectively. The $N_R \times N_T$ matrix $\mathrm{H}[d]$ has components $\mathrm{H}_{n_r,n_t}[d]$ which represents the length D channel impulse response between transmit antenna n_t and receive antenna n_r at delay d, where $D = \lceil f_s \tau_{max} \rceil$. Delay can be considered as another domain in the system model. So for a time-invariant channel, at any time instant k the relation in (19) can be expressed using tensor model as [63]:

$$\underline{\mathbf{y}}[k] = \mathcal{H} *_2 \mathbf{X}[k] + \underline{\mathbf{n}}[k] \tag{20}$$

where all the individual vectors $\underline{\mathbf{x}}[k], \underline{\mathbf{x}}[k-1], \ldots, \underline{\mathbf{x}}[k-(D-1)]$ from (19) form the columns of the matrix $\mathbf{X}[k]$ of size $N_T \times D$ and all the individual matrices $\mathrm{H}[1], \ldots, \mathrm{H}[D]$ from (19) are stacked together where they form the slices of the order-3 channel tensor $\mathcal{H}$ of size $N_R \times N_T \times D$. For $D = 1$, $\mathcal{H}$ reduces to a matrix and $\mathbf{X}[k]$ to a vector in (20). As D increases, the tensor framework allows capturing the time dispersion in the system model by increasing the dimension of the third domain in the channel and the input. Now assuming channel is time variant, the discretized input/output relation can be given as [23]:

$$\underline{\mathbf{y}}[k] = \sum_{d=1}^{D} \mathrm{H}[k,d]\underline{\mathbf{x}}[k-(d-1)] + \underline{\mathbf{n}}[k], \quad k = 0, 1, \ldots, (N-1) \tag{21}$$

where each element $\mathrm{H}_{n_r,n_t}[k,d]$ represents complex channel gain between n_tth transmit and n_rth receive antenna for delay d at time instant k. In [23], assuming a cyclic prefix addition to each input block, (21) is expressed in matrix notation as $\underline{\mathbf{y}}' = \mathrm{H}'\underline{\mathbf{x}}' + \underline{\mathbf{n}}'$ over a time block of N symbol durations by appending vectors $\underline{\mathbf{y}}[k]$, $\underline{\mathbf{x}}[k-(d-1)]$ and $\underline{\mathbf{n}}[k]$ for different k into vectors $\underline{\mathbf{y}}'$, $\underline{\mathbf{x}}'$ and $\underline{\mathbf{n}}'$ of size $N \cdot N_R$, $N \cdot N_T$ and $N \cdot N_R$ respectively, and the channel matrix $\mathrm{H}[k,d]$ into a larger matrix H' of size $N \cdot N_R \times N \cdot N_T$. However, appending the vectors implies making the two distinct domains indistinguishable in the system formulation. Hence a more obvious and intuitive way to represent such a system would be using tensors where the channel can be expressed as $N_R \times N \times N_T \times D'$ tensor where $D' = N + D - 1$. We do not assume any cyclic prefix addition here. Since output at index k i.e. $\underline{\mathbf{y}}[k]$ will depend on inputs $\underline{\mathbf{x}}[k], \underline{\mathbf{x}}[k-1], \ldots, \underline{\mathbf{x}}[k-(D-1)]$, so corresponding to output being indexed by N time indices $0, 1, \ldots, (N-1)$, the input will be indexed by $N+D-1$ time indices $-(D-1), \ldots, (N-1)$. So in the system model (20), we include a domain of length N time slots in the channel tensor and the output tensor to account for time variation and increase the delay domain length to $N + D - 1$ in the channel tensor and the input tensor. Thereby, our system model becomes:

$$\mathbf{Y} = \mathcal{H} *_2 \mathbf{X} + \mathbf{N} \tag{22}$$

where all the individual vectors $\underline{\mathbf{y}}[k]$ from (21) for $k = 0, \ldots, (N-1)$ form the columns of the matrix $\mathbf{Y}$ of size $N_R \times N$. Similarly vectors $\underline{\mathbf{n}}[k]$ from (21) form the columns of the matrix $\mathbf{N}$ of size $N_R \times N$ and vectors $\underline{\mathbf{x}}[d']$ where $d' = k - (d-1)$ form the columns of the matrix $\mathbf{X}$ of size $N_T \times D'$. All the individual matrices $\mathrm{H}[k,d]$ from (21) are now sub-tensors of the order-4 channel tensor $\mathcal{H}$ of size $N_R \times N \times N_T \times D'$ where $\mathcal{H}_{:,k,:,d} = \mathrm{H}[k,d]$. We can see how the tensor system model in (20) simply evolved in (22) to account for time variation of the channel as well.

B. TENSOR MODEL APPLIED TO MU-MIMO OFDM

The model presented in (16) can be used for a wide variety of systems. For instance, the system model used in [44] for image restoration is a specific case of the system model of (16). To stress the relevance of the proposed tensor model, particularly in multi-domain communication systems, we now present an example of a multi-user MIMO OFDM system which can be represented using the tensor framework.

OFDM is one of the most popular multi-carrier schemes and has been used extensively with MIMO in 4G standards and Wi-Fi [64]. A conventional model for a MIMO OFDM system in the frequency domain is given by [23]:

$$\check{\underline{\mathbf{y}}}[p] = \check{\mathrm{H}}[p,p]\check{\underline{\mathbf{x}}}[p] + \sum_{q=0, q\neq p}^{N_{sc}-1} \check{\mathrm{H}}[p,q]\check{\underline{\mathbf{x}}}[q] + \check{\underline{\mathbf{n}}}[p] \tag{23}$$

where $\check{\underline{\mathbf{y}}}[p]$, $\check{\underline{\mathbf{x}}}[p]$ and $\check{\underline{\mathbf{n}}}[p]$ are the frequency domain received, transmitted and noise symbol vectors at sub-carrier p, while

N_{sc} denotes the number of sub-carriers. The model represented by (23) can be obtained from (21) by taking the $N_{sc}-$point discrete Fourier Transform (DFT) of $\{\underline{\mathbf{y}}[k]\}$ where $DFT\{\underline{\mathbf{y}}[k]\} = \underline{\check{\mathbf{y}}}[p]$, $DFT\{\underline{\mathbf{n}}[k]\} = \underline{\check{\mathbf{n}}}[p]$ and $DFT\{\underline{\mathbf{x}}[k]\} = \underline{\check{\mathbf{x}}}[p]$. The frequency domain $N_R \times N_T$ channel matrix between transmit sub-carrier q and receive sub-carrier p is $\check{\mathrm{H}}[p, q]$, whose individual elements $\check{\mathrm{H}}_{n_r,n_t}[p, q]$ can be obtained from the discrete time varying channel between the n_rth receive antenna and n_tth transmit antenna, $\mathrm{H}_{n_r,n_t}[k, d]$ (based on DFT of (21)) as:

$$\check{\mathrm{H}}_{n_r,n_t}[p, q] = \frac{1}{N_{sc}} \sum_{k=0}^{N_{sc}-1} \sum_{d=0}^{D-1} \mathrm{H}_{n_r,n_t}[k, d] e^{j2\pi k(q-p)/N_{sc}} e^{-j2\pi qd/N_{sc}} \tag{24}$$

where $1 \leq n_r \leq N_R$, $1 \leq n_t \leq N_T$ and $0 \leq p, q \leq N_{sc}-1$. Using tensors, we can represent the frequency domain input/output relation in MIMO OFDM of (23) as:

$$\check{\mathbf{Y}} = \check{\mathcal{H}} *_2 \check{\mathbf{X}} + \check{\mathbf{N}} \tag{25}$$

where each vector $\underline{\check{\mathbf{y}}}[p]$ and $\underline{\check{\mathbf{n}}}[p]$ from (23) for $p = 0, \ldots, N_{sc}-1$ form the columns of matrices $\check{\mathbf{Y}}$ and $\check{\mathbf{N}}$ of size $N_R \times N_{sc}$, and vectors $\underline{\check{\mathbf{x}}}[p]$ form the columns of matrix $\check{\mathbf{X}} \in \mathbb{C}^{N_T \times N_{sc}}$. The input and output are connected by an order-4 tensor channel $\check{\mathcal{H}} \in \mathbb{C}^{N_R \times N_{sc} \times N_T \times N_{sc}}$ where each element $\check{\mathrm{H}}_{n_r,n_t}[p, q]$ from (24) is now an element in the channel tensor as $\check{\mathcal{H}}_{n_r,p,n_t,q}$. We can expand the MIMO OFDM system model to include users as an additional domain in the model which will lead to a sixth order tensor channel. In the case of multi-user MIMO OFDM, the frequency domain channel matrix is often represented as an $N_R \times N_T$ matrix corresponding to a specific user and a specific sub-carrier [65], [66]. To account for inter-carrier interference (ICI) as well, the channel matrix could be represented as $\check{\mathrm{H}}[u, p, q] \in \mathbb{C}^{N_R \times N_T}$ corresponding to the uth user for transmit sub-carrier q and receive sub-carrier p. Consider a multi-user MIMO OFDM downlink system where a base station is catering to U users having N_R receive antennas each. The system model is a generalization of (23) and it is given by

$$\underline{\check{\mathbf{y}}}[u, p] = \check{\mathrm{H}}[u, p, p]\underline{\check{\mathbf{x}}}[p] + \underbrace{\sum_{q=0, q\neq p}^{N_{sc}-1} \check{\mathrm{H}}[u, p, q]\underline{\check{\mathbf{x}}}[q]}_{\text{ICI for } u\text{th user}} + \underline{\check{\mathbf{n}}}[u, p] \tag{26}$$

for $p, q = 0, \ldots, N_{sc}-1$ and $u = 1, \ldots, U$. The entities $\underline{\check{\mathbf{y}}}[u, p] \in \mathbb{C}^{N_R \times 1}$ and $\underline{\check{\mathbf{n}}}[u, p] \in \mathbb{C}^{N_R \times 1}$ represent the received signal vector and noise vector on sub-carrier p for the uth user, respectively. Also, $\underline{\check{\mathbf{x}}}[q] \in \mathbb{C}^{N_T \times 1}$ denotes the transmit vector from the base station at sub-carrier q, which is given by [67]:

$$\underline{\check{\mathbf{x}}}[q] = \sum_{u'=1}^{U} \mathrm{M}[u', q]\underline{\check{\mathbf{d}}}[u', q] \tag{27}$$

where $\mathrm{M}[u', q] \in \mathbb{C}^{N_T \times N_T}$ denotes the precoding matrix used to transmit data vector $\underline{\check{\mathbf{d}}}[u', q] \in \mathbb{C}^{N_T \times 1}$ to user u' on sub-carrier q. Hence the system model of (26) becomes

$$\underline{\check{\mathbf{y}}}[u, p] = \sum_{q=0}^{N_{sc}-1} \check{\mathrm{H}}[u, p, q]\Big(\sum_{u'=1}^{U} \mathrm{M}[u', q]\underline{\check{\mathbf{d}}}[u', q]\Big) + \underline{\check{\mathbf{n}}}[u, p] \tag{28}$$

Let $\check{\mathrm{H}}[u, p, q] \cdot \mathrm{M}[u', q] = \mathrm{G}[u, u', p, q] \in \mathbb{C}^{N_R \times N_T}$ denote the equivalent channel between transmit data vector $\underline{\check{\mathbf{d}}}[u', q]$ and receive vector $\underline{\check{\mathbf{y}}}[u, p]$, then the input/output relation can be written as:

$$\underline{\check{\mathbf{y}}}[u, p] = \sum_{q=0}^{N_{sc}-1} \sum_{u'=1}^{U} \mathrm{G}[u, u', p, q]\underline{\check{\mathbf{d}}}[u', q] + \underline{\check{\mathbf{n}}}[u, p]. \tag{29}$$

In this case, the output and noise tensors can be rearranged into order-3 tensors $\check{\mathbf{\mathcal{Y}}}, \check{\mathbf{\mathcal{N}}} \in \mathbb{C}^{U \times N_R \times N_{sc}}$. The components $\underline{\check{\mathbf{y}}}_{n_r}[u, p]$ and $\underline{\check{\mathbf{n}}}_{n_r}[u, p]$ are mapped to elements of third order tensors, denoted by $\check{\mathbf{\mathcal{Y}}}_{u,n_r,p}$ and $\check{\mathbf{\mathcal{N}}}_{u,n_r,p}$ respectively. Similarly, the input can be rearranged as an order-3 tensor $\check{\mathbf{\mathcal{D}}} \in \mathbb{C}^{U \times N_T \times N_{sc}}$ where $\underline{\check{\mathbf{d}}}_{n_t}[u', q]$ is mapped to $\check{\mathbf{\mathcal{D}}}_{u',n_t,q}$. Subsequently, the channel can be represented as an order-6 tensor $\check{\mathcal{G}} \in \mathbb{C}^{U \times N_R \times N_{sc} \times U \times N_T \times N_{sc}}$ where each element of matrix $\mathrm{G}_{n_r,n_t}[u, u', p, q]$ from (28) is mapped to an element $\check{\mathcal{G}}_{u,n_r,p,u',n_t,q}$ of the sixth order tensor channel. The system model then becomes:

$$\check{\mathbf{\mathcal{Y}}} = \check{\mathcal{G}} *_3 \check{\mathbf{\mathcal{D}}} + \check{\mathbf{\mathcal{N}}} \tag{30}$$

The tensor model represented by (30) can be considered as an evolution of the common matrix MIMO model in the space domain only, to a tensor MIMO model that in addition to space it encapsulates also the frequency and user domains. Similarly, other systems such as MIMO DSL [68], MIMO CDMA [69], MIMO FBMC [26] and MIMO GFDM [24] can be represented using the tensor based system model.

Representing (29) using (30), or (23) using (25) gives us a unified framework which encapsulates the multitude of signalling domains into well structured tensor entities. The matrix representations from (29) and (23) often leads to a per user or per sub-carrier processing of the signals, where the inter-domain interferences are treated in combination with the additive noise term. On the contrary, the tensor model from (30) and (25) can be employed to design transceivers jointly across domains such as users, sub-carriers, and antennas, using tools from tensor algebra. We present several numerical examples in sections VI-B and VII illustrating the advantages of the joint domain processing for multi-domain communication systems. Note that an alternate representation of (29) can also be achieved via a large matrix channel based model, where the input and output signals are concatenated across users, sub-carrier and antenna domains to form vectors. However, such representation will obscure the indices u', n_t, q and u, n_r, p into single indices with no well defined physical meaning. Such a lower order representation of high order entities makes it difficult to incorporate domain specific

constraints such as those explained in section V-A which are well handled using tensors. As a prelude to the capacity of tensor channels, we first present some basic information theoretic notions for tensors in the next section.

IV. INFORMATION THEORETIC NOTIONS FOR TENSORS

A tensor is said to be *random* if its components are random variables. Expectation of a random tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ denoted by $\bar{\mathcal{M}} = \mathbb{E}[\mathcal{X}] \in \mathbb{C}^{I_1 \times \cdots \times I_N}$, is a tensor with each component consisting of the expected value of the corresponding component of $\mathcal{X}$.

COVARIANCE AND PSEUDO-COVARIANCE

Covariance of a random tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ can be represented using a tensor $\mathcal{Q} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ defined as $\mathcal{Q} = \mathbb{E}[(\mathcal{X} - \bar{\mathcal{M}}) \circ (\mathcal{X} - \bar{\mathcal{M}})^*]$. It can be shown that a covariance tensor is Hermitian, i.e. $\mathcal{Q} = \mathcal{Q}^H$. The pseudo-covariance of $\mathcal{X}$ is also an order $2N$ tensor defined as $\mathbb{E}[(\mathcal{X} - \bar{\mathcal{M}}) \circ (\mathcal{X} - \bar{\mathcal{M}})]$. A random tensor is called *proper* if its pseudo-covariance is an all zero tensor [14].

TENSOR GAUSSIAN DISTRIBUTION

The probability density function (pdf) of a random tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ is a scalar function of all its elements. Hence the pdf of a tensor having proper complex Gaussian entries can be specified by vectorizing the tensor and using the pdf of a proper complex Gaussian vector. However, using the properties of the Einstein product, the pdf can also be expressed without vectorizing as explained in [14]. The pdf of a proper complex Gaussian distributed tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ is given as:

$$p_{\mathcal{X}}(\mathcal{X}) = \frac{\exp\left\{ -(\mathcal{X} - \bar{\mathcal{M}})^* *_N \mathcal{Q}^{-1} *_N (\mathcal{X} - \bar{\mathcal{M}}) \right\}}{(\pi)^{I_1 I_2 \ldots I_N} \det(\mathcal{Q})} \tag{31}$$

where $\bar{\mathcal{M}} = \mathbb{E}[\mathcal{X}]$ is the order-N mean tensor and $\mathcal{Q} = \mathbb{E}[(\mathcal{X} - \bar{\mathcal{M}}) \circ (\mathcal{X} - \bar{\mathcal{M}})^*]$ is the order-$2N$ covariance tensor.

A. DIFFERENTIAL ENTROPY OF A CIRCULAR COMPLEX GAUSSIAN DISTRIBUTED TENSOR

A complex random tensor $\mathcal{X}$ is defined as *circular* if it is rotationally invariant, i.e. $\mathcal{X}$ and $\mathcal{Y} = e^{j\alpha}\mathcal{X}$ have the same probability distribution for any given real α. A complex Gaussian random vector is circular if and only if it is zero mean and proper [70]. This statement can be extended to tensor case also:

Lemma 2: A complex Gaussian random tensor is circular if and only if it is zero mean and proper.

The proof of Lemma 2 directly follows from the definition of proper and circular tensors. The distribution of a circularly symmetric complex Gaussian tensor $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ with covariance $\mathcal{Q}$ is given by (31), with $\bar{\mathcal{M}} = 0$. Subsequently, the differential entropy of such a tensor is given by:

$$\begin{aligned}
\mathcal{H}(\mathcal{X}) &= \mathbb{E}\big[-\log p_{\mathcal{X}}(\mathcal{X}) \big] \\
&= \mathbb{E}\left[-\log \left\{ \frac{\exp\left(-\mathcal{X}^* *_N \mathcal{Q}^{-1} *_N \mathcal{X} \right)}{(\pi)^{I_1 I_2 \ldots I_N} \det(\mathcal{Q})} \right\} \right] \\
&= \log\Big((\pi)^{I_1 \ldots I_N} \det(\mathcal{Q}) \Big) \\
&\quad + (\log e) \mathbb{E}\Big[\big((\mathcal{X}^* *_N \mathcal{Q}^{-1}) *_N \mathcal{X} \big) \Big] \\
&= \log\Big((\pi)^{I_1 \ldots I_N} \det(\mathcal{Q}) \Big) \\
&\quad + (\log e) \mathbb{E}\Big[\operatorname{tr}\big(\mathcal{X} \circ (\mathcal{X}^* *_N \mathcal{Q}^{-1}) \big) \Big] \text{(from (12))} \\
&= \log\Big((\pi)^{I_1 \ldots I_N} \det(\mathcal{Q}) \Big) \\
&\quad + (\log e) \operatorname{tr}\Big(\underbrace{\mathbb{E}[\mathcal{X} \circ \mathcal{X}^*]}_{\text{identity tensor}} *_N \mathcal{Q}^{-1} \Big) \\
&\qquad \text{(from associativity rule, (8))} \\
&= \log\Big((e\pi)^{I_1 \ldots I_N} \det(\mathcal{Q}) \Big)
\end{aligned} \tag{32}$$

The following lemma regarding the differential entropy of a circularly symmetric complex Gaussian random tensor is proven in [45]:

Lemma 3: Let $\mathcal{X} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ be a circularly symmetric complex Gaussian random tensor with covariance tensor $\mathcal{Q}$. Let $\mathcal{Y} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ be another zero-mean random tensor with same covariance tensor. Then, $\mathcal{H}(\mathcal{X}) \geq \mathcal{H}(\mathcal{Y})$, i.e. for a given covariance tensor, a circularly symmetric Gaussian distribution is the entropy maximizer.

B. MUTUAL INFORMATION

For the system model defined in (16), the output covariance tensor is

$$\begin{aligned}
\mathcal{Q}_{\mathcal{Y}} &= \operatorname{cov}(\mathcal{Y}) = \mathbb{E}[\mathcal{Y} \circ \mathcal{Y}^*] \\
&= \mathbb{E}\big[(\mathcal{H} *_N \mathcal{X} + \mathcal{N}) \circ (\mathcal{H}^* *_N \mathcal{X}^* + \mathcal{N}^*) \big] \qquad (33) \\
\mathcal{Q}_{\mathcal{Y}} &= \underbrace{\mathbb{E}\big[(\mathcal{H} *_N \mathcal{X}) \circ (\mathcal{H}^* *_N \mathcal{X}^*) \big]}_{\text{Main term}} + \underbrace{\mathbb{E}\big[\mathcal{N} \circ \mathcal{N}^* \big]}_{\text{Noise Covariance}} \\
&\quad + \underbrace{\mathbb{E}\big[(\mathcal{H} *_N \mathcal{X}) \circ \mathcal{N}^* \big] + \mathbb{E}\big[\mathcal{N} \circ (\mathcal{H}^* *_N \mathcal{X}^*) \big]}_{\text{cross terms}} \qquad (34)
\end{aligned}$$

Assuming $\mathcal{X}$ and $\mathcal{N}$ are zero mean and independent, it can be shown that the cross terms are zero. Based on commutativity rule (9), we get $\mathcal{H}^* *_N \mathcal{X}^* = \mathcal{X}^* *_N (\mathcal{H}^*)^T = \mathcal{X}^* *_N \mathcal{H}^H$. Using the associativity rule (8), we get:

$$\begin{aligned}
\mathcal{Q}_{\mathcal{Y}} &= \mathbb{E}\big[(\mathcal{H} *_N \mathcal{X}) \circ (\mathcal{X}^* *_N \mathcal{H}^H) \big] + \mathcal{Q}_{\mathcal{N}} \\
&= \big(\mathcal{H} *_N \mathbb{E}[\mathcal{X} \circ \mathcal{X}^*] *_N \mathcal{H}^H \big) + \mathcal{Q}_{\mathcal{N}} \\
&= \mathcal{H} *_N \mathcal{Q}_{\mathcal{X}} *_N \mathcal{H}^H + \mathcal{Q}_{\mathcal{N}}
\end{aligned} \tag{35}$$

where $\mathcal{Q}_{\mathcal{X}}$ and $\mathcal{Q}_{\mathcal{N}}$ are the input and noise covariance tensors respectively. The following two lemmas can be proven using the definition of circularly symmetric complex Gaussian tensors:

Lemma 4: If $\mathcal{X} \in \mathbb{C}^{I_1 \times I_2 \times \cdots \times I_N}$ is a circularly symmetric complex Gaussian tensor, then so is $\mathcal{Y} = \mathcal{H} *_N \mathcal{X}$ for any deterministic tensor, $\mathcal{H} \in \mathbb{C}^{K_1 \times K_2 \times \cdots \times K_P \times I_1 \times I_2 \times \cdots \times I_N}$.

Lemma 5: If $\mathcal{X}$ and $\mathcal{Y}$ are independent circularly symmetric complex Gaussian tensors of same order and size, then $\mathcal{Z} = \mathcal{X} + \mathcal{Y}$ is also circularly symmetric complex Gaussian.

Assuming noise to be a zero-mean Gaussian distributed tensor with covariance $\mathcal{Q}_{\mathcal{N}}$ that is independent of the input tensor $\mathcal{X}$, we can write the mutual information between input and output tensors as:

$$\begin{aligned}\mathcal{I}(\mathcal{X};\mathcal{Y}) &= \mathcal{H}(\mathcal{Y}) - \mathcal{H}(\mathcal{Y}|\mathcal{X}) = \mathcal{H}(\mathcal{Y}) - \mathcal{H}(\mathcal{N}) \\ &= \mathcal{H}(\mathcal{Y}) - \log\left((e\pi)^{J_1 J_2 \cdots J_M} \det(\mathcal{Q}_{\mathcal{N}})\right) \end{aligned} \tag{36}$$

Based on Lemma 3 and received covariance derived in (35), we can write:

$$\mathcal{H}(\mathcal{Y}) \leq \log\left((e\pi)^{J_1 J_2 \cdots J_M} \det(\mathcal{H} *_N \mathcal{Q}_{\mathcal{X}} *_N \mathcal{H}^H + \mathcal{Q}_{\mathcal{N}})\right) \tag{37}$$

$$\Rightarrow \mathcal{I}(\mathcal{X};\mathcal{Y}) \leq \log\left[\frac{\det\left(\mathcal{H} *_N \mathcal{Q}_{\mathcal{X}} *_N \mathcal{H}^H + \mathcal{Q}_{\mathcal{N}}\right)}{\det(\mathcal{Q}_{\mathcal{N}})}\right] \tag{38}$$

where equality is achieved only if $\mathcal{Y}$ is Gaussian.

V. CAPACITY OF A FIXED TENSOR CHANNEL

Finding the Shannon capacity of the tensor channel requires the maximization of the mutual information between the input and the output tensors over input distributions under possible constraints. In this work, we assume that the channel tensor is known and the noise tensor is zero-mean Gaussian distributed having independent components with variance σ^2, and hence the noise covariance tensor is given by $\mathcal{Q}_{\mathcal{N}} = \sigma^2 \mathcal{I}_M$. For simplicity we assume $\sigma^2 = 1$. Now let us consider the mutual information inequality of (38) where equality is achieved only if $\mathcal{Y}$ is Gaussian. In our tensor channel model (16), since $\mathcal{N}$ is zero-mean circularly symmetric complex Gaussian, then using Lemma 4 and 5 implies that $\mathcal{Y}$ will also be zero-mean circularly symmetric complex Gaussian if $\mathcal{X}$ is so. Hence for the purpose of maximizing mutual information, we take $\mathcal{X}$ as zero mean circularly symmetric complex Gaussian with covariance $\mathcal{Q}_{\mathcal{X}} = \mathcal{Q}$. Thus with noise covariance tensor as identity tensor $\mathcal{I}_M$, we get

$$\mathcal{I}(\mathcal{X};\mathcal{Y}) = \log\left[\det\left(\mathcal{H} *_N \mathcal{Q} *_N \mathcal{H}^H + \mathcal{I}_M\right)\right] \tag{39}$$

and the capacity is given by,

$$\begin{aligned} &\max_{\mathcal{Q}} \left(\log\left[\det\left(\mathcal{H} *_N \mathcal{Q} *_N \mathcal{H}^H + \mathcal{I}_M\right)\right]\right) \\ &s.t. \quad f(\mathcal{Q}) \leq 0, \quad \mathcal{Q} \succeq 0. \end{aligned} \tag{40}$$

where the inequality constraint $f(\mathcal{Q}) \leq 0$ can represent a family of power constraints.

A. FAMILY OF POWER CONSTRAINTS

In a practical system, the transmit power constraints can span multiple domains. For instance, in a transmission scheme employing the space, time and frequency domains, instead of imposing power across the tensor symbol, the power constraint might be on each antenna, or each antenna and time slot, or each antenna, time slot and frequency bin. In our framework, we have the flexibility of mathematically representing any such power constraint.

First we introduce some notations for a simplification. We denote the sequence of indices $(i_1, i_2, \ldots, i_N)$ as $\underline{i}$. Let $\underline{i}_c$ denotes the sequence of indices indicating tensor symbol elements under power constraint and let $\underline{i}_r$ represents the rest of the indices in $\underline{i}$ which are not in $\underline{i}_c$. For example, in an order-5 tensor of size $I_1 \times I_2 \times I_3 \times I_4 \times I_5$, we have $\underline{i} = (i_1, i_2, i_3, i_4, i_5)$. Let the domains which are under individual power constraints be 2 and 3, then $\underline{i}_c = (i_2, i_3)$ and $\underline{i}_r = (i_1, i_4, i_5)$. With this choice of notations, we will write $\sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3}$ as simply $\sum_{\underline{i}_c}$. Notations corresponding to cases when either $\underline{i}_c$ or $\underline{i}_r$ is empty are explained in Table 1.

TABLE 1. Simplified notation for indices.

Cases	$\sum_{\underline{i}_r} \mathcal{Q}_{\underline{i},\underline{i}}$ denotes	$P_{\underline{i}_c}$ denotes	Interpretation
$\underline{i}_c$ is empty and $\underline{i}_r = \underline{i}$	$\sum_{\underline{i}} \mathcal{Q}_{\underline{i},\underline{i}} = \mathrm{tr}(\mathcal{Q})$	P	sum power constraint
$\underline{i}_r$ is empty and $\underline{i}_c = \underline{i}$	$\mathcal{Q}_{i_1,\ldots,i_N,i_1,\ldots,i_N}$	$P_{i_1,\ldots,i_N}$	per element constraints

Using these simplified notations, we will now describe a family of optimization problems to find capacity which can cover different types of power constraints, as follows:

$$\max_{\mathcal{Q}} \left(\log\left[\det\left(\mathcal{H} *_N \mathcal{Q} *_N \mathcal{H}^H + \mathcal{I}_M\right)\right]\right) \tag{41}$$

$$s.t. \sum_{\underline{i}_r} \mathcal{Q}_{\underline{i},\underline{i}} \leq P_{\underline{i}_c} \quad \forall \underline{i}_c, \tag{42}$$

$$\mathcal{Q} \succeq 0. \tag{43}$$

To understand how the above framework represents a large variety of constraints, let us consider a few specific cases. The case $\underline{i}_c = \underline{i}$, hence $\underline{i}_r$ is empty, will represent the situation where we have per element power constraints with $P_{\underline{i}_c} = P_{i_1,\ldots,i_N}$ and (42) becomes:

$$\mathcal{Q}_{i_1,\ldots,i_N,i_1,\ldots,i_N} \leq P_{i_1,\ldots,i_N}, \quad \forall i_1, i_2, \ldots, i_N \tag{44}$$

When $\underline{i}_c = i_K$, where $K \leq N$, we have the case with per domain element constraint for Kth domain where each element i_K has a different budget of P_{i_K}, i.e. $P_{\underline{i}_c} = P_{i_K}$ and (42) becomes:

$$\sum_{i_1=1}^{I_1} \cdots \sum_{i_{K-1}=1}^{I_{K-1}} \sum_{i_{K+1}=1}^{I_{K+1}} \cdots \sum_{i_N=1}^{I_N} \mathcal{Q}_{i_1,\ldots,i_K,\ldots,i_N,i_1,\ldots,i_K,\ldots,i_N} \leq P_{i_K}, \quad \forall i_K \tag{45}$$

Now let's assume we have power constraints on two domains K and L such that $K < L \leq N$. Then $\underline{i}_c = (i_K, i_L)$ and $\underline{i}_r = (i_1, \ldots, i_{K-1}, i_{K+1}, \ldots i_{L-1}, i_{L+1}, \ldots, i_N)$. In this case $P_{\underline{i}_c} = P_{i_K,i_L}$ and (42) becomes:

$$\sum_{i_1=1}^{I_1} \cdots \sum_{i_{K-1}=1}^{I_{K-1}} \sum_{i_{K+1}=1}^{I_{K+1}} \cdots \sum_{i_{L-1}=1}^{I_{L-1}} \sum_{i_{L+1}=1}^{I_{L+1}} \cdots \sum_{i_N=1}^{I_N} \mathcal{Q}_{i_1,\ldots,i_K,\ldots,i_L,\ldots,i_N,i_1,\ldots,i_K,\ldots,i_L,\ldots,i_N} \leq P_{i_K,i_L}, \quad \forall (i_K, i_L) \tag{46}$$

Similarly, we can represent constraints on any number of domains. Lastly, let's assume that $\underline{i}_c$ is empty, hence $\underline{i}_r = \underline{i}$ and the power constraint translates to the sum power constraint, i.e. $P_{\underline{i}_c} = P$ and we get:

$$\sum_{i_1=1}^{I_1} \cdots \sum_{i_N=1}^{I_N} \mathcal{Q}_{i_1,\ldots,i_N,i_1,\ldots,i_N} = \text{tr}(\mathcal{Q}) \leq P \tag{47}$$

All these power constraints are linear, and the objective function in (41) is concave. Notice that the feasible set for this optimization problem is the set of positive semi-definite tensors satisfying the given power constraints which are linear. Hence the feasible set is convex. Thereby (41), (42) and (43) represent a family of convex optimization problems which can be solved using the Karush-Kuhn-Tucker (KKT) conditions [71]. Furthermore, since $P_{\underline{i}_c}$ are finite and non-negative, a simple choice of covariance tensor belonging to the feasible set could be a pseudo-diagonal tensor with all non-negative entries such that they satisfy the power constraints. So the feasible set is a non-empty convex set, and hence by Slater's condition [71], strong duality holds and the optimal solution always exist. Next we will find the optimal solution using the KKT conditions.

B. SOLUTION USING KKT CONDITIONS

Let $\mathcal{M} \succeq 0$ be the Lagrange multiplier tensor for the positive semi-definite constraint from (43) of size $I_1 \times \ldots I_N \times I_1 \times \ldots I_N$. Let $\mu_{\underline{i}_c} \geq 0$ for all $\underline{i}_c$ be the Lagrange multipliers corresponding to all the linear constraints from (42). Then the Lagrangian functional can be defined as:

$$\begin{aligned}\mathcal{L}(\mathcal{Q}, \{\mu_{\underline{i}_c}\}, \mathcal{M}) = &-\log[\det(\mathcal{H} *_N \mathcal{Q} *_N \mathcal{H}^H + \mathcal{I}_M)] \\ &+ \sum_{\underline{i}_c} \mu_{\underline{i}_c} (\sum_{\underline{i}_r} \mathcal{Q}_{\underline{i},\underline{i}} - P_{\underline{i}_c}) - \text{tr}(\mathcal{M} *_N \mathcal{Q}).\end{aligned} \tag{48}$$

We arrange the values $\{\mu_{\underline{i}_c}\}$ in a pseudo-diagonal tensor $\mathcal{B}$ of same size as the input covariance such that its non-zero entries are $\mathcal{B}_{\underline{i},\underline{i}} = \mu_{\underline{i}_c}, \forall \underline{i}_r$. For instance, if $\underline{i}_c = (i_1, i_2)$, then $\mathcal{B}_{i_1,\ldots,i_N,i_1,\ldots,i_N} = \mu_{i_1,i_2}$ for all $(i_3, \ldots, i_N)$. Then we get

$$\begin{aligned}\sum_{\underline{i}_c} \mu_{\underline{i}_c} \cdot \sum_{\underline{i}_r} \mathcal{Q}_{\underline{i},\underline{i}} &= \sum_{\underline{i}_c} \sum_{\underline{i}_r} \mu_{\underline{i}_c} \cdot \mathcal{Q}_{\underline{i},\underline{i}} \\ &= \sum_{\underline{i}} \mathcal{B}_{\underline{i},\underline{i}} \cdot \mathcal{Q}_{\underline{i},\underline{i}} = \text{tr}(\mathcal{B} *_N \mathcal{Q}).\end{aligned} \tag{49}$$

Based on (49), we can re-write the Lagrangian from (48) as:

$$\begin{aligned}\mathcal{L}(\mathcal{Q}, \{\mu_{\underline{i}_c}\}, \mathcal{M}) = &-\log[\det(\mathcal{H} *_N \mathcal{Q} *_N \mathcal{H}^H + \mathcal{I}_M)] \\ &- \sum_{\underline{i}_c} \mu_{\underline{i}_c} P_{\underline{i}_c} + \text{tr}(\mathcal{B} *_N \mathcal{Q}) - \text{tr}(\mathcal{M} *_N \mathcal{Q}).\end{aligned} \tag{50}$$

Based on (14), (15) and the definition of tensor gradient as presented in section II-B, the gradient of Lagrangian with respect to $\mathcal{Q}$ can be written as:

$$\nabla_{\mathcal{Q}} \mathcal{L} = -\mathcal{H}^H *_M (\mathcal{H} *_N \mathcal{Q} *_N \mathcal{H}^H + \mathcal{I}_M)^{-1} *_M \mathcal{H} + \mathcal{B} - \mathcal{M}. \tag{51}$$

Equating $\nabla_{\mathcal{Q}} \mathcal{L}$ from (51) to $0_{\mathcal{T}}$, we get the first KKT condition as

$$\mathcal{H}^H *_M (\mathcal{H} *_N \mathcal{Q} *_N \mathcal{H}^H + \mathcal{I}_M)^{-1} *_M \mathcal{H} = \mathcal{B} - \mathcal{M}. \tag{52}$$

The KKT equations also include complementary slackness condition corresponding to each constraint and its associated Lagrange multiplier [71]. For the linear constraints in (42), the definition of complementary slackness leads to:

$$\mu_{\underline{i}_c} (\sum_{\underline{i}_r} \mathcal{Q}_{\underline{i},\underline{i}} - P_{\underline{i}_c}) = 0, \quad \forall \underline{i}_c. \tag{53}$$

For the constraint in (43), based on the approach taken for semi-definite programming [71], the complementary slackness can be written as $\text{tr}(\mathcal{M} *_N \mathcal{Q}) = 0$. Since $\mathcal{M}, \mathcal{Q} \succeq 0$, similar to the matrix case as presented in [72], we have $\text{tr}(\mathcal{M} *_N \mathcal{Q}) = 0 \Rightarrow \mathcal{M} *_N \mathcal{X} = 0_{\mathcal{T}}$. Also, since $\text{tr}(\mathcal{M} *_N \mathcal{Q}) = \text{tr}(\mathcal{M} *_N \mathcal{Q}^{1/2} *_N \mathcal{Q}^{1/2}) = \text{tr}(\mathcal{Q}^{1/2} *_N \mathcal{M} *_N \mathcal{Q}^{1/2})$, the complementary slackness for the positive semi-definite constraint is written as:

$$\mathcal{Q}^{1/2} *_N \mathcal{M} *_N \mathcal{Q}^{1/2} = 0_{\mathcal{T}} \tag{54}$$

The tensor KKT conditions for the problem in (41)-(43) are given by (52), (53) and (54).

Notice that all the entries of $\mathcal{B}$, i.e $\mu_{\underline{i}_c}$, will be strictly greater than 0 at optimum because the inequality constraint must be met with equality at optimum. So $\mathcal{B}$ is a positive definite tensor, i.e. $\mathcal{B} \succ 0$ and hence invertible. Also since $\mu_{\underline{i}_c} > 0$, (53) can be written as:

$$\sum_{\underline{i}_r} \mathcal{Q}_{\underline{i},\underline{i}} - P_{\underline{i}_c} = 0, \quad \forall \underline{i}_c. \tag{55}$$

Let us define a tensor $\bar{\mathcal{H}} \in \mathbb{C}^{I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N}$ and its tensor EVD as:

$$\bar{\mathcal{H}} \triangleq \mathcal{B}^{-1/2} *_N (\mathcal{H}^H *_M \mathcal{H}) *_N \mathcal{B}^{-1/2} = \mathcal{V} *_N \bar{\mathcal{D}} *_N \mathcal{V}^H. \tag{56}$$

We can use the result of Theorem 1 from [73] by extending it to tensor case to solve (52) and (54) subject to $\mathcal{Q} \succeq 0$, $\mathcal{M} \succeq 0$ and $\mathcal{B} \succ 0$ to obtain the optimal $\mathcal{Q}$. The tensor version of Theorem 1 from [73] along with its proof has been included in Appendix A. Based on the results from Appendix A, the optimal $\mathcal{Q}$ is given by

$$\mathcal{Q}_{opt} = \mathcal{B}^{-1/2} *_N \mathcal{V} *_N \left(\mathcal{I}_N - \bar{\mathcal{D}}^{-1}\right)^+ *_N \mathcal{V}^H *_N \mathcal{B}^{-1/2} \tag{57}$$

where $\bar{\mathcal{D}}$ and $\mathcal{V}$ are obtained through the tensor EVD of $\bar{\mathcal{H}}$ (56) and $(\mathcal{Z})^+$ denotes a pseudo-diagonal tensor whose all the pseudo-diagonal entries are non-negative, i.e. $(\mathcal{Z}_{i_1,\ldots,i_N})^+ = \max\{0, \mathcal{Z}_{i_1,\ldots,i_N}\}$. Hence we can calculate the capacity as:

$$\begin{aligned}C &= \log\left[\det\left(\mathcal{H} *_N \mathcal{Q}_{opt} *_N \mathcal{H}^H + \mathcal{I}_M\right)\right] \\ &= \log\left[\det\left(\mathcal{Q}_{opt} *_N \mathcal{H}^H *_M \mathcal{H} + \mathcal{I}_N\right)\right] \\ &= \log\left[\det(\mathcal{B}^{-1/2} *_N \mathcal{V} *_N (\mathcal{I}_N - \bar{\mathcal{D}}^{-1})^+ *_N \mathcal{V}^H \right. \\ &\qquad \left. *_N \mathcal{B}^{-1/2} *_N \mathcal{H}^H *_M \mathcal{H} + \mathcal{I}_N)\right]\end{aligned}$$

$$= \log\Big[\det(\mathcal{V} *_N (\mathcal{I}_N - \bar{\mathcal{D}}^{-1})^+ *_N \mathcal{V}^H *_N \underbrace{\mathcal{B}^{-1/2} *_N \mathcal{H}^H *_M \mathcal{H} *_N \mathcal{B}^{-1/2}}_{\bar{\mathcal{H}}} + \mathcal{I}_N)\Big] \text{ (using (13), (56))}$$
$$= \log\Big[\det(\mathcal{V} *_N ((\mathcal{I}_N - \bar{\mathcal{D}}^{-1})^+ *_N \bar{\mathcal{D}} + \mathcal{I}_N) *_N \mathcal{V}^H)\Big]. \tag{58}$$

Since the determinant is the product of eigenvalues, we get:

$$C = \sum_{i_1,\ldots,i_N} \log((1 - \bar{d}^{-1}_{i_1,\ldots,i_N})^+ \cdot \bar{d}_{i_1,\ldots,i_N} + 1)$$
$$= \sum_{i_1,\ldots,i_N} (\log(\bar{d}_{i_1,\ldots,i_N}))^+ \tag{59}$$

where $\bar{d}_{i_1,\ldots,i_N}$ are the non-zero eigenvalues of $\bar{\mathcal{H}}$. Note that the optimum covariance tensor from (57) depends not only on the eigenvalues, but also the eigentensors of $\bar{\mathcal{H}}$. Hence ensuring an optimum input covariance leads to not only an optimum power allocation scheme, but also a joint multi-domain precoding at the transmitter which is required for achieving capacity. We will now simplify the expression for covariance under high SNR assumption and develop an algorithm to approximate the optimum covariance and capacity using (55), (57) and (59).

The inverse of the tensor $(\mathcal{H}^H *_M \mathcal{H})$ exists if the transformed matrix $f_{I_1,\ldots,I_N|I_1,\ldots,I_N}(\mathcal{H}^H *_M \mathcal{H})$ is full rank. In case the inverse does not exist, then a minimum norm least square solution can be adopted which aims to find an approximate tensor $\mathcal{T}$ such that $||(\mathcal{H}^H *_M \mathcal{H}) *_N \mathcal{T} - \mathcal{I}_N||^2$ is minimized [18]. Assuming high SNR, we can ignore $()^+$ and write (57) as:

$$\mathcal{Q} = \mathcal{B}^{-1/2} *_N \Big(\mathcal{I}_N - \underbrace{\mathcal{V} *_N \bar{\mathcal{D}}^{-1} *_N \mathcal{V}^H}_{\bar{\mathcal{H}}^{-1}}\Big) *_N \mathcal{B}^{-1/2}$$
$$= \mathcal{B}^{-1/2} *_N \big(\mathcal{I}_N - \mathcal{B}^{1/2} *_N (\mathcal{H}^H *_M \mathcal{H})^{-1} *_N \mathcal{B}^{1/2}\big) *_N \mathcal{B}^{-1/2} \quad \text{(using (56))}$$
$$= \mathcal{B}^{-1} - (\mathcal{H}^H *_M \mathcal{H})^{-1} \tag{60}$$

and (58) as:

$$C = \log\Big[\det(\mathcal{V} *_N ((\mathcal{I}_N - \bar{\mathcal{D}}^{-1}) *_N \bar{\mathcal{D}} + \mathcal{I}_N) *_N \mathcal{V}^H)\Big]$$
$$= \log\Big[\det(\mathcal{V} *_N \bar{\mathcal{D}} *_N \mathcal{V}^H)\Big] = \log\Big[\det(\bar{\mathcal{H}})\Big]$$
$$= \log\Big[\det(\mathcal{B}^{-1/2} *_N (\mathcal{H}^H *_M \mathcal{H}) *_N \mathcal{B}^{-1/2})\Big]$$
$$= \log\Big[\det(\mathcal{B}^{-1} *_N (\mathcal{H}^H *_M \mathcal{H}))\Big] \quad \text{(using (11))}$$
$$= \log\Big[\det(\mathcal{B}^{-1}) \cdot \det(\mathcal{H}^H *_M \mathcal{H})\Big]. \tag{61}$$

Thus under high SNR approximation, we can find the elements of $\mathcal{B}$ by substituting $\mathcal{Q}$ from (60) into (55) to get:

$$\sum_{\underline{i}_r} \big(\mathcal{B}^{-1} - (\mathcal{H}^H *_M \mathcal{H})^{-1}\big)_{\underline{i},\underline{i}} = P_{\underline{i}_c} \quad \forall \underline{i}_c \tag{62}$$
$$\sum_{\underline{i}_r} (\mathcal{B}^{-1})_{\underline{i},\underline{i}} - \sum_{\underline{i}_r} ((\mathcal{H}^H *_M \mathcal{H})^{-1})_{\underline{i},\underline{i}} = P_{\underline{i}_c} \quad \forall \underline{i}_c. \tag{63}$$

Let $N_{\underline{i}_r}$ denote the number of values that $\underline{i}_r$ can take. For example, if $\underline{i}_r = (i_1, \ldots, i_N)$, then $N_{\underline{i}_r} = I_1 \cdots I_N$, if $\underline{i}_r = (i_K, i_L)$, then $N_{\underline{i}_r} = I_K \cdot I_L$. Since $\mathcal{B}$ contains $\mu_{\underline{i}_c}$ on its pseudo-diagonal with each $\mu_{\underline{i}_c}$ appearing exactly $N_{\underline{i}_r}$ times, from (63) we can write:

$$N_{\underline{i}_r} \cdot \mu^{-1}_{\underline{i}_c} - \sum_{\underline{i}_r} ((\mathcal{H}^H *_M \mathcal{H})^{-1})_{\underline{i},\underline{i}} = P_{\underline{i}_c} \quad \forall \underline{i}_c \tag{64}$$
$$\mu_{\underline{i}_c} = \frac{N_{\underline{i}_r}}{P_{\underline{i}_c} + \sum_{\underline{i}_r} ((\mathcal{H}^H *_M \mathcal{H})^{-1})_{\underline{i},\underline{i}}} \quad \forall \underline{i}_c \tag{65}$$

which gives us all the entries of $\mathcal{B}$. The proposed solution in (65), (61) and (60) assumes high SNR as we have ignored the $()^+$ operation. To extend the solution for any SNR, we now present a scaling approach to approximate the input covariance tensor at any SNR setting, and verify that the resulting covariance satisfies the constraints. If the covariance $\mathcal{Q}$ obtained from (60) has negative eigenvalues, then we force the negative eigenvalues of $\mathcal{Q}$ to be zero. If the tensor EVD of $\mathcal{Q}$ is given as $\mathcal{U} *_N \mathcal{D} *_N \mathcal{U}^H$, the pseudo-diagonal elements of $\mathcal{Q}$ can be written as, $\mathcal{Q}_{\underline{i},\underline{i}} = \sum_{\underline{i}'} \mathcal{U}_{\underline{i},\underline{i}'} \mathcal{D}_{\underline{i}',\underline{i}'} \mathcal{U}^H_{\underline{i}',\underline{i}}$ Thus the brute force approach of setting the negative values in $\mathcal{D}$ to zero can result into larger values at the pseudo-diagonal of $\mathcal{Q}$. This in turn can make the solution infeasible, i.e. the power constraint $P_{\underline{i}_c}$ from (42) might not be met and we may get a different power allotted say $P'_{\underline{i}_c}$ where $P'_{\underline{i}_c} \geq P_{\underline{i}_c}$. Note that

$$P'_{\underline{i}_c} = \sum_{\underline{i}_r} \mathcal{Q}_{\underline{i},\underline{i}}, \quad \forall \underline{i}_c. \tag{66}$$

So we scale the resulting $\mathcal{Q}$ using another pseudo-diagonal tensor $\mathcal{S}$ such that the power constraints remain satisfied, i.e. $\mathcal{Q}_{scaled} = \mathcal{S} *_N \mathcal{Q} *_N \mathcal{S}^H$ where the pseudo-diagonal entries of scaling tensor $\mathcal{S}$ are given as:

$$\mathcal{S}_{\underline{i},\underline{i}} = \sqrt{\frac{P_{\underline{i}_c}}{P'_{\underline{i}_c}}}, \quad \forall \underline{i}_r \tag{67}$$

such that the pseudo-diagonal entries of $\mathcal{Q}_{scaled}$ become:

$$(\mathcal{Q}_{scaled})_{\underline{i},\underline{i}} = \mathcal{Q}_{\underline{i},\underline{i}} \cdot \frac{P_{\underline{i}_c}}{P'_{\underline{i}_c}}. \tag{68}$$

Hence, based on (66) we have

$$\sum_{\underline{i}_r} (\mathcal{Q}_{scaled})_{\underline{i},\underline{i}} = \sum_{\underline{i}_r} \mathcal{Q}_{\underline{i},\underline{i}} \cdot \frac{P_{\underline{i}_c}}{P'_{\underline{i}_c}} = P_{\underline{i}_c}. \tag{69}$$

Thus the choice of scaling operation ensures that $\mathcal{Q}_{scaled}$ satisfies the power constraints. A similar technique has been used for matrix-field water-filling in [74] where the diagonal entries of the covariance matrix are scaled to ensure that the resulting matrix is positive semi-definite. Such a scaling approach simplifies the computation of the input covariance but also makes it suboptimal and hence leads to an approximation of the capacity. However, this approximation gets better as SNR grows and is exact at sufficiently high SNR. The procedure is systematically presented in Algorithm 1. Finding the capacity is a convex optimization problem, hence can be solved using software tools such as CVX [75], which can be compared with the capacity obtained via the approximation in Algorithm 1 to assess the validity of the proposed approach.

We present such a comparison in the numerical examples (Figure 9) to illustrate the accuracy of the proposed approach.

Algorithm 1 Finding the Input Covariance Tensor.

Input $\mathcal{H}$, P_{i_c}, N_{i_r}
Initialize $flag \leftarrow 0$
Find pseudo-diagonal elements of $\mathcal{B}$ by calculating μ_{i_c} using (65)
Calculate $\mathcal{Q}$ using (60).
Perform tensor EVD of $\mathcal{Q} = \mathcal{U} *_N \mathcal{D} *_N \mathcal{U}^H$
for all $i_1, i_2, \ldots, i_N$
    if $\mathcal{D}_{i_1,\ldots,i_N} < 0$
        $\mathcal{D}_{i_1,\ldots,i_N} \leftarrow 0$
        $flag \leftarrow 1$
    end if
end for
if $flag == 1$
    Update $\mathcal{Q} \leftarrow \mathcal{U} *_N \mathcal{D} *_N \mathcal{U}^H$
    Calculate P'_{i_c} using (66)
    Find pseudo-diagonal tensor $\mathcal{S}$ using (67)
    Update $\mathcal{Q} \leftarrow \mathcal{S} *_N \mathcal{Q} *_N \mathcal{S}^H$
end if
return $\mathcal{Q}$

C. COMPLEXITY ANALYSIS OF ALGORITHM 1

Algorithm 1 essentially approximates the optimal input covariance tensor using (60) and a scaling process which ensures a feasible solution. The algorithm requires fixed amount of computational resources and can be deployed off line. Thus it is of interest to analyze the computational resources required to execute this algorithm. Modern day computational strategies often employ cloud services provided by suppliers such as AWS and Azure. Such cloud services often provide both the infrastructure and the platform/software as services on demand [76], [77], where the cost depends on the amount of required computational resources and time of execution. Given that Algorithm 1 requires extensive tensor operations which scale with the tensor size, using cloud services to implement it is an appealing option. Such cloud services provide several parallel and distributed computing infrastructures for faster and efficient computations [78]. Thus depending on the platform being considered and the amount of multi-core processors being employed for such implementation, the time of execution of the algorithm can significantly differ. However, independent of the computing infrastructure available, a suitable measure of the computational complexity can be specified in terms of the required number of mathematical operations to be performed in a given algorithm, as used in [79], [80]. Hence in this section, we analyze the computational complexity of Algorithm 1 in terms of the required number of flops for a given step. A flop is defined as a single floating point operation such as an addition, multiplication, subtraction, division, comparison ($>$, $<$, $==$), or a scalar square root, etc.

The complexity of Algorithm 1 primarily depends on finding the tensor inversion and the tensor EVD for a tensor of size $I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N$. Such tensor operations can be performed using various algorithms by employing the Einstein product as discussed in [18], [44], [49], [81]. Consider the definition of Einstein product across N modes between tensor $\mathcal{A}$ of size $I_1 \times \cdots \times I_P \times K_1 \times \cdots \times K_N$ and tensor $\mathcal{B}$ of size $K_1 \times \cdots \times K_N \times J_1 \times \cdots \times J_M$ from (3). Based on (3), computing a single element in $\mathcal{A} *_N \mathcal{B}$ requires $K_1 \cdot K_2 \cdots K_N$ multiplications and also $K_1 \cdot K_2 \cdots K_N - 1$ additions. There are total $I_1 \cdot I_2 \cdots I_P \cdot J_1 \cdot J_2 \cdots J_M$ elements in the tensor $\mathcal{A} *_N \mathcal{B}$. For ease of notation, we use $I = \prod_{i=1}^{P} I_i$, $J = \prod_{j=1}^{M} J_j$, and $K = \prod_{k=1}^{N} K_k$. So there are total IJ elements in $\mathcal{A} *_N \mathcal{B}$ where each element is computed using K multiplications and $K - 1$ additions. Subsequently finding all the elements of $\mathcal{A} *_N \mathcal{B}$ requires IJK multiplications and $(K-1)IJ$ additions. Hence a total of $IJK + (K-1)IJ = (2K-1)IJ$ flops are required for computing all the elements in $\mathcal{A} *_N \mathcal{B}$. Note that for any operation where the number of required flops is polynomial in its size n, i.e. the operation requires $a_0 n^p + a_1 n^{p-1} + \cdots + a_p$ flops (for fixed constants $a_0, a_1, \ldots, a_p$), we represent the complexity only using the highest power in n as $\mathcal{O}(n^p)$. Hence the complexity of the Einstein product which requires $(2K-1)IJ$ flops is given as $\mathcal{O}(IKJ)$ which is same as $\mathcal{O}((I_1 \cdots I_P) \cdot (K_1 \cdots K_N) \cdot (J_1 \cdots J_M))$. Thus for the specific case where both $\mathcal{A}$ and $\mathcal{B}$ are order $2N$ tensors of size $I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N$ each, the complexity of $\mathcal{A} *_N \mathcal{B}$ is given as $\mathcal{O}((I_1 \cdots I_N)^3)$.

Now assume channel $\mathcal{H}$ is of size $J_1 \times \cdots \times J_M \times I_1 \times \cdots \times I_N$ and N_{i_c} denotes the number of values that i_c can take. In Algorithm 1, the first two steps are input initialization. Step 3 requires computing the Einstein product over the common M modes of $\mathcal{H}^H$ and $\mathcal{H}$, which based on the discussion in the previous paragraph has a complexity of $\mathcal{O}((I_1 \cdots I_N)^2 (J_1 \cdots J_M))$. Further within step 3, it is required to find the inverse of $\mathcal{H}^H *_M \mathcal{H}$ which is an order $2N$ tensor. The inverse of an order $2N$ tensor can be calculated using the higher order bi-conjugate gradient (HOBG) method described in [18] or Newton's method (NM) which solves an iterative equation involving Einstein product between tensors of size $I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N$. Thus each iteration in NM has a computational cost of $\mathcal{O}((I_1 \cdots I_N)^3)$ [81]. It is important to note that the complexity of such iterative methods depends on the number of iterations, which in turn depends on the desired accuracy level set to achieve convergence. It was shown in [81] that NM requires a lower number of iterations as compared to HOBG to reach the same accuracy. The NM requires $\mathcal{O}(\log(I_1 \cdots I_N))$ iterations to converge to a fixed error bound [14]. Furthermore, [14] shows a method to reduce the per iteration complexity of NM, and thus perform tensor inversion in $\mathcal{O}(\log^2(I_1 \cdots I_N))$ parallel time units. Also, non-iterative methods such as Gauss elimination based on triangular decomposition of tensors [49] can be used for tensor inversion which requires a computational complexity of $\mathcal{O}((I_1 \cdots I_N)^3)$. Hence, the worst case complexity of tensor inversion without any use of parallel

processors is $\mathcal{O}((I_1 \cdots I_N)^3)$. Eventually step 3 calculates each $\mu_{\underline{i}_c}$ using (65) which requires $N_{\underline{i}_r} + 1$ additions and 1 division, and this needs to be done for all the $N_{\underline{i}_c}$ values that $\underline{i}_c$ can take. Thus this step requires $(N_{\underline{i}_r} + 2) \cdot N_{\underline{i}_c}$ flops and its complexity is $\mathcal{O}(N_{\underline{i}_r} \cdot N_{\underline{i}_c})$. Note that since $N_{\underline{i}_r} \cdot N_{\underline{i}_c} = I_1 \cdots I_N$, the complexity of finding $\mu_{\underline{i}_c}$ can be written as $\mathcal{O}(I_1 \cdots I_N)$. Step 4 which finds $\mathcal{Q}$ using (60) subtracts an order $2N$ tensor from a pseudo-diagonal tensor. Since the number of pseudo-diagonal elements are $I_1 \cdots I_N$, step 4 performs $I_1 \cdots I_N$ subtractions and thus has a complexity of $\mathcal{O}(I_1 \cdots I_N)$. Further step 5 finds the tensor EVD of an order $2N$ tensor $\mathcal{Q}$. The complexity of finding the EVD of a tensor of size $I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N$ using the Einstein product properties is $\mathcal{O}((I_1 \cdots I_N)^3)$ [16]. Algorithms which generalize the matrix eigen decomposition approaches to tensor EVD using the Einstein product properties can be found in [44], [49], [16, Algorithm C.2]. Steps 6 to 11 essentially perform the operation $\max(0, \mathcal{D}_{i_1,\ldots,i_N})$ on each of the $I_1 \cdots I_N$ pseudo-diagonal elements of the tensor $\mathcal{D}$. Hence it has a complexity of $\mathcal{O}(I_1 \cdots I_N)$. Step 12 is just a single scalar comparison, and step 13 updates $\mathcal{Q}$ using the Einstein product between tensors of order $2N$ with size $I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N$ for which the complexity is $\mathcal{O}((I_1 \cdots I_N)^3)$. Step 14 finds $P'_{\underline{i}_c}$ for all the $N_{\underline{i}_c}$ values of $\underline{i}_c$. Thus it performs $N_{\underline{i}_r}$ additions for each of the $N_{\underline{i}_c}$ values that $\underline{i}_c$ can take. Hence step 14 requires $N_{\underline{i}_r} \cdot N_{\underline{i}_r}$ flops and thus has a complexity of $\mathcal{O}(N_{\underline{i}_r} \cdot N_{\underline{i}_r})$ which is same as $\mathcal{O}(I_1 \cdots I_N)$. Step 15 calculates the scaling factor for all $\underline{i}_c$, thus performs $N_{\underline{i}_c}$ divisions and square roots. Hence it has a complexity of $\mathcal{O}(N_{\underline{i}_c})$. Finally, step 16 updates $\mathcal{Q}$ using the Einstein product between tensors of order $2N$ and thus has a complexity of $\mathcal{O}((I_1 \cdots I_N)^3)$.

TABLE 2. Computational complexity of Algorithm 1.

Step	Operation	Complexity
3	Find $\mathcal{H}^H *_M \mathcal{H}$	$\mathcal{O}((I_1 \cdots I_N)^2 \cdot J_1 \cdots J_M)$
3	Find $(\mathcal{H}^H *_M \mathcal{H})^{-1}$	$\mathcal{O}((I_1 \cdots I_N)^3)$
3	Find $\mu_{\underline{i}_c}$ using (65)	$\mathcal{O}(I_1 \cdots I_N)$
4	Find $\mathcal{Q}$ using (60)	$\mathcal{O}(I_1 \cdots I_N)$
5	EVD of $\mathcal{Q}$	$\mathcal{O}((I_1 \cdots I_N)^3)$
6-11	Check $\mathcal{D}$	$\mathcal{O}(I_1 \cdots I_N)$
13	Update $\mathcal{Q}$	$\mathcal{O}((I_1 \cdots I_N)^3)$
14	Find $P'_{\underline{i}_c}$ using (66)	$\mathcal{O}(I_1 \cdots I_N)$
15	Find $\mathcal{S}$ using (67)	$\mathcal{O}(N_{\underline{i}_c})$
16	Update $\mathcal{Q}$	$\mathcal{O}((I_1 \cdots I_N)^3)$

Table 2 summarizes these step by step computational complexity cost of Algorithm 1. The first column in Table 2 indicates the step number from Algorithm 1, second column describes the operation and third column states the complexity. We can observe that all the entries in complexity column of Table 2, have a complexity order of 3 (cubic) or less in $I_1 \cdots I_N$ except the first operation which has a complexity of $\mathcal{O}((I_1 \cdots I_N)^2 \cdot J_1 \cdots J_M)$. Hence on summing all the entries of the third column in Table 2, we see that the overall complexity of Algorithm 1 is given as $\mathcal{O}((I_1 \cdots I_N)^2 \cdot J_1 \cdots J_M) + \mathcal{O}((I_1 \cdots I_N)^3)$. Further in the case when $J_1 \cdots J_M \leq I_1 \cdots I_N$, the complexity of Algorithm 1 can be written as $\mathcal{O}((I_1 \cdots I_N)^3)$.

Note that the steps in Algorithm 1 with complexity of $\mathcal{O}((I_1 \cdots I_N)^3)$ primarily rely on the Einstein product operation. However, since this algorithm can be executed on multi-core computer systems, the time complexity of performing all the operations in Einstein product can be significantly reduced by making use of parallel processing. Several flops in the Einstein product can be executed simultaneously on a parallel computing platform. To see this, assume tensors $\mathcal{X}$ and $\mathcal{Y}$ of size $I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N$ each and $\mathcal{Z} = \mathcal{X} *_N \mathcal{Y}$. Then all the elements of $\mathcal{Z}$ can be computed using multiple processors as shown in Figure 1.

The white rectangular nodes in Figure 1 correspond to the individual multiplication operations. All the white nodes need to be added to compute a single element (the gray rectangular nodes) in tensor $\mathcal{Z}$. The addition of white nodes to generate the gray nodes can be done using a binary tree approach as shown in Figure 1 where all the addition operations at a given level of the tree are performed simultaneously. The figure illustrates this process for a single gray node, but further all the gray nodes can be computed simultaneously using similar binary tree approach if multiple processors are available.

For a specific gray node, the number of nodes in a binary tree at each level starting from root (gray node) are $2^0, 2^1, 2^2, \ldots, 2^h$ where h is the depth of the tree and 2^h is the number of leaf nodes. We can have similar binary trees for each of the gray nodes. Since in Figure 1, the number of leaf nodes (white nodes) are $I_1 \cdots I_N$, we get that the depth of the tree corresponding to each gray node is $\lceil \log(I_1 \cdots I_N) \rceil$. We use the ceil operator as $I_1 \cdots I_N$ may not always be a power of 2. Hence we can say that the height of the tree is $\mathcal{O}(\log(I_1 \cdots I_N))$. Since all the gray nodes can be computed simultaneously, all the individual elements of $\mathcal{X} *_N \mathcal{Y}$ can be calculated in $\mathcal{O}(\log(I_1 \cdots I_N))$ parallel time units. Such a parallel processing method can significantly reduce the time complexity of calculating the Einstein product, and subsequently other tensor operations which rely on the Einstein product such as tensor inversion and EVD. The other steps in Algorithm 1 which have a complexity of $\mathcal{O}(I_1 \cdots I_N)$ or less (step 15), can also be performed faster using parallel processors. For example in step 3, all the $\mu_{\underline{i}_c}$ of (65) for different $\underline{i}_c$ can be calculated simultaneously on parallel processors. Similarly all the $I_1 \cdots I_N$ operations in steps 4 and 6-11, can be performed simultaneously. In steps 14 and 15, the $P'_{\underline{i}_c}$ from (66) and the scaling factors from (67) for all the $\underline{i}_c$ can also be computed simultaneously. Hence Algorithm 1 can be suitably adapted to run on parallel processing multi-core computer systems depending on the number of processors available. Several platforms such as MATLAB provide support for parallel implementation of such algorithms. A more detailed study into the parallelization of the proposed algorithm for faster time complexity has been left for future

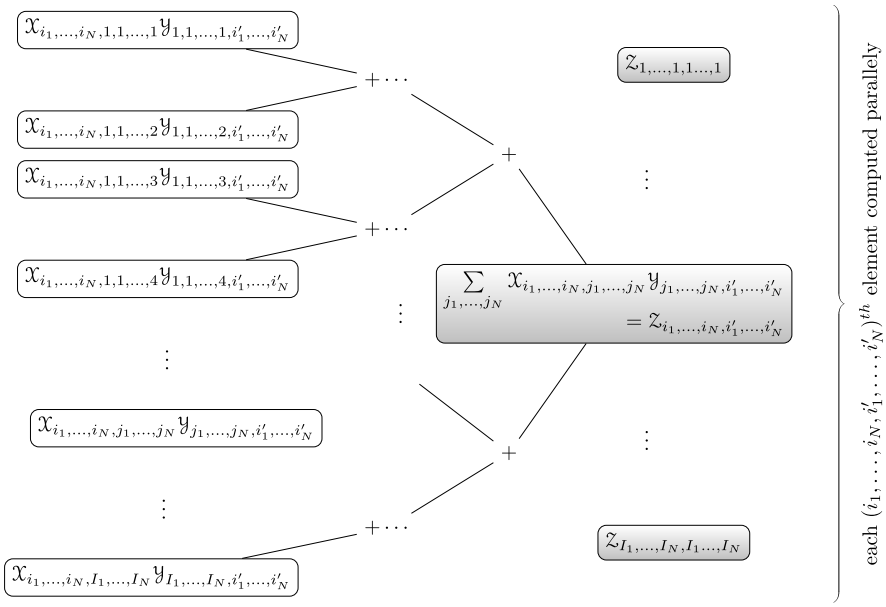

FIGURE 1. Parallel execution of Einstein product.

investigation. In general, developing fast and efficient algorithms for several tensor functions depending on the Einstein product and its properties is an active area of research in the numerical tensor algebra community [82]–[87].

D. COMPARING DIFFERENT CONSTRAINTS

Let the set of all possible positive semi-definite tensors of size $I_1 \times \cdots \times I_N \times I_1 \times \cdots \times I_N$ be represented by $\mathbb{Q}$. Let the feasible set for the optimization problem in (41)-(43) for two different settings $\underline{i}_{c1}$ and $\underline{i}_{c2}$ be $\mathbb{Q}_1$ and $\mathbb{Q}_2$ respectively. Assume $\underline{i}_{c1}$ is a subsequence of $\underline{i}_{c2}$. For instance, let $\underline{i}_{c2} = (i_1, i_2, i_3)$, and $\underline{i}_{c1} = (i_1, i_2)$. So $\mathbb{Q}_1$ represents a set of all positive semi-definite tensors $\mathcal{Q}$ which satisfies

$$\sum_{i_3,i_4,\ldots,i_N} \mathcal{Q}_{i_1,\ldots,i_N,i_1,\ldots,i_N} \leq P_{i_1,i_2} \quad \forall (i_1, i_2) \tag{70}$$

and $\mathbb{Q}_2$ represents a set of all positive semi-definite tensors $\mathcal{Q}$ which satisfies

$$\sum_{i_4,\ldots,i_N} \mathcal{Q}_{i_1,\ldots,i_N,i_1,\ldots,i_N} \leq P_{i_1,i_2,i_3} \quad \forall (i_1, i_2, i_3). \tag{71}$$

In (70), the first two domains are under power constraints, whereas in (71), the power constraints span the third domain as well with $\sum_{i_3} P_{i_1,i_2,i_3} = P_{i_1,i_2}$. Summing over i_3 in (71) gives (70). Hence every $\mathcal{Q}$ that satisfies (71) will also satisfy (70), showing that the set of $\mathcal{Q}$ satisfying (71) is a subset of the set of $\mathcal{Q}$ satisfying (70), i.e. $\mathbb{Q}_2 \subseteq \mathbb{Q}_1$.

Let the optimal value of the objective function in the optimization problem in (41)-(43) for set $\mathbb{Q}_1$ be C_1 and for $\mathbb{Q}_2$ be C_2. From the basic principles of convex optimization [71], it is known that a globally optimal point is also locally optimal. Hence if C_1 is the maximum of the objective function over the set of constraints $\mathbb{Q}_1$, then C_1 is also the maximum of the objective function over $\mathbb{Q}_2$ since $\mathbb{Q}_2 \subseteq \mathbb{Q}_1$. Hence $C_2 \leq C_1$, where equality is possible if the optimal Q lies in the set $\mathbb{Q}_2$. This holds for any configuration of $\underline{i}_{c1}$ and $\underline{i}_{c2}$ so far as $\underline{i}_{c1}$ is a subsequence of $\underline{i}_{c2}$. Essentially, as more domains are put under constraints, the feasible set for the optimization problem shrinks, possibly lowering the achievable capacity.

For instance, consider a 2×2 input where the two domains are antenna and time slots. Let the capacity achieved under total power constraint P, be C_1. Let the capacity achieved under per antenna power constraints of P_1 for antenna 1 and P_2 for antenna 2 such that $P_1 + P_2 = P$, be C_2. Since the set of feasible solution with per antenna power constraints is a subset of the set of feasible solution with total power constraint we have $C_2 \leq C_1$. Similarly, capacity achieved under power constraints per element, i.e. $P_{11}, P_{12}, P_{21}, P_{22}$ where $P_{i,j}$ represents power budget on the ith antenna and the jth time slot, such that $P_{11}+P_{12} = P_1$, $P_{21}+P_{22} = P_2$, be C_3, then $C_3 \leq C_2$. This has also been shown in [88] for a MIMO channel where capacity under per antenna power constraint is always smaller than the capacity under sum power constraint.

E. CAPACITY UNDER SUM POWER CONSTRAINT

Under the sum power constraint of (47), $\underline{i}_c$ is empty and hence there is a single Lagrange multiplier μ associated with the constraint (47). Hence the tensor $\mathcal{B}$, which contains the Lagrange multipliers on its pseudo-diagonal, will be a scaled identity tensor. Substituting $\mathcal{B} = \mu \mathcal{I}_N$ in (56) gives $\bar{\mathcal{H}} = \mu^{-1} \cdot (\mathcal{H}^H *_M \mathcal{H}) = \mathcal{V} *_N (\mu^{-1} \cdot \mathcal{D}) *_N \mathcal{V}^H$, and subsequently (57) becomes

$$\mathcal{Q} = \mathcal{V} *_N \left(\mu^{-1}\mathcal{I}_N - \mathcal{D}^{-1}\right)^+ *_N \mathcal{V}^H. \tag{72}$$

Substituting $\bar{d}_{i_1,\ldots,i_N} = d_{i_1,\ldots,i_N}/\mu$ in (59) gives:

$$C = \sum_{i_1,\ldots,i_N} \log\Big[\Big(\frac{1}{\mu} - \frac{1}{d_{i_1,\ldots,i_N}}\Big)^+ \cdot d_{i_1,\ldots,i_N} + 1\Big] = \sum_{i_1,\ldots,i_N} \Big(\log\Big(\frac{d_{i_1,\ldots,i_N}}{\mu}\Big)\Big)^+ \tag{73}$$

where a^+ denotes $\max\{0, a\}$, $d_{i_1,\ldots,i_N}$ are the non-zero eigenvalues of $\mathcal{H}^H *_M \mathcal{H}$ and $1/\mu$ is chosen to satisfy (55)

$$\text{tr}(\mathcal{Q}) = \sum_{i_1,\ldots,i_N} \Big(\frac{1}{\mu} - \frac{1}{d_{i_1,\ldots,i_N}}\Big)^+ = P. \tag{74}$$

The optimum covariance derived in (72) is a generalization of water-filling solution for the MIMO matrix channel to multiple domains. Hence under sum power constraint, we perform the tensor EVD of $\mathcal{H}^H *_M \mathcal{H}$ to get $\mathcal{V}$ and $\mathcal{D}$. Further, we use (74) to find μ and subsequently use (73) to find the capacity.

F. MULTIPLEXING GAIN

We can also characterize the capacity contribution by each constrained domain separately and the multiplexing gain under various power constraints. For a fixed tensor channel, the tensor $\mathcal{B}^{-1}$ is pseudo-diagonal with entries $\mu_{\underline{i}_c}^{-1}$, hence $\det(\mathcal{B}^{-1}) = \prod_{\underline{i}} \mu_{\underline{i}_c}^{-1} = \prod_{\underline{i}_r}(\prod_{\underline{i}_c} \mu_{\underline{i}_c}^{-1}) = \Big(\prod_{\underline{i}_c} \mu_{\underline{i}_c}^{-1}\Big)^{N_{\underline{i}_r}}$. For instance, assume that out of N input domains, elements of the first domain are under individual power constraints such that $\underline{i}_c = (i_1)$ and $\underline{i}_r = (i_2, \ldots, i_N)$, then

$$\det(\mathcal{B}^{-1}) = \prod_{i_1,i_2,\ldots,i_N} \mu_{i_1}^{-1} = \Big(\prod_{i_1} \mu_{i_1}^{-1}\Big)^{I_2 \cdot I_3 \cdots I_N} \tag{75}$$

Also $\det(\mathcal{H}^H *_M \mathcal{H}) = \prod_{\underline{i}} d_{\underline{i}}$ where $d_{\underline{i}}$ are eigenvalues of $\mathcal{H}^H *_M \mathcal{H}$. Hence using (61) we have

$$C = \log\Big[\prod_{\underline{i}} \frac{d_{\underline{i}}}{\mu_{i_1}}\Big] = \sum_{\underline{i}} \log \frac{d_{\underline{i}}}{\mu_{i_1}} = \sum_{i_1} \Big(\sum_{i_2,\ldots,i_N} \log \frac{d_{\underline{i}}}{\mu_{i_1}}\Big) = \sum_{i_1} C_{i_1} \tag{76}$$

where $C_{i_1} = \sum_{i_2,\ldots,i_N} \log \frac{d_{\underline{i}}}{\mu_{i_1}}$ can be seen as the contribution of the i_1th element of the constrained domain to the overall capacity. For instance if the first domain refers to space domain, then C_{i_1} is the capacity contribution of the i_1th antenna. For any general case where $\underline{i}_c$ contains the indices of domains under constraint, we can write:

$$C = \log\Big[\prod_{\underline{i}_c,\underline{i}_r} \frac{d_{\underline{i}}}{\mu_{\underline{i}_c}}\Big] = \sum_{\underline{i}_c} \Big(\sum_{\underline{i}_r} \log \frac{d_{\underline{i}}}{\mu_{\underline{i}_c}}\Big) = \sum_{\underline{i}_c} C_{\underline{i}_c} \tag{77}$$

where $C_{\underline{i}_c} = \sum_{\underline{i}_r} \log \frac{d_{\underline{i}}}{\mu_{\underline{i}_c}}$ can be seen as the contribution of the $\underline{i}_c$th element of the constrained domains to the overall capacity. Substituting $\mu_{\underline{i}_c}$ from (65) into $C_{\underline{i}_c}$, we can further write:

$$C_{\underline{i}_c}(P_{\underline{i}_c}) = \sum_{\underline{i}_r} \log \frac{d_{\underline{i}}}{\mu_{\underline{i}_c}} \tag{78}$$

$$= \sum_{\underline{i}_r} \log\Big(d_{\underline{i}} \cdot \frac{P_{\underline{i}_c} + \sum_{\underline{i}_r}((\mathcal{H}^H *_M \mathcal{H})^{-1})_{\underline{i},\underline{i}}}{N_{\underline{i}_r}}\Big) \tag{79}$$

$$= \sum_{\underline{i}_r} \Big[\log\Big(P_{\underline{i}_c} + \sum_{\underline{i}_r}((\mathcal{H}^H *_M \mathcal{H})^{-1})_{\underline{i},\underline{i}}\Big) + \log\Big(\frac{d_{\underline{i}}}{N_{\underline{i}_r}}\Big)\Big] \tag{80}$$

We can write the multiplexing gain provided by $\underline{i}_c$th constrained domain as:

$$\chi_{\underline{i}_c} = \lim_{P_{\underline{i}_c} \to \infty} \frac{C_{\underline{i}_c}(P_{\underline{i}_c})}{\log(P_{\underline{i}_c})} \tag{81}$$

Since for large $P_{\underline{i}_c}$, we have $\log\Big(P_{\underline{i}_c} + \sum_{\underline{i}_r}((\mathcal{H}^H *_M \mathcal{H})^{-1})_{\underline{i},\underline{i}}\Big) \approx \log(P_{\underline{i}_c})$, hence using (81) and (80) we get

$$\chi_{\underline{i}_c} = \lim_{P_{\underline{i}_c} \to \infty} \frac{\sum_{\underline{i}_r}\Big[\log\Big(P_{\underline{i}_c}\Big) + \log\Big(\frac{d_{\underline{i}}}{N_{\underline{i}_r}}\Big)\Big]}{\log(P_{\underline{i}_c})} \tag{82}$$

$$= \lim_{P_{\underline{i}_c} \to \infty} \frac{N_{\underline{i}_r}\log(P_{\underline{i}_c}) + \sum_{\underline{i}_r}\log\Big(\frac{d_{\underline{i}}}{N_{\underline{i}_r}}\Big)}{\log(P_{\underline{i}_c})} \tag{83}$$

$$= N_{\underline{i}_r} \tag{84}$$

In general, since $N_{\underline{i}_r}$ represents the product of dimensions of the domains not under individual constraints, it increases exponentially in the number of unconstrained domains. As a result, for tensor channels the multiplexing gain increases exponentially with the increase in the number of unconstrained domains. Note that in deriving the multiplexing gain, for simplicity we have assumed that all the eigenvalues of $\mathcal{H}^H *_M \mathcal{H}$ are non-zero, which may not always be the case. In general, the capacity is a function of the given channel's specific singular values, some of which may be zero. In that case depending on which and how many singular values are zero, the multiplexing gain will be different and may not increase exponentially with increase in domains.

Based on (84), the multiplexing gain achieved under sum power constraint is $I_1 \cdots I_N$. However (84) assumes that inverse of $(\mathcal{H}^H *_M \mathcal{H})$ exists which will be the case if all its eigenvalues are non-zero. If that is not the case, then an approximation of the inverse $(\mathcal{H}^H *_M \mathcal{H})$ can be calculated using the higher order bi-conjugate gradient method described in [18]. Next we analyze the multiplexing gain, also known as pre-log, associated with a tensor channel under sum power constraint. Note that the number of non-zero eigenvalues of $(\mathcal{H}^H *_M \mathcal{H})$ will be same as the number of non-zero singular values of $\mathcal{H}$. Let $I_1 \cdot I_2 \cdots I_N = I$ and $J_1 \cdot J_2 \cdots J_M = J$, then the number of non-zero singular

values R are less than or equal to I and J, i.e. $R \leq \min\{I, J\}$ where equality is met if all the the singular values of $\mathcal{H}$ are non-zero.

The pre-log, denoted by χ, is calculated at a very high SNR, for which water-filling reduces to uniform power allocation over the non-zero eigen channels, i.e. $\left(\frac{1}{\mu} - \frac{1}{d_{i_1,\ldots,i_N}}\right)^+ \approx \frac{P}{R}$. Hence using (73), the capacity becomes

$$C = \sum_{\substack{i_1,\ldots,i_N \\ d_{i_1,\ldots,i_N} \neq 0}} \left(\log\left(1 + \frac{P}{R} d_{i_1,\ldots,i_N}\right)\right) \tag{85}$$

As $P \to \infty$, (85) simplifies to:

$$C = \sum_{\substack{i_1,\ldots,i_N \\ d_{i_1,\ldots,i_N} \neq 0}} \log\left(\frac{P}{R} d_{i_1,\ldots,i_N}\right) \tag{86}$$

$$= \sum_{\substack{i_1,\ldots,i_N \\ d_{i_1,\ldots,i_N} \neq 0}} \left[\log(P) + \log\left(\frac{d_{i_1,\ldots,i_N}}{R}\right)\right] \tag{87}$$

$$= R\log(P) + \sum_{\substack{i_1,\ldots,i_N \\ d_{i_1,\ldots,i_N} \neq 0}} \log\left(\frac{d_{i_1,\ldots,i_N}}{R}\right) \tag{88}$$

$$\Rightarrow \chi = \lim_{P\to\infty} \frac{C}{\log(P)} = R \tag{89}$$

Assuming all the singular values of the tensor channel are non-zero, we have $R = \min\{I_1 \cdot I_2 \cdots I_N, J_1 \cdot J_2 \cdots J_M\}$. For a conventional MIMO matrix channel, capacity pre-log is known to be less than or equal to the minimum of the number of transmit and receive antennas [21] which is the specific case of the tensor pre-log. For MIMO matrix channel $N = M = 1$, and we have $\chi = \min\{I_1, J_1\}$ where I_1 and J_1 are the number of transmit and receive antennas respectively. For a tensor case, it is interesting to see that assuming equal dimension sizes on each domain i.e. $I_1 = I_2 = \cdots = I_N = J_1 = J_2 \cdots J_M = L$, then the capacity pre-log can be given as:

$$\begin{aligned} \chi &= \min\{I_1 \cdot I_2 \cdots I_N, J_1 \cdot J_2 \cdots J_M\} \\ &= \min\{L^N, L^M\} = L^{\min\{N,M\}} \end{aligned} \tag{90}$$

which is exponential in the number of domains.

VI. NUMERICAL EXAMPLES AND APPLICATIONS

In this section, we present numerical examples to illustrate previous results. We also present an example of a MIMO GFDM system modelled using the tensor framework.

A. EXAMPLES OF ERGODIC CAPACITY WHEN CHANNEL REALIZATIONS ARE KNOWN

Our results assume that the channel is deterministic. For the numerical examples, rather than using a specific channel tensor, we generate the channel using the Rayleigh model as in [21], [89], [90]. The channel tensor consists of realizations of i.i.d. circularly symmetric complex Gaussian entries of zero mean and unit variance such that $\mathbb{E}[|\mathcal{H}_{j_1,\ldots,j_M,i_1\ldots,i_N}|^2] = 1$ [89]. These channel realizations are known at the transmitter and receiver.

Let the capacity of a deterministic tensor channel $\mathcal{H}$ be denoted as $C(\mathcal{H})$. Assume now that we have capacities of K such channels denoted by $C(\mathcal{H}^{(k)})$ for $k = 1, \ldots, K$ where the tensor $\mathcal{H}^{(k)}$ consist of realizations of complex Gaussian random variables. The average capacity of K such deterministic channels is given by $\bar{C}_K = \frac{1}{K}\sum_{k=1}^{K} C(\mathcal{H}^{(k)})$. Due to law of large numbers, as $K \to \infty$, we have $\bar{C}_K \to \mathbb{E}[C(\mathcal{H})]$ where $\mathcal{H}$ is a tensor of Gaussian random variables. All the numerical results presented in this section present $\bar{C}_K$ for $K = 100$, which can be interpreted as the ergodic capacity of a random tensor channel when its realizations are known at the transmitter and the receiver. The SNR is defined as P/σ^2 as used in [21], where P is the sum power constraint or the total transmit power and the noise tensor contains i.i.d. circularly symmetric complex Gaussian entries with zero mean and variance $\sigma^2 = 1$.

1) COMPARISON WITH UNIFORM POWER ALLOCATION

If the channel is not known then the ergodic capacity of a random tensor channel, $\mathbb{E}[C(\mathcal{H})]$ will be a function of the pdf of $\mathcal{H}$. For a MIMO matrix channel, capacity when the channel is Rayleigh distributed with channel state information available at only the receiver is achieved by uniform power allocation at the transmitter [21]. In this paper we present a comparison between uniform power allocation and the tensor water-filling power allocation for a tensor channel. Figure 2 presents capacity in bits-per-tensor-symbol under sum power constraint for 2 different sizes of tensor channel with uniform power allocation across all the transmit tensor elements and optimal tensor water-filling approach. Similar to the MIMO matrix case, it can be seen that uniform power allocation is suboptimal but at high SNR, it gives similar capacity as achieved by optimal power allocation. Also, it can be seen that the total capacity is higher when we increase the number of input and output domains. Next, we present examples to illustrate the impact of different domains and their dimensions on tensor channel capacity.

2) CAPACITY FOR DIFFERENT SIZES OF CHANNEL TENSOR

In Figure 3 we present the channel capacity in bits/tensor-symbol for a fourth order channel tensor under a sum power constraint at SNR of 10 dB. Input and output are order-2 tensors of size $X \times Y$ each, corresponding to a $X \times Y \times X \times Y$ tensor channel. It is seen that increasing X and Y individually causes an increase in total capacity. To understand the effect of different types of channel, we also find capacity when the channel is normalized such that the total receive average power is identical to the transmitted power. Such a normalization has been suggested in [91], [92] for the MIMO channel case. In tensor channels, such a normalization is achieved when the individual entries of the channel tensor $\mathcal{H} \in \mathbb{C}^{J_1 \times \cdots \times J_M \times I_1 \times \cdots \times I_N}$ are generated as circular complex

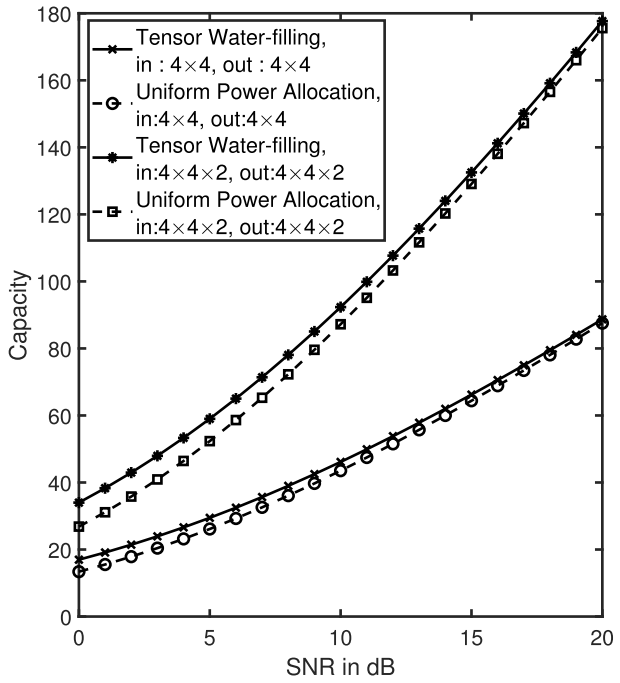

FIGURE 2. Capacity [bits/tensor-symbol] vs SNR with uniform power allocation.

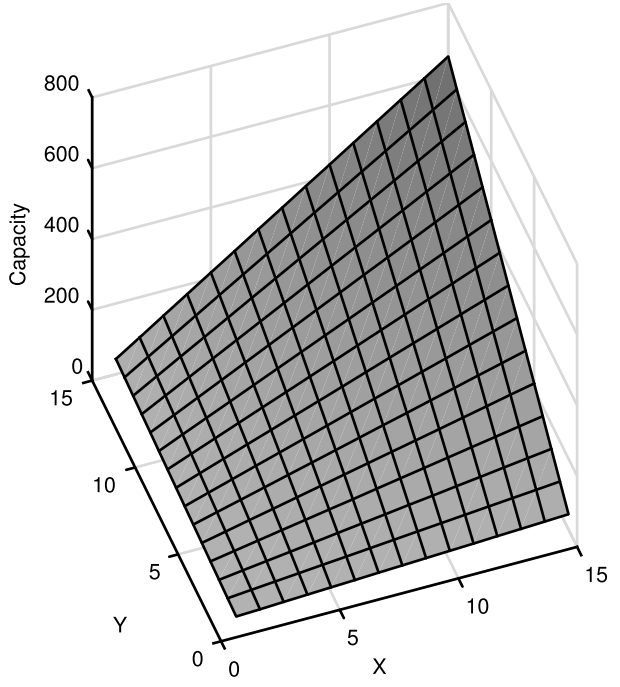

FIGURE 3. Capacity [bits/tensor-symbol] vs X vs Y for channel with $X \times Y$ size input and output tensor.

Gaussian with zero mean and variance $1/(J_1 \cdots J_M)$. Figure 4 shows the capacity of such a normalized tensor channel of size $X \times Y \times X \times Y$ with both input and output of size $X \times Y$ each, under sum power constraint at a fixed SNR of 10 dB. On increasing the size of input and output tensors, the rate of increase in capacity is lower as compared to Figure 3,

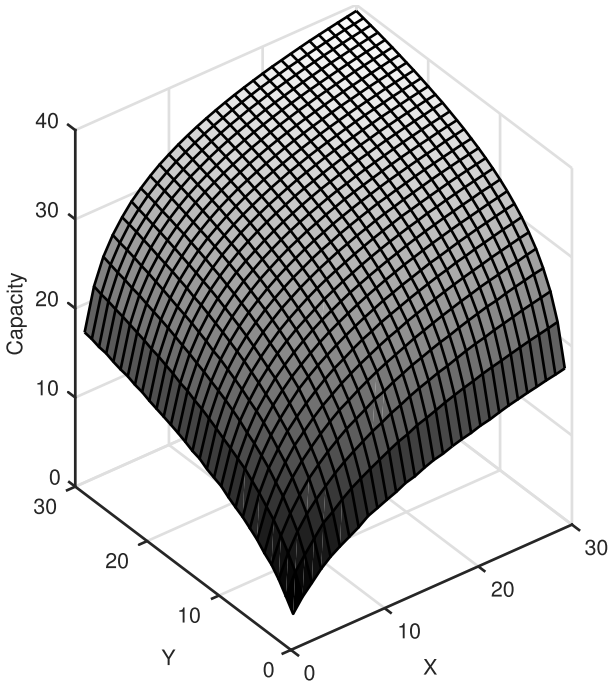

FIGURE 4. Capacity [bits/tensor-symbol] for normalized channel vs X vs Y where $X \times Y$ is the size of input and output tensor.

and the capacity tends to reach a saturation for large values of X and Y. For the case where channel gain is unity, and hence transmit and receive signal power are same, capacity is plotted against received SNR for a $2\times2\times2\times2$ tensor channel in Figure 5. A comparison with corresponding scalar and matrix channels with the same received signal power is also presented. The gain in the capacity achieved by moving from scalar to a tensor channel can be attributed to the multiplexing gain provided by the tensor channel which increases with the number of domains.

For the rest of the numerical examples, the more widely used model from [21], [89] where the channel entries are circular complex Gaussian with zero mean and unit variance is employed, as for Figure 3. In Figure 6, we present the capacity for a fourth order tensor channel under sum power constraint where the output is fixed as a 2×2 tensor and input is $X \times Y$ tensor at 10 dB SNR, under sum power constraint. The rate of increase in capacity for X and Y is slower as compared to Figure 3 for higher values of X and Y as the output tensor size is fixed as 2×2. So the capacity pre-log which is bounded by $\min\{X \cdot Y, 2 \cdot 2\} = 4$, does not increase with increasing X and Y. We see that increasing the size of individual domains of the input tensor does not provide significant gain if the number and size of the corresponding domains of the output tensor are fixed.

3) CAPACITY UNDER DIFFERENT DOMAIN POWER CONSTRAINTS

In this section we compare different possible power constraints. In order to approximate the optimum input

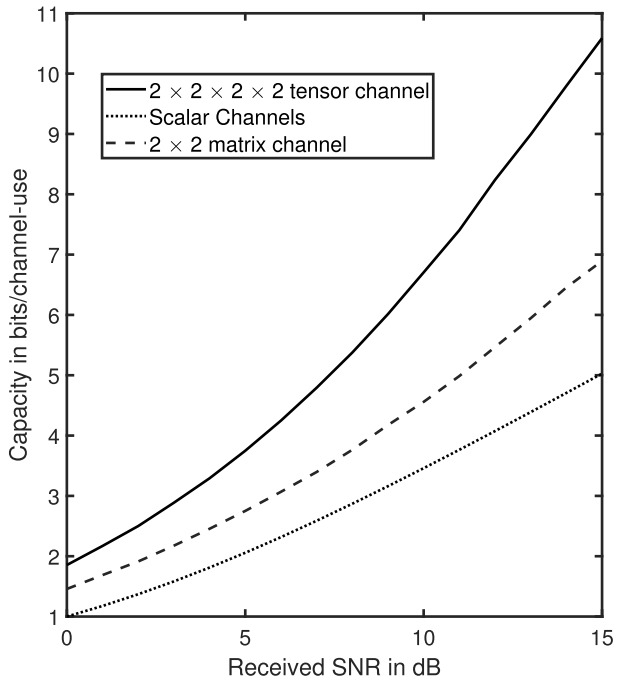

FIGURE 5. Channel capacity for tensor, matrix and scalar channel with same received signal power.

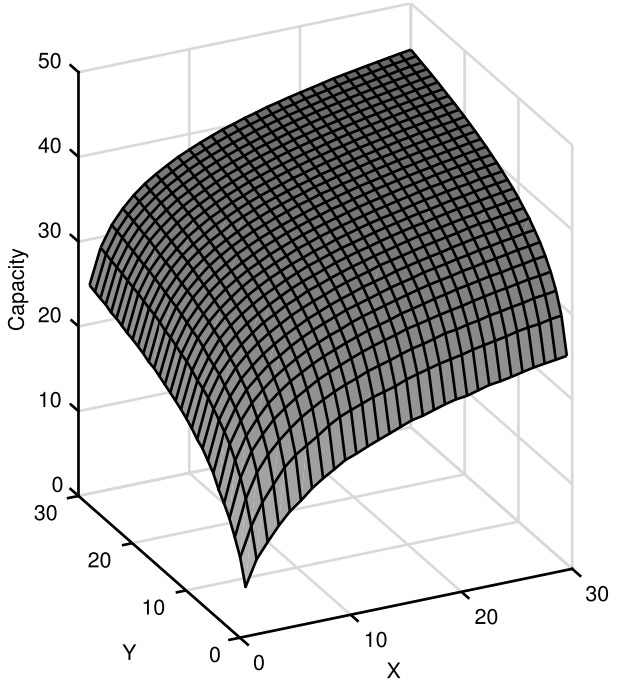

FIGURE 6. Capacity [bits/tensor-symbol] vs X vs Y for channel with $X \times Y$ size input tensor and 2×2 output tensor.

covariance and thereby capacity, under per domain element and per element power constraints, we use Algorithm 1.

Figure 7 illustrates the capacity under sum power constraint and per domain element power constraints for a $2 \times 2 \times 2 \times 2$ channel with power constraint on input tensor of size 2×2. The power budgets on one of the domains of the input tensor are $P_1 = x \cdot P$ and $P_2 = (1 - x) \cdot P$. To put it in context, let the two domains be space and time. Then this constraint reflects that the power on first time slot for both the antennas is P_1 and on second slot for both the antennas is P_2 with total available power $P_1 + P_2 = P$. The plot in Figure 7 is presented for capacity against $x = P_1/P$ at 10 dB SNR. The flat line represents the capacity under sum power constraint which shows no variation with x, and the curved line shows the capacity with per domain element power constraints. As can be observed from Figure 7, the capacity under per domain element constraints is always lower than the capacity under sum power constraint, and these become very close to each other when $x \approx 0.5$, i.e. uniform power is allotted to the elements of the constrained domain. Note that such a behaviour is observed over an average of 100 channel realizations. For a given specific realization, the two curves may not meet at $x = 0.5$. For the MIMO case, a similar numerical result has been presented in [88] under per antenna power constraints.

In Figure 8 we present the capacity under per element power constraints and compare it with sum power constraint. If total available power is P, then as before $P_1 = x \cdot P$ and $P_2 = (1 - x) \cdot P$. Further, $P_{11} = y \cdot P_1, P_{12} = (1 - y) \cdot P_1, P_{21} = y \cdot P_2$ and $P_{22} = (1 - y) \cdot P_2$. Thus, with $0 \leq x, y \leq 1$, $P_{11}, P_{12}, P_{21}, P_{22}$ denote the individual power constraints on all the four elements of the input tensor such that $P_{11}+P_{12}+P_{21}+P_{22} = P$. With different choices of x and y, we achieve different per element power constraints such that total power remains P. The capacity with per element power constraints against x and y at SNR of 15 dB is presented in Figure 8. The flat surface represents capacity under sum power constraint which shows no variation with x and y, and the curved surface shows capacity under per element power constraints. It can be seen that for different values of x and y, the capacity achieved under per element constraints can be significantly lower than the capacity achieved under sum power constraint.

Note that Algorithm 1 only approximates the optimum input covariance and thus does not provide the exact capacity at low SNRs. However, since the problem at hand is a convex optimization problem, several standard software tools for numerical optimization can be used to find the capacity. To analyze how well the scaling approximation in Algorithm 1 works, we present a comparison between the capacity calculated through the convex optimization software tool CVX [75], and the capacity approximated through Algorithm 1. Figure 9 presents the capacity for sum power constraint, per domain element power constraints where a single domain of dimension 2 is constrained with power budgets P_1, P_2, and per element power constraints where all the four elements have different power budgets $P_{11}, P_{12}, P_{21}, P_{22}$ against SNR for $x = y = 0.1$. We present such results for capacity calculated by using two approaches. First approach uses Algorithm 1 and the graphs are presented using solid curves. Second approach uses CVX and the graphs are presented using dashed curves. As can be seen, the capacity

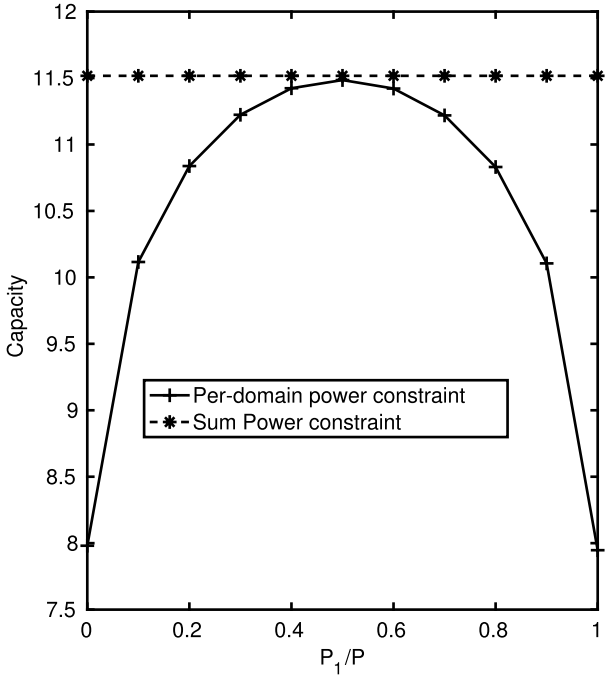

FIGURE 7. Capacity [bits/tensor-symbol] for a 2 × 2 × 2 × 2 tensor channel under per domain element power constraints and total power constraint at 10 dB SNR.

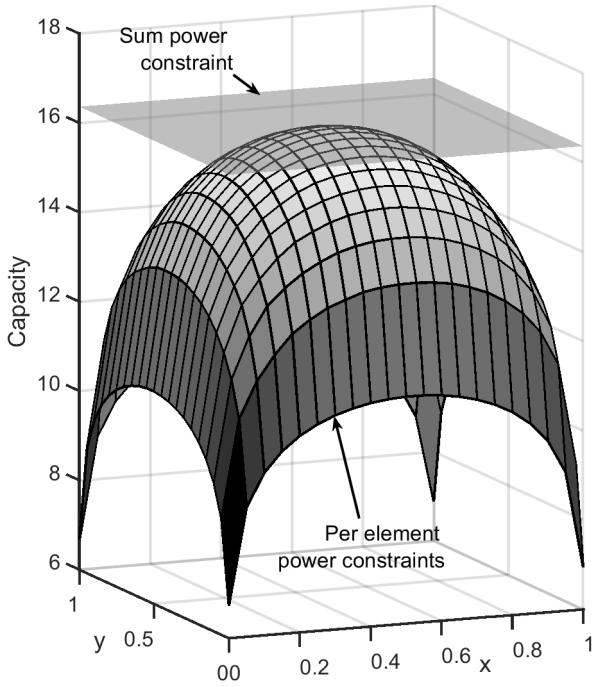

FIGURE 8. Capacity [bits/tensor-symbol] for per element power constraints vs x vs y at 15 dB SNR.

calculated via the approximation of Algorithm 1 matches very closely to the one calculated via CVX, and is almost indistinguishable at moderate to high SNR. This shows that Algorithm 1 provides a reasonably good approximation to the optimal solution at low SNR, while providing an exact solution at high SNR. Furthermore, it can be seen that the capacity under per domain element and per element constraints is always upper bounded by the capacity under sum power constraint. The capacity increases with SNR for all three cases, but the performance difference also gradually increases between the three solid curves, with sum power constraint performing the best, followed by per domain element and lastly, the per element constraint.

Figure 10 compares capacity under sum power and per domain element power constraints with $x = 0.1$ for different N where both input and output are order N. The channel is an order $2N$ tensor and the size of each domain is 2. The capacity increases exponentially with N in both the cases. However, the capacity under per domain element constraint is always upper bounded by the capacity under sum power constraint.

4) CORRELATED TENSOR CHANNEL

Consider an order N random tensor $\mathcal{H} \in \mathbb{C}^{I_1 \times \cdots \times I_N}$ with i.i.d. zero mean and unit variance entries. Let $\Psi^{(n)} \in \mathbb{C}^{I_n \times I_n}$ for $n = 1, \ldots, N$ be a sequence of Hermitian matrices such that $\Psi^{(n)} = \mathrm{A}^{(n)}\mathrm{A}^{(n)H}$ where $\mathrm{A}^{(n)} \in \mathbb{C}^{I_n \times I_n}$ is the square root matrix of $\Psi^{(n)}$. The mode-n product of tensor $\mathcal{H}$ across all the modes with these matrices (also known as Tucker product) is expressed as [93]:

$$\mathcal{H}^{corr} = \mathcal{H} \times_1 \mathrm{A}^{(1)} \times_2 \mathrm{A}^{(2)} \times_3 \cdots \times_N \mathrm{A}^{(N)} \tag{91}$$

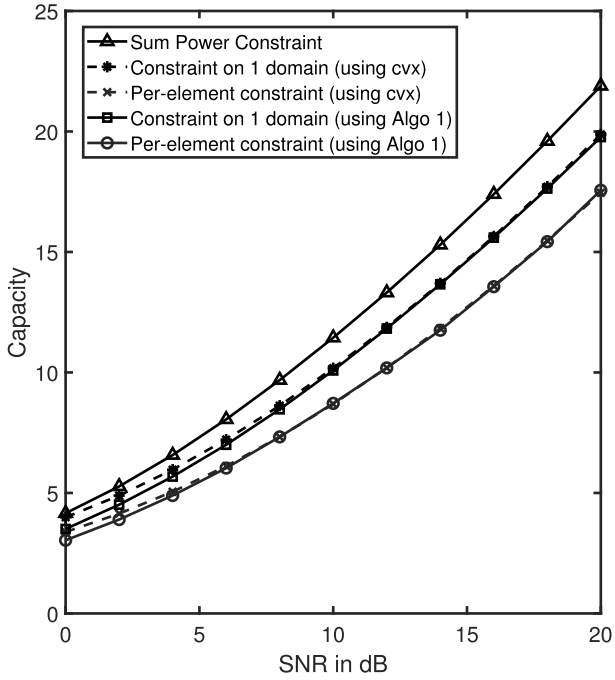

FIGURE 9. Capacity [bits/tensor-symbol] for different constraints vs SNR using Algorithm 1 and CVX.

where the entries of $\mathcal{H}^{corr}$ are correlated. Let $\mathrm{vec}(\mathcal{H})$ be denoted as $\underline{\mathbf{h}}$, then using the property of mode-n product from [93], [94, Lemma 2.1], we can write (91) using Kronecker product denoted by $\otimes$ as:

$$\mathrm{vec}(\mathcal{H}^{corr}) = (\mathrm{A}^{(N)} \otimes \cdots \otimes \mathrm{A}^{(1)})\underline{\mathbf{h}}. \tag{92}$$

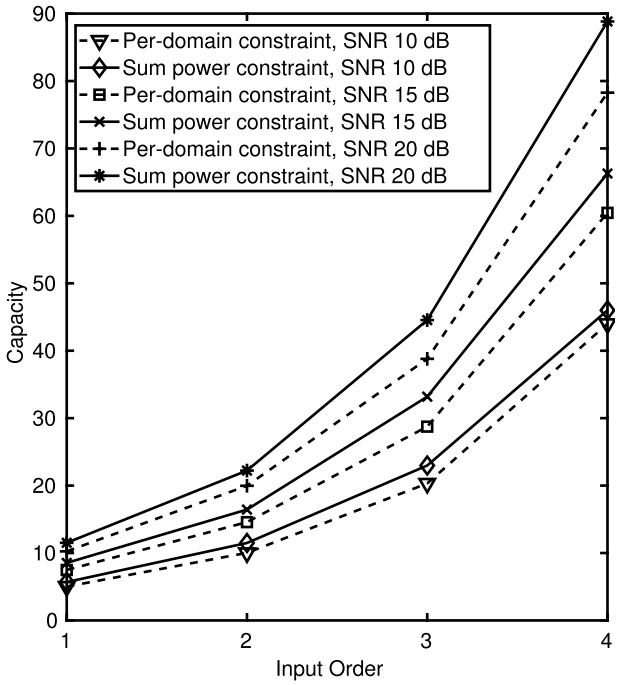

FIGURE 10. Capacity [bits/tensor-symbol] vs input order comparing sum power and per domain element power constraints with $P_1/P = 0.1$.

Then the correlation matrix of the vectorized channel is given as :

$$\mathbb{E}[\text{vec}(\mathcal{H}^{corr})\,\text{vec}(\mathcal{H}^{corr})^H]$$
$$= \mathbb{E}[(\mathrm{A}^{(N)} \otimes \cdots \otimes \mathrm{A}^{(1)})\underline{\mathbf{h}} \cdot \underline{\mathbf{h}}^H (\mathrm{A}^{(N)} \otimes \cdots \otimes \mathrm{A}^{(1)})^H] \quad (93)$$
$$= (\mathrm{A}^{(N)} \otimes \cdots \otimes \mathrm{A}^{(1)})\mathbb{E}[\underline{\mathbf{h}} \cdot \underline{\mathbf{h}}^H](\mathrm{A}^{(N)} \otimes \cdots \otimes \mathrm{A}^{(1)})^H \quad (94)$$
$$= (\mathrm{A}^{(N)} \otimes \cdots \otimes \mathrm{A}^{(1)})(\mathrm{A}^{(N)} \otimes \cdots \otimes \mathrm{A}^{(1)})^H \quad (95)$$
$$= (\mathrm{A}^{(N)}\mathrm{A}^{(N)H} \otimes \cdots \otimes \mathrm{A}^{(1)}\mathrm{A}^{(1)H}) \quad (96)$$
$$= \Psi^{(N)} \otimes \cdots \otimes \Psi^{(1)} \quad (97)$$

where (95) to (96) follow from matrix Kronecker product properties [95, Corollary 4]. Hence the correlation matrix of the vectorized tensor is given in terms of the Kronecker product of different mode-n factor correlation matrices denoted by $\Psi^{(n)}$. Such a model is called separable and it is considered for real random variables in [96]. While (97) expresses the correlation as a matrix by vectorizing $\mathcal{H}^{corr}$, it is shown in [96, Proposition 2.1] that the correlation of $\mathcal{H}^{corr}$ from (91) can also be expressed as an order $2N$ tensor obtained via the outer product of the factor matrices $\Psi^{(n)}$ defined as $\bar{\mathcal{R}} = \Psi^{(1)} \circ \cdots \circ \Psi^{(N)}$. Note that the correlation tensor when defined as $\mathcal{R} = \mathbb{E}[\mathcal{H}^{corr} \circ \mathcal{H}^{corr*}]$ is just a permuted version of $\bar{\mathcal{R}}$, where $\mathcal{R}_{i_1,\ldots,i_N,i'_1,\ldots,i'_N} = \bar{\mathcal{R}}_{i_1,i'_1,\ldots,i_N,i'_N} = \mathbb{E}[\mathcal{H}^{corr}_{i_1,\ldots,i_N} \cdot \mathcal{H}^{corr*}_{i'_1,\ldots,i'_N}] = \Psi^{(1)}_{i_1,i'_1} \cdots \Psi^{(N)}_{i_N,i'_N}$. Hence the separable model implies that each element in the correlation tensor can be written in terms of the product of the elements of the factor matrices. The proof in [96] is for real tensors, but can be easily extended to complex tensors as well.

The well known MIMO matrix Kronecker correlation model forms a special case of (91) where the tensor $\mathcal{H}$ is order-2 and the factor matrices $\mathrm{A}^{(1)}$ and $\mathrm{A}^{(2)}$ represent the square root of row and column correlation matrices respectively [97]. The MIMO Kronecker model may not be very accurate in all scenarios, but has been widely used because of its tractable analytic form, see for example [98]–[101].

Now consider an order-4 tensor channel of size $3 \times 3 \times 3 \times 3$ corresponding to an order-2 input and order-2 output. We generate such a channel $\mathcal{H}^{corr}$ with correlated elements using (91), where the entries of $\mathcal{H}$ are i.i.d zero mean complex Gaussian with unit variance. For numerical illustration, we consider the correlation matrices generated using the exponential model with different correlation factor ρ_n where the elements of the correlation matrix $\Psi^{(n)}$ are defined as $\Psi^{(n)}_{i,j} = \rho_n^{|i-j|}$ for $\rho_n \in [0, 1)$ [102]. The four correlation matrices $\Psi^{(1)}$, $\Psi^{(2)}$, $\Psi^{(3)}$ and $\Psi^{(4)}$ are generated using the exponential model with correlation coefficients ρ_1, ρ_2, ρ_3 and ρ_4 respectively. Assuming that the channel realization is known at the transmitter and the receiver, we find the capacity of such a channel with correlated elements under sum power constraint. Figure 11 presents capacity at 10 dB SNR for different values of correlation coefficients where the receive domains are correlated with $\rho_1 = \rho_2 = \rho_R$ and the transmit domains are correlated with $\rho_3 = \rho_4 = \rho_T$. The plot shows that the capacity decreases with increase in ρ_T and ρ_R, and it is least when ρ_T and ρ_R approach 1. Capacity is largest when both ρ_T and ρ_R approach zero in which case the correlation matrices are identity and the channel has only uncorrelated elements across all the domains.

Next we investigate the impact of correlation on the tensor channel capacity when the correlation spans over a variable number of domains. Figure 12 presents the capacity against SNR for different number of domains having correlated entries. Capacity is lowest when ρ_n is non-zero (0.7 in the figure) for all the domains (i.e. for $n = 1, 2, 3, 4$) and is highest when ρ_n is zero for all n, i.e. all entries are uncorrelated. Further it can be observed that the capacity difference between the various cases presented in Figure 12 is more significant at higher SNR. It is seen that the capacity decreases with increase in the number of domains having non-zero correlation factor. Such a loss of capacity with increase in the domains having correlation is further illustrated in Figure 13 for various tensor channel order.

Figure 13 presents the capacity against SNR for order $2M$ correlated tensor channels with order M input and output having individual dimensions of 3. The factor correlation matrices along each of the $2M$ modes are based on the exponential model with correlation factor ρ. The plot is presented for $M = 2, 3, 4, 5$ and $\rho = 0.4, 0.7$. Note that different values of M lead to different tensor channel sizes with order 4, 6, 8, and 10. Hence for a meaningful comparison of the impact of correlation, the capacity plotted in Figure 13 is

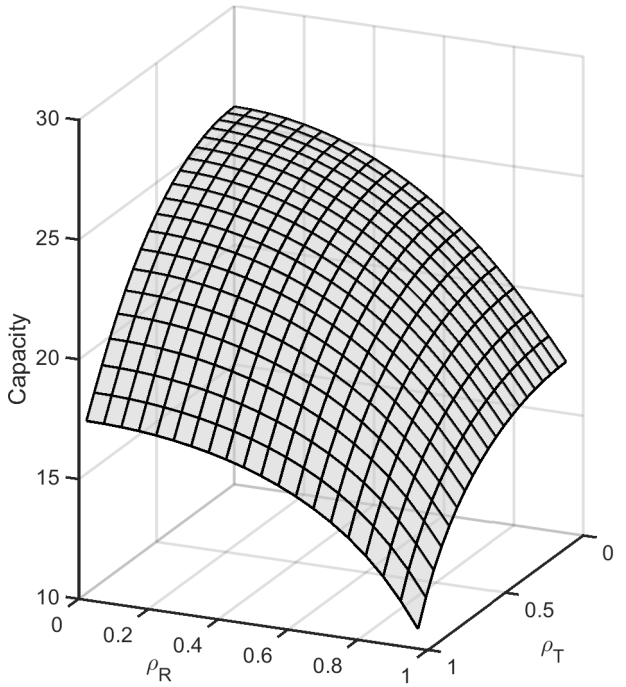

FIGURE 11. Capacity [bits/tensor-symbol] vs correlation coefficients for tensor channel with correlated entries.

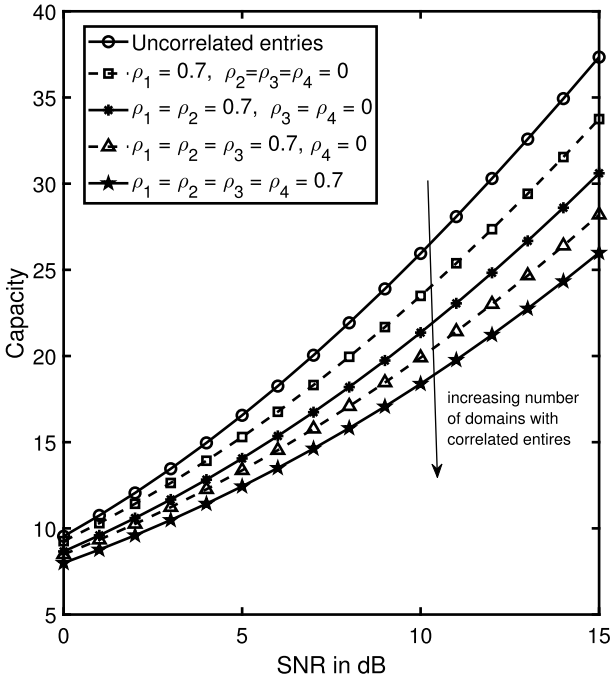

FIGURE 12. Capacity [bits/tensor-symbol] vs SNR for tensor channel with correlated entries.

normalized with respect to the number of elements in the transmit tensor symbol which is 3^M. As can be seen in Figure 13, for a fixed ρ, as the number of domains of the tensor channel with all correlated elements increases, the capacity per element decreases. Further, this loss is more significant for higher values of ρ (increased correlation) as can be seen by comparing the plots for $\rho = 0.7$ and 0.4.

B. TENSOR BASED MIMO GFDM SYSTEM

Generalized Frequency Division Multiplexing (GFDM) is a multi-carrier modulation scheme where each GFDM symbol consists of complex valued data symbols $d_{k,m}$ distributed across K sub-carriers and M time-slots known as sub-symbols. Thus each GFDM block consists of $N = KM$ complex data symbols. Consider a data vector $\underline{\mathbf{d}} \in \mathbb{C}^{N\times 1}$ which contains the complex data symbols $d_{k,m}$. The transmitted signal $\underline{\mathbf{x}} \in \mathbb{C}^{N\times 1}$ is produced as $\underline{\mathbf{x}} = \mathrm{A}\underline{\mathbf{d}}$, where $\mathrm{A} \in \mathbb{C}^{KM\times KM}$ is the GFDM modulation matrix. The modulator matrix is defined as $\mathrm{A} = (\underline{\mathrm{g}}_{0,0}, \ldots, \underline{\mathrm{g}}_{K-1,0}, \underline{\mathrm{g}}_{0,1}, \ldots, \underline{\mathrm{g}}_{0,M-1}, \ldots, \underline{\mathrm{g}}_{K-1,M-1})$, where $\underline{\mathrm{g}}_{k,m} = (g_{k,m}[0], \ldots, g_{k,m}[N-1])^T \in \mathbb{C}^{N\times 1}$ with $g_{k,m}[n]$ denoting the shifted version of the prototype filter impulse response $g[n]$ to the mth sub-symbol modulated on the kth sub-carrier. A detailed description of the modulator structure can be found in [103]. After cyclic prefix addition to ensure no inter-block interference, the transmission through a wireless channel is modelled as $\underline{\mathbf{y}} = \mathrm{H}\underline{\mathbf{x}} + \underline{\mathbf{n}}$ where $\mathrm{H} \in \mathbb{C}^{N\times N}$ is the circular channel convolution matrix and $\underline{\mathbf{y}}, \underline{\mathbf{n}} \in \mathbb{C}^{N\times 1}$ represent the received signal and noise vectors respectively [103]. The extension to a MIMO GFDM system

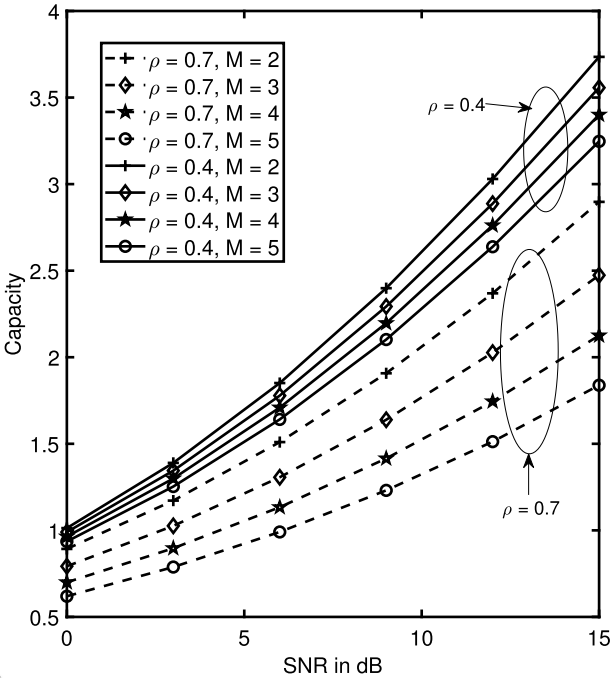

FIGURE 13. Capacity [bits/tensor element] vs SNR for different order tensor channels with correlated entries.

is considered in [24], [25] where the data vectors for different antennas are concatenated together. Hence the data corresponding to all the sub-carriers, sub-symbols and antennas for the transmitter and the receiver are arranged in transmit and receive vectors, and the channel is arranged as a matrix

to represent the system model. However, a more natural way of representing the signals and the channel in GFDM is using tensors since such a model can retain the distinction between the different domains namely sub-carriers, sub-symbols and antennas at each stage. Such a model has been considered in [37], [45].

Assume a MIMO GFDM system with N_R receive and N_T transmit antennas. Let a vector of $N = KM$ complex data symbols be considered as a single GFDM data stream. Consider a case where S such independent streams of complex data symbols denoted by $\underline{\mathbf{d}}^{(s)} \in \mathbb{C}^{N\times 1}$ for $s = 1, \ldots, S$ are each transmitted using K sub-carriers and M sub-symbols. Assume that $S = N_T = N_R$ and each of the S independent streams of data is transmitted concurrently using a different antenna among the N_T transmit antennas. For instance, if we have $S = N_T = 2$, then data stream 1 is mapped to antenna 1 and data stream 2 is mapped to antenna 2. Let the transmit data symbol corresponding to kth sub-carrier, mth sub-symbol and sth transmit stream represented as $d_{s,k,m}$, be an element of a third order tensor $\mathcal{D} \in \mathbb{C}^{S\times K\times M}$. Similarly the received signal and noise tensors can be represented using third order tensors $\tilde{\mathcal{D}} \in \mathbb{C}^{S\times K\times M}$ and $\mathcal{N} \in \mathbb{C}^{S\times K\times M}$ respectively. Subsequently the channel that couples the input $\mathcal{D}$ with the output $\tilde{\mathcal{D}}$ can be seen as an order-6 tensor $\mathcal{H} \in \mathbb{C}^{S\times K\times M\times S\times K\times M}$. The tensor channel considered here is the equivalent channel obtained from the cascading of the transmit filter, physical channel and the receive filter. More details on the representation of the signal and the channel model can be found in [37], [45]. The overall system model is represented as

$$\tilde{\mathcal{D}} = \mathcal{H} *_3 \mathcal{D} + \mathcal{N} \tag{98}$$

and it is illustrated in Figure 14 which represents the system model for a MIMO GFDM system with 2 transmit and 2 receive antennas. In Figure 14 a matrix is shown as a square and a higher order tensor as a double-line square with its order written on top right corner. A third order tensor is represented as a cube with staggered edges. The data corresponding to each antenna for all the K sub-carriers and M sub-symbols is represented as a $K \times M$ matrix, where the matrices corresponding to each antenna form slices of the third order input tensor. The covariance of the input tensor $\mathcal{D}$ can be written as a sixth order tensor $\mathcal{Q}$ of size $S \times K \times M \times S \times K \times M$. We simulate such a system with two transmit and two receive antennas, considering three different cases:

case 1: the entire transmit tensor is under sum power constraint P, i.e. $\mathrm{tr}(\mathcal{Q}) \le P$. We obtain the optimal covariance $\mathcal{Q}^{(opt1)}$ in this case which performs power allocation and precoding at the transmitter using: $\mathcal{Q}^{(opt1)} = \mathcal{V}_{\mathcal{H}} *_3 \left(\mu^{-1}\mathcal{I} - \mathcal{D}_{\mathcal{H}}^{-1}\right)^{+} *_3 \mathcal{V}_{\mathcal{H}}^{H}$, where $\mathcal{V}_{\mathcal{H}}$ and $\mathcal{D}_{\mathcal{H}}$ are obtained from the tensor EVD of $(\mathcal{H}^H *_3 \mathcal{H})$, and $\mathcal{I}$ is an identity tensor of size $S \times K \times M \times S \times K \times M$. The coefficient μ is calculated using tensor water-filling to ensure $\mathrm{tr}(\mathcal{Q}^{(opt1)}) = P$.

case 2: the two transmit antennas have different power budgets, $P_1 = x \cdot P$ and $P_2 = (1-x) \cdot P$. Hence the constraint on the covariance can be written as $\sum_{k,m} \mathcal{Q}_{1,k,m,1,k,m} \le P_1$ and $\sum_{k,m} \mathcal{Q}_{2,k,m,2,k,m} \le P_2$. We obtain the input covariance $\mathcal{Q}^{(opt2)}$ in this case using Algorithm 1 with the channel tensor $\mathcal{H}$, power budgets P_1 and P_2, and $N_{i_r} = K \cdot M$ as inputs.

case 3: the constraint is same as case 2, but now we perform per antenna power allocation and precoding. The optimum covariance in this case is determined by independently calculating the covariance for the data transmitted on the two antennas. We find $\mathcal{Q}^{(opt3)}_{1,:,:,1,:,:} = \mathcal{Q}^{(1)}$ and $\mathcal{Q}^{(opt3)}_{2,:,:,2,:,:} = \mathcal{Q}^{(2)}$, where $\mathcal{Q}^{(1)}$ and $\mathcal{Q}^{(2)}$ are order-4 tensors of size $K \times M \times K \times M$ representing the covariance of the matrix data transmitted on antenna 1 and 2 respectively. Similarly channel sub-tensors are defined as $\mathcal{H}^{(1)} = \mathcal{H}_{1,:,:,1,:,:}$ and $\mathcal{H}^{(2)} = \mathcal{H}_{2,:,:,2,:,:}$. Notice that the constraints $\sum_{k,m} \mathcal{Q}_{1,k,m,1,k,m} \le P_1$ and $\sum_{k,m} \mathcal{Q}_{2,k,m,2,k,m} \le P_2$ can be equivalently written as $\mathrm{tr}(\mathcal{Q}^{(1)}) \le P_1$ and $\mathrm{tr}(\mathcal{Q}^{(2)}) \le P_2$ respectively. We find the optimal $\mathcal{Q}^{(i)}$ for $i = 1, 2$ using tensor water-filling as $\mathcal{Q}^{(i)} = \mathcal{V}^{(i)} *_2 \left(\mu_i^{-1}\mathcal{I} - \mathcal{D}^{(i)^{-1}}\right)^{+} *_2 \mathcal{V}^{(i)^H}$, where $\mathcal{V}^{(i)}$ and $\mathcal{D}^{(i)}$ are obtained from the tensor EVD of $(\mathcal{H}^{(i)^H} *_2 \mathcal{H}^{(i)})$. The coefficient μ_i is calculated using the tensor water-filling to ensure $\mathrm{tr}(\mathcal{Q}^{(i)}) \le P_i$.

For the simulation, the input tensor $\mathcal{D}$ contains entries drawn from a 4QAM constellation. At the transmitter, a raised cosine transmit pulse shaping filter with roll off factor 1 is employed as used in [25], [37]. The receive filter is matched to the transmit filter. The number of sub-symbols used is $M = 5$ and the number of sub-carriers used is $K = 8$. The channel between transmit and receive filter is generated as complex Gaussian with i.i.d. zero mean and unit variance entries. The overall channel is normalized to ensure that the received power is the same as the transmit power P. Hence the received SNR per bit is calculated as P/b where b is the number of bits in each received tensor symbol, obtained as $b =$ (Number of elements in each tensor) $\times$ (bits per element). The noise added is AWGN with zero mean and unit variance such that the noise covariance is the identity tensor.

The capacity of the channel $\mathcal{H}$ from (98) under three different cases of input covariance with $x = 0.1$ for case 2 and 3 is shown in Figure 15. As can be seen the capacity achieved for case 1, i.e. sum power constraint is always higher than the other two cases. For case 2, where individual antennas have different power budgets and joint domain processing is used to generate the optimal covariance outperforms case 3 where separate covariances are obtained for antenna 1 and 2 independently.

We also test the BER performance when the input is generated from a 4QAM constellation for the three different cases and the results are shown in Figure 16 where BER is plotted against received SNR per bit in dB. At the transmitter, the covariance of the input $\mathcal{D} \in \mathbb{C}^{2\times 8\times 5}$ before precoding is given by an identity tensor. The input $\mathcal{D}$ is contracted with the square root of the optimum covariance as $\mathcal{Q}^{(opt)^{1/2}} *_3 \mathcal{D}$. This ensures that the covariance of the input tensor after

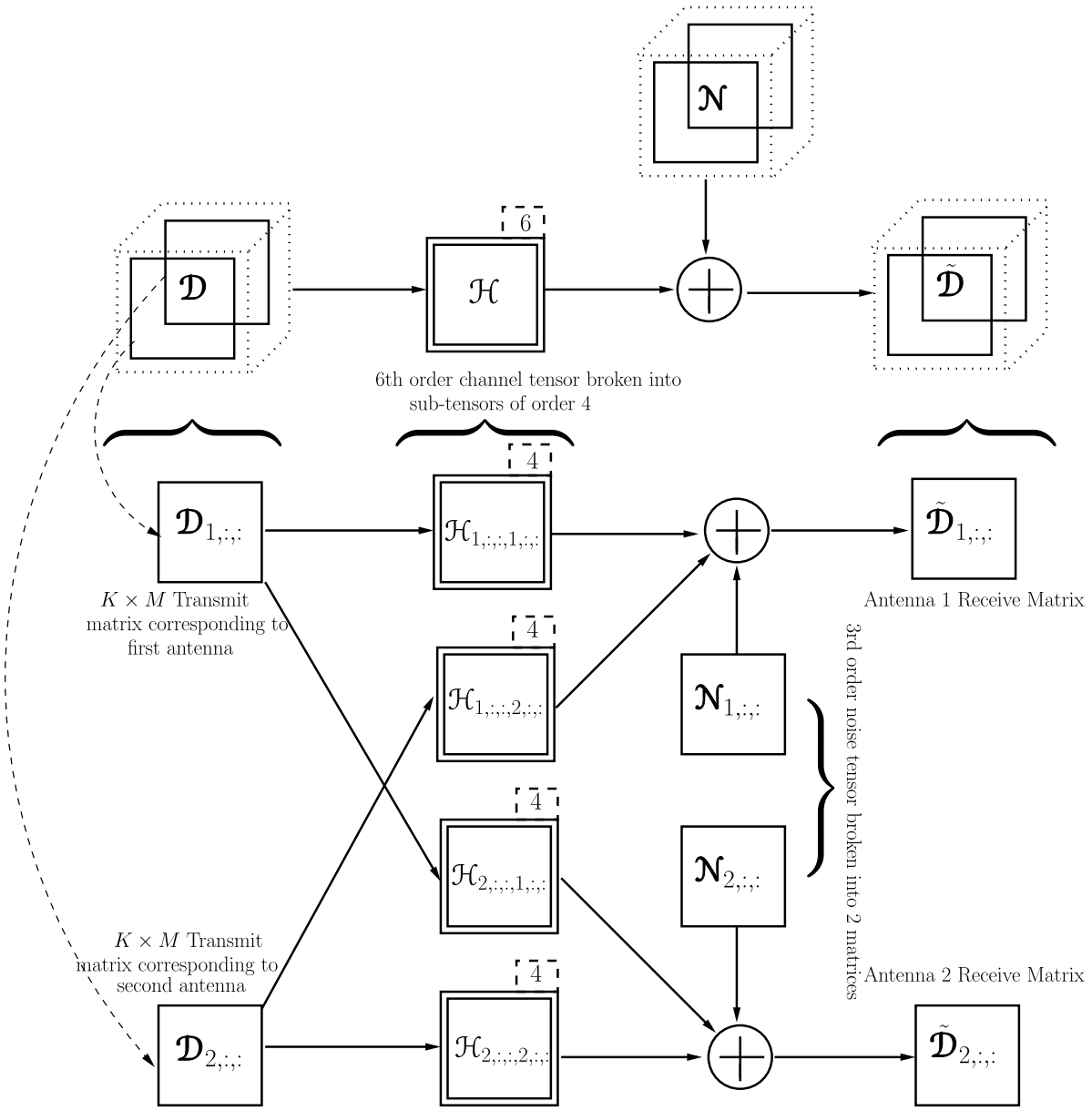

FIGURE 14. Tensor system model for MIMO GFDM with 2 antennas ($N_T = N_R = 2$).

precoding, defined as $\mathbb{E}[(\mathcal{Q}^{(opt)^{1/2}} *_3 \mathcal{D}) \circ (\mathcal{Q}^{(opt)^{1/2}} *_3 \mathcal{D})^*]$ is given by $\mathcal{Q}^{(opt)} \in \mathbb{C}^{2\times8\times5\times2\times8\times5}$. Hence, based on (98), we get $\tilde{\mathcal{D}} = \mathcal{H} *_3 \mathcal{Q}^{(opt)^{1/2}} *_3 \mathcal{D} + \mathcal{N}$. To assess the BER performance, we assume a multi-linear minimum mean square error (MMSE) receiver [45]. At the receiver, $\tilde{\mathcal{D}}$ from (98) is passed through a multi-linear MMSE filter, $\mathcal{G}_{MMSE}$ to obtain an estimate $\hat{\mathcal{D}} = \mathcal{G}_{MMSE} *_3 \tilde{\mathcal{D}}$ of the transmitted tensor $\mathcal{D}$. The filter $\mathcal{G}_{MMSE}$ is chosen such that $\mathbb{E}[||\mathcal{D} - \hat{\mathcal{D}}||^2]$ is minimized, and is given by [14]:

$$\mathcal{G}_{MMSE} = \mathcal{W}^H *_3 (\mathcal{W} *_3 \mathcal{W}^H + \mathcal{I})^{-1} \tag{99}$$

where $\mathcal{W} = \mathcal{H} *_3 \mathcal{Q}^{(opt)^{1/2}} \in \mathbb{C}^{2\times8\times5\times2\times8\times5}$ and $\mathcal{I}$ is an identity tensor of size $2 \times 8 \times 5 \times 2 \times 8 \times 5$. The estimate is passed through a QAM demodulator to recover the transmitted symbols. Figure 16 presents results based on Monte-Carlo simulations with averaging over 100 different channel realizations. For evaluating bit error rate (BER), at least 100 errors were collected at the receiver for each channel realization. It can be seen that the best performance is achieved in case 1, where there is no individual power constraint on domains. A comparison between cases 2 and 3 shows that with per antenna power constraints, joint processing across the domains (as in case 2) performs better than the per antenna processing (as in case 3).

VII. MULTI-USER MIMO CAPACITY

In this section, we consider the application of the tensor framework to multi-user MIMO channels. We present

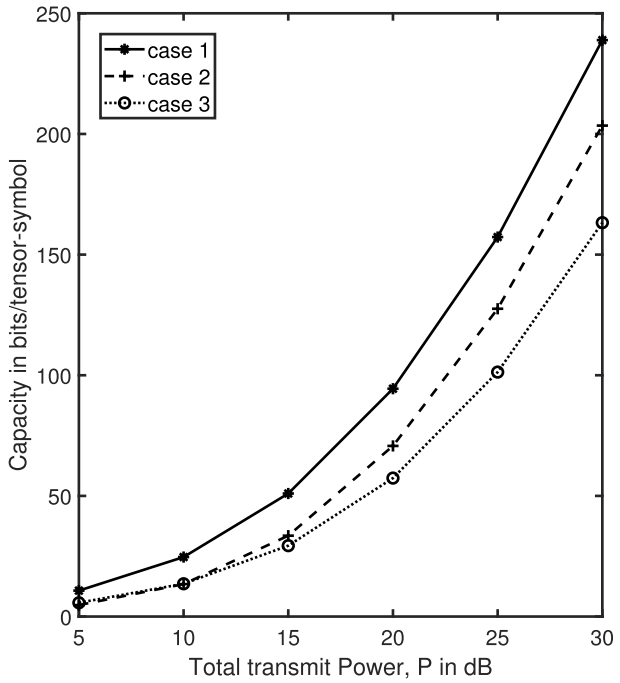

FIGURE 15. Capacity [bits/tensor-symbol] vs transmit power for MIMO GFDM under sum power constraint, and per antenna power constraints with and without joint domain processing.

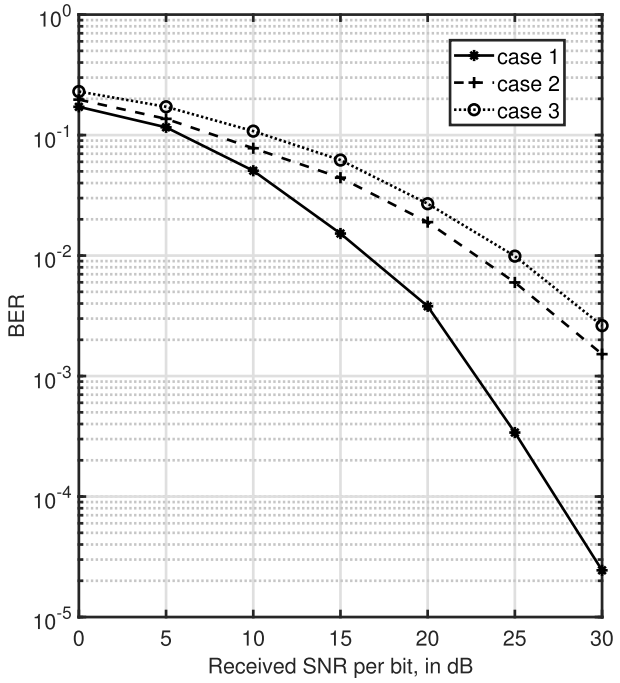

FIGURE 16. BER plot for MIMO GFDM under sum power constraint, and per antenna power constraints with and without joint domain processing.

numerical examples comparing the tensor approach with other results found in literature. We consider K-user Gaussian multiple access channels and interference channels.

A. MULTIPLE ACCESS CHANNELS

Consider a multi-user MIMO network where a base station (BS) equipped with N_R antennas is receiving information from K users with N_T antennas each. If we denote the uplink channel matrix between the kth user and the base station by $\mathrm{H}^{(k)} \in \mathbb{C}^{N_R \times N_T}$, then the discrete time received signal $\underline{\mathrm{y}} \in \mathbb{C}^{N_R \times 1}$ at the BS can be written as [42]:

$$\underline{\mathrm{y}} = \sum_{k=1}^{K} \mathrm{H}^{(k)} \underline{\mathrm{x}}^{(k)} + \underline{\mathrm{n}} \tag{100}$$

where $\underline{\mathrm{x}}^{(k)} \in \mathbb{C}^{N_T \times 1}$ is the signal transmitted by the kth user and $\underline{\mathrm{n}} \in \mathbb{C}^{N_R \times 1}$ is the received noise vector which is assumed circularly symmetric complex Gaussian with identity covariance matrix. In such a multiple access channel (MAC), each user k is subject to an individual power constraint P_k. If the transmit covariance matrix of user k is denoted as $\mathrm{Q}^{(k)}$, then the constraint is represented as $\mathrm{tr}(\mathrm{Q}^{(k)}) \leq P_k$ for $k = 1, \ldots, K$.

A multi-user channel with K users is characterized by a K-dimensional achievable rate region $\mathcal{C}_{\mathcal{R}}$, known as capacity region [42], where each point in the region $(R_1, R_2, \ldots, R_K)$ represents the achievable rates R_k at which user k can send information with arbitrarily low error probability. We assume that all the channel matrices are known to the receiver and all the transmitters. To denote the convex hull of the union of sets we use the symbol $\bar{\bigcup}$. With power constraints $(P_1, P_2, \ldots, P_K)$, the capacity region of MIMO MAC is given by [42]:

$$\begin{aligned} \mathcal{C}_{\mathcal{R}} = \bar{\bigcup}_{\mathrm{tr}(\mathrm{Q}^{(k)}) \leq P_k, \forall k} \Big\{ & (R_1, \ldots, R_K) : \\ & 0 \leq \sum_{k \in \mathcal{S}} R_k \leq \log\det\Big(\mathrm{I} + \sum_{k \in \mathcal{S}} \mathrm{H}^{(k)} \mathrm{Q}^{(k)} \mathrm{H}^{(k)H}\Big) \\ & \forall \mathcal{S} \subseteq \{1, \ldots, K\} \Big\} \end{aligned} \tag{101}$$

where $\mathcal{S}$ denotes a subset of the set of users. Each set of covariance matrices $(\mathrm{Q}^{(1)}, \ldots, \mathrm{Q}^{(K)})$ satisfying the power constraints corresponds to a K-dimensional polyhedron [104]. The capacity region is the convex hull of the union of all such polyhedrons. Note that for Gaussian MIMO MAC, the capacity region is defined using only the union of rate regions and the convex hull is not needed [42]. It is shown in [105] that for Gaussian MIMO MAC, the boundary points of the capacity region can be characterized by maximizing a weighted sum rate $\sum_k \nu_k R_k$ for all non-negative ν_k such that $\sum_k \nu_k = 1$, and thus finding boundary points can be cast into a convex optimization problem.

Essentially, (101) represents a set of bounds on individual rates $R_1, R_2, \ldots, R_K$, and combination of rates such as $R_1 + R_2, R_1 + R_3, R_2 + R_3, R_1 + R_2 + R_3$ and so on, including the sum rate $R_1 + R_2 + \cdots + R_K$. For a two users case, this can be represented by (102), as shown at the bottom of the next page. For a given choice of covariance matrices $\mathrm{Q}^{(1)}$ and $\mathrm{Q}^{(2)}$ which satisfy the power constraints, (102) represents an upper bound on R_1, R_2 and

$R_1 + R_2$. The maximum of $\log\det(\mathrm{I} + \mathrm{H}^{(1)}\mathrm{Q}^{(1)}\mathrm{H}^{(1)H})$ and $\log\det(\mathrm{I} + \mathrm{H}^{(2)}\mathrm{Q}^{(2)}\mathrm{H}^{(2)H})$ from(102) are the individual achievable capacities by user 1 and user 2 respectively assuming the other user is silent. The maximum of $\log\det(\mathrm{I}+\mathrm{H}^{(1)}\mathrm{Q}^{(1)}\mathrm{H}^{(1)H}+\mathrm{H}^{(2)}\mathrm{Q}^{(2)}\mathrm{H}^{(2)H})$ is the sum capacity achievable when both users are transmitting. Note that the choice of covariance matrices $\mathrm{Q}^{(1)}$ and $\mathrm{Q}^{(2)}$ satisfying the power constraints which achieves the sum capacity may not achieve the individual capacities. Similarly the choice of $\mathrm{Q}^{(1)}$ and $\mathrm{Q}^{(2)}$ which achieves the individual capacities may not achieve the sum capacity. In general, different transmit choices lead to different pairs of covariance matrices $(\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)})$ such that the capacity region is given by the convex hull of the union of an infinite number of rate regions each corresponding to a different set $(\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)})$. The optimal choice of $\mathrm{Q}^{(1)}$ and $\mathrm{Q}^{(2)}$ which achieves the sum capacity is found by an iterative water-filling approach [105] which sequentially finds covariance for each user with the single-user classical water-filling method assuming interference from other users as noise. A detailed step by step algorithm can be found in [105].

Such an approach, however, assumes that different users transmit independently. The iterative water-filling algorithm treats the information available about other users' interfering channels as noise. In the presence of complete channel state information, a better transmit strategy which would provide higher achievable rates would be to allow users to coordinate for transmission. Hence a joint signal transmit strategy can expand the capacity region. Using the tensor framework, we can find the capacity region assuming user coordination. First, let us represent (100) using the tensor system model. We define the multi-user MIMO tensor channel as a third order tensor $\mathcal{H} \in \mathbb{C}^{N_R \times N_T \times K}$ where $\mathcal{H}_{:,:,k} = \mathrm{H}^{(k)}$. The input signal can be defined as a matrix $\mathbf{X} \in \mathbb{C}^{N_T \times K}$, where each $\underline{\mathbf{x}}^{(k)}$ of (100) forms a column of the matrix $\mathbf{X}$. Hence the system model in (100) can be represented as:

$$\underline{\mathbf{y}} = \mathcal{H} *_2 \mathbf{X} + \underline{\mathbf{n}} \tag{103}$$

The input covariance is represented as an order-4 tensor $\mathcal{Q} \in \mathbb{C}^{N_T \times K \times N_T \times K}$. Assuming the noise to be circularly symmetric complex Gaussian with identity covariance matrix, denoted by I, the output covariance can be written as $(\mathcal{H} *_2 \mathcal{Q} *_2 \mathcal{H}^H + \mathrm{I})$. Subsequently, the sum capacity of such a system with user coordination can be calculated from the following optimization problem:

$$\max_{\mathcal{Q}} \log\left[\det\left(\mathcal{H} *_2 \mathcal{Q} *_2 \mathcal{H}^H + \mathrm{I}\right)\right] \tag{104}$$

$$s.t. \sum_{n=1}^{N_T} \mathcal{Q}_{n,k,n,k} \le P_k \quad \forall k, \tag{105}$$

$$\mathcal{Q} \succeq 0. \tag{106}$$

where $\sum_{n=1}^{N_T} \mathcal{Q}_{n,k,n,k} \le P_k$ represents the individual power constraints for different users. The optimal $\mathcal{Q}$ that achieves capacity can be approximated using Algorithm 1. Note the difference in this tensor formulation and the iterative water-filling approach used in vector formulation is that the latter assumes different users transmit independently despite having perfect channel state information. With independent transmissions, the iterative water-filling maximizes the function $\log\det(\mathrm{I} + \sum_{k=1}^{K} \mathrm{H}^{(k)}\mathrm{Q}^{(k)}\mathrm{H}^{(k)H})$ subject to $\mathrm{tr}(\mathrm{Q}^{(k)}) \le P_k$ and $\mathrm{Q}^{(k)} \succeq 0$ for $k = 1, \ldots, K$ [105]. Note that this objective function is same as the upper bound on the sum rate from (101) for $S = \{1, \ldots, K\}$. The vector based iterative water-filling treats inter-user interference as noise since it attempts to optimize the sum rate over all choices of separate covariance matrix for each user as shown in [105]. On the other hand, the tensor approach solves the problem in (104)-(106) and aims to find a joint covariance across all the users. Thereby, the tensor approach suggests a joint transmit scheme for all the users wherein the inter-user interference is not treated as noise since the interference term also carries signal information. The maximum sum rate achieved by all the K users given in (104) is the sum capacity of the K users MIMO MAC under user coordination. Similarly the sum capacity achieved by a subset $\mathcal{S}$ of the all the users $\mathcal{U} = \{1, \ldots, K\}$ is given as the maximum of $\log\det(\mathrm{I} + \mathcal{H}^{(\mathcal{S})} *_2 \mathcal{Q}^{(\mathcal{S})} *_2 \mathcal{H}^{(\mathcal{S})H})$ over the choice of positive semi-definite $\mathcal{Q}^{(\mathcal{S})}$ which satisfies the power constraints. The tensor $\mathcal{H}^{(\mathcal{S})} \in \mathbb{C}^{N_R \times N_T \times |\mathcal{S}|}$ with $|\mathcal{S}|$ denoting the cardinality of $\mathcal{S}$, contains matrices of size $N_R \times N_T$ as slices corresponding to only those users which are included in $\mathcal{S}$. Similarly $\mathcal{Q}^{(\mathcal{S})} \in \mathbb{C}^{N_T \times |\mathcal{S}| \times N_T \times |\mathcal{S}|}$ is the covariance tensor of $\mathbf{X}^{(\mathcal{S})} \in \mathbb{C}^{N_T \times |\mathcal{S}|}$ which contains columns of only those users which are included in $\mathcal{S}$.

Hence the tensor framework allows to define a capacity region with user coordination as:

$$\begin{aligned}\mathcal{C}_{\mathcal{R}} = \bar{\bigcup_{\sum_{n=1}^{N_T} \mathcal{Q}^{(\mathcal{S})}_{n,i,n,i} \le \underline{\mathrm{p}}_i^{(\mathcal{S})}, \forall i}} \Big\{ & (R_1, \ldots, R_K) : \\ 0 \le \sum_{k \in \mathcal{S}} R_k \le \log\det & \left(\mathrm{I} + \mathcal{H}^{(\mathcal{S})} *_2 \mathcal{Q}^{(\mathcal{S})} *_2 \mathcal{H}^{(\mathcal{S})H}\right) \\ & \forall \mathcal{S} \subseteq \{1, \ldots, K\} \Big\}\end{aligned} \tag{107}$$

where $\mathcal{S}$ contains the list of users being considered. The vector $\underline{\mathrm{p}}^{(\mathcal{S})}$ contains the power budgets of the users included

$$\mathcal{C}_{\mathcal{R}} = \bar{\bigcup_{\substack{\mathrm{tr}(\mathrm{Q}^{(1)}) \le P_1, \\ \mathrm{tr}(\mathrm{Q}^{(2)}) \le P_2}}} \left\{ \begin{array}{c} 0 \le R_1 \le \log\det\left(\mathrm{I} + \mathrm{H}^{(1)}\mathrm{Q}^{(1)}\mathrm{H}^{(1)H}\right) \\ 0 \le R_2 \le \log\det\left(\mathrm{I} + \mathrm{H}^{(2)}\mathrm{Q}^{(2)}\mathrm{H}^{(2)H}\right) \\ 0 \le R_1 + R_2 \le \log\det\left(\mathrm{I} + \mathrm{H}^{(1)}\mathrm{Q}^{(1)}\mathrm{H}^{(1)H} + \mathrm{H}^{(2)}\mathrm{Q}^{(2)}\mathrm{H}^{(2)H}\right) \end{array} \right\} \tag{102}$$

in $\mathcal{S}$, with its components $\mathrm{p}_i^{(\mathcal{S})}$ denoting the power budget of the ith user in set $\mathcal{S}$ for $i = 1, \ldots, |\mathcal{S}|$. Note that the expression $\log\det(\mathrm{I} + \mathcal{H}^{(\mathcal{S})} *_2 \mathcal{Q}^{(\mathcal{S})} *_2 \mathcal{H}^{(\mathcal{S})H})$ in (107) is same as the objective function in (104) when $\mathcal{S}$ contains all the users, i.e. $\mathcal{S} = \{1, \ldots, K\}$.

As an example, let us consider a three users scenario in which case $\mathcal{S}$ can assume the following sets:{1}, {2}, {3}, {1, 2}, {2, 3}, {1, 3}, {1, 2, 3}. Subsequently, we can expand (107) as in (108), as shown at the bottom of the page. For the first bound on R_1, we have $\mathcal{S} = \{1\}$, i.e. we consider only first user and find what is the maximum rate that user 1 can transmit given the power constraint on user 1 and that all other users are silent. In this case, $\mathcal{H}^{(1)} \in \mathbb{C}^{N_R \times N_T \times 1}$ is essentially the matrix $\mathrm{H}^{(1)}$ between the user 1 and base station, and $\mathcal{Q}^{(1)} \in \mathbb{C}^{N_T \times 1 \times N_T \times 1}$ is the covariance matrix $\mathrm{Q}^{(1)}$ of user 1. Hence this reduces to single user MIMO channel where the optimal $\mathrm{Q}^{(1)}$ which achieves maximum rate can be found using classical water-filling. Similarly, conditions two and three in (108) correspond to the bounds on the rates achieved by user 2 and user 3 (R_2 and R_3) respectively when all other users are silent. Hence, the first three conditions are same as the one derived from the vector case (101). Condition four corresponds to $\mathcal{S} = \{1, 2\}$, thus gives a bound on the sum rate that user 1 and user 2 can together achieve given that user 3 is silent. In this case $\mathcal{H}^{(1,2)} \in \mathbb{C}^{N_R \times N_T \times 2}$ is a sub-tensor of $\mathcal{H}$ as $\mathcal{H}^{(1,2)} = \mathcal{H}_{:,:,1:2}$. The covariance tensor $\mathcal{Q}^{(1,2)} \in \mathbb{C}^{N_T \times 2 \times N_T \times 2}$ is a sub-tensor of $\mathcal{Q}$ given by $\mathcal{Q}^{(1,2)} = \mathcal{Q}_{:,1:2,:,1:2}$. The power constraints are defined as $\sum_n \mathcal{Q}^{(1,2)}_{n,1,n,1} \leq P_1$ and $\sum_n \mathcal{Q}^{(1,2)}_{n,2,n,2} \leq P_2$ where P_1, P_2 are power budgets for user 1 and 2 respectively. Note that the bound on sum rate $R_1 + R_2$ achieved using this method assumes that user 1 and 2 perform a joint transmission. Hence the sum rate achieved using the tensor approach will be different than the one obtained from iterative water-filling which assumes independent transmission. Similarly, condition five represents the bound on sum rate $R_2 + R_3$ that can be achieved when user 2 and 3 transmit together keeping user 1 silent. In this case $\mathcal{H}^{(2,3)} \in \mathbb{C}^{N_R \times N_T \times 2}$ is a sub-tensor of $\mathcal{H}$ given by $\mathcal{H}^{(2,3)} = \mathcal{H}_{:,:,2:3}$. The covariance tensor $\mathcal{Q}^{(2,3)} \in \mathbb{C}^{N_T \times 2 \times N_T \times 2}$ is a sub-tensor of $\mathcal{Q}$ given by $\mathcal{Q}^{(2,3)} = \mathcal{Q}_{:,2:3,:,2:3}$. The power constraints are defined as $\sum_n \mathcal{Q}^{(2,3)}_{n,1,n,1} \leq P_2$ and $\sum_n \mathcal{Q}^{(2,3)}_{n,2,n,2} \leq P_3$ where P_2, P_3 are power budgets for user 2 and 3 respectively. Similarly condition six represents the bound on sum rate $R_1 + R_3$ that can be achieved when user 1 and 3 transmit together keeping user 2 silent. The last condition represents the bound on sum rate when all the three users are transmitting. Note that the covariance tensors satisfying the power constraints of all the equations in (108) can be seen as sub-tensors of the covariance tensor $\mathcal{Q}$. For instance, $\mathcal{Q}_{:,1,:,1} = \mathrm{Q}^{(1)}$ represents the covariance matrix of user 1. But the optimal choice of covariance tensor $\mathcal{Q}$ that achieves the sum capacity under user cooperation can not be obtained from only individual covariance tensors $\mathrm{Q}^{(i)}$ for different users. For instance, the optimal $\mathrm{Q}^{(1)}$ that maximizes $\log\det(\mathrm{I} + \mathrm{H}^{(1)}\mathrm{Q}^{(1)}\mathrm{H}^{(1)H})$ may not be the sub-tensor of the optimal $\mathcal{Q}$ that maximizes $\log\det(\mathrm{I} + \mathcal{H} *_2 \mathcal{Q} *_2 \mathcal{H}^H)$. The capacity region under user coordination is thus given by the convex hull of the union of all the rate regions over all the choices of covariance tensors which satisfy the power constraints.

With additional constraints on the covariance tensor

$$\mathcal{Q}^{(\mathcal{S})}_{n,i,n',i'} = 0, \quad \text{for } i \neq i', \tag{109}$$

we can use (107) to represent the capacity region without user coordination as in (101). To show this, we consider the expression $\mathcal{H}^{(\mathcal{S})} *_2 \mathcal{Q}^{(\mathcal{S})} *_2 \mathcal{H}^{(\mathcal{S})H}$ from (107) for any given $\mathcal{S}$ and show that with (109) it reduces to $\sum_{k\in\mathcal{S}} \mathrm{H}^{(k)}\mathrm{Q}^{(k)}\mathrm{H}^{(k)H}$ as in (101). We can write:

$$\begin{aligned}&(\mathcal{H}^{(\mathcal{S})} *_2 \mathcal{Q}^{(\mathcal{S})} *_2 \mathcal{H}^{(\mathcal{S})H})_{j,j'} \\ &\quad= \sum_{n',i'}\Big(\sum_{n,i} \mathcal{H}^{(\mathcal{S})}_{j,n,i}\mathcal{Q}^{(\mathcal{S})}_{n,i,n',i'}\Big)\mathcal{H}^{(\mathcal{S})H}_{n',i',j'}\end{aligned} \tag{110}$$

With (109), entries of $\mathcal{Q}^{(\mathcal{S})}$ are zero for all $i \neq i'$. Hence we can write (110) as

$$\begin{aligned}&(\mathcal{H}^{(\mathcal{S})} *_2 \mathcal{Q}^{(\mathcal{S})} *_2 \mathcal{H}^{(\mathcal{S})H})_{j,j'} \\ &\quad= \sum_{i=1}^{|\mathcal{S}|}\Bigg(\sum_{n'=1}^{N_T}\Big(\sum_{n=1}^{N_T} \mathcal{H}^{(\mathcal{S})}_{j,n,i}\mathcal{Q}^{(\mathcal{S})}_{n,i,n',i}\Big)\mathcal{H}^{(\mathcal{S})H}_{n',i,j'}\Bigg)\end{aligned} \tag{111}$$

Note that each user in the set of all users $\mathcal{U} = \{1, \ldots, K\}$ is known by its index k, i.e. first user, second user, kth user and so on. The variable i denotes the index of a user in set $\mathcal{S}$ where $\mathcal{S} \subseteq \mathcal{U}$. For instance if $\mathcal{S} = \{3, 4, \ldots, K\}$, then the third user ($k = 3$) is at index $i = 1$ in $\mathcal{S}$. Hence we replace the

$$\mathcal{C}_{\mathcal{R}} = \bigcup_{\substack{\mathrm{tr}(\mathrm{Q}^{(i)}) \leq P_i,\ i=1,2,3 \\ \sum_n \mathcal{Q}^{(1,2)}_{n,1,n,1} \leq P_1,\ \sum_n \mathcal{Q}^{(1,2)}_{n,2,n,2} \leq P_2 \\ \sum_n \mathcal{Q}^{(2,3)}_{n,1,n,1} \leq P_2,\ \sum_n \mathcal{Q}^{(2,3)}_{n,2,n,2} \leq P_3 \\ \sum_n \mathcal{Q}^{(1,3)}_{n,1,n,1} \leq P_1,\ \sum_n \mathcal{Q}^{(1,3)}_{n,2,n,2} \leq P_3 \\ \sum_n \mathcal{Q}_{n,k,n,k} \leq P_k,\ k=1,2,3}} \left\{ \begin{array}{c} 0 \leq R_1 \leq \log\det\left(\mathrm{I} + \mathrm{H}^{(1)}\mathrm{Q}^{(1)}\mathrm{H}^{(1)H}\right) \\ 0 \leq R_2 \leq \log\det\left(\mathrm{I} + \mathrm{H}^{(2)}\mathrm{Q}^{(2)}\mathrm{H}^{(2)H}\right) \\ 0 \leq R_3 \leq \log\det\left(\mathrm{I} + \mathrm{H}^{(3)}\mathrm{Q}^{(3)}\mathrm{H}^{(3)H}\right) \\ 0 \leq R_1 + R_2 \leq \log\det\left(\mathrm{I} + \mathcal{H}^{(1,2)} *_2 \mathcal{Q}^{(1,2)} *_2 \mathcal{H}^{(1,2)H}\right) \\ 0 \leq R_2 + R_3 \leq \log\det\left(\mathrm{I} + \mathcal{H}^{(2,3)} *_2 \mathcal{Q}^{(2,3)} *_2 \mathcal{H}^{(2,3)H}\right) \\ 0 \leq R_1 + R_3 \leq \log\det\left(\mathrm{I} + \mathcal{H}^{(1,3)} *_2 \mathcal{Q}^{(1,3)} *_2 \mathcal{H}^{(1,3)H}\right) \\ 0 \leq R_1 + R_2 + R_3 \leq \log\det\left(\mathrm{I} + \mathcal{H} *_2 \mathcal{Q} *_2 \mathcal{H}^H\right) \end{array} \right\} \tag{108}$$

index i with the user number k. The entities $\mathcal{H}^{(\mathcal{S})}_{:,:,i}$ and $\mathcal{Q}^{(\mathcal{S})}_{:,i,:,i}$ are the channel sub-tensor and covariance matrix of the user i in set $\mathcal{S}$, and equivalently of the kth user in set of all users. Hence we get $\mathcal{H}^{(\mathcal{S})}_{j,n,i} = \mathrm{H}^{(k)}_{j,n}$ for $k \in \mathcal{S}$, and the covariance as $\mathcal{Q}_{n,i,n',i} = \mathrm{Q}^{(k)}_{n,n'}$ for $k \in \mathcal{S}$. Thus we can write (111) as:

$$(\mathcal{H}^{(\mathcal{S})} *_2 \mathcal{Q}^{(\mathcal{S})} *_2 \mathcal{H}^{(\mathcal{S})H})_{j,j'} = \sum_{k\in\mathcal{S}} \Big(\sum_{n'=1}^{N_T} (\sum_{n=1}^{N_T} \mathrm{H}^{(k)}_{j,n} \mathrm{Q}^{(k)}_{n,n'}) \mathrm{H}^{(k)H}_{n,j'} \Big) \quad (112)$$

$$= \sum_{k\in\mathcal{S}} (\mathrm{H}^{(k)} \mathrm{Q}^{(k)} \mathrm{H}^{(k)H})_{j,j'} \quad (113)$$

Substituting (113) with the additional constraints (109) into (107) gives us the capacity region from (101) which assumes no user coordination.

Next we present a few numerical examples to illustrate the concepts. We consider a multi-user MIMO MAC scenario where K users with $N_T = 2$ antennas each are communicating with a base station having $N_R = 10$ antennas. The noise vector at the receiver is circularly symmetric complex Gaussian with zero mean and identity covariance matrix. The channel entries are realizations of circularly symmetric complex Gaussian random variables with zero mean and unit variance and these realizations are known at the transmitters and receiver. The results presented are averaged over 100 different channel realizations. Each user has an individual power constraint P_k. The total transmit power is $P = \sum_k P_k$. We assume all the users have the same power constraint, i.e. $P_k = P/K$ and plot the sum capacity obtained through the tensor approach achieved by K users against total power in Figure 17. The sum capacity can be found by solving the optimization problem in (104)-(106). We apply the proposed solution from Algorithm 1 to approximate the optimal input covariance tensor which achieves the sum capacity. The covariance tensor obtained from Algorithm 1 is then used to approximate the sum capacity given by $\log\det(\mathcal{H} *_2 \mathcal{Q} *_2 \mathcal{H}^H + \mathrm{I})$. This approach assumes that all the users coordinate for transmission. It can be seen that for a fixed number of users, the sum capacity increases with an increase in the total transmit power. Furthermore, for a fixed total transmit power, the sum capacity increases when the number of users increases. Especially at higher transmit powers, the increased number of users lead to a significant increase in the sum capacity.

The results of Figure 17 can be contrasted with the iterative water-filling approach from literature where different users despite having channel state information of other users transmit independently. Figure 18 presents the sum capacity obtained with coordinated users and independent users against the number of users for two different values of total power. The sum capacity under independent users is calculated using the iterative water-filling approach from [105, Algorithm 1]. It can be observed that as the number of users increases, there is a significant difference in achievable rate of coordinated users as compared to independent users. This shows that user cooperation captured in the input covariance tensor structure can improve the sum capacity substantially.

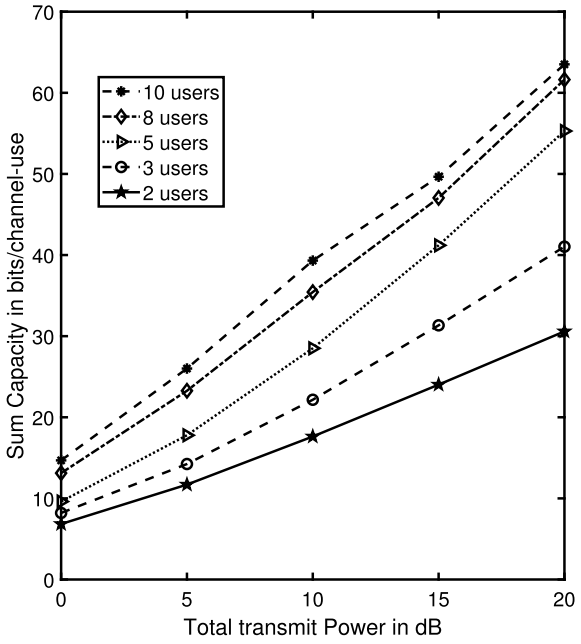

FIGURE 17. Sum capacity vs transmit power for MU MIMO MAC.

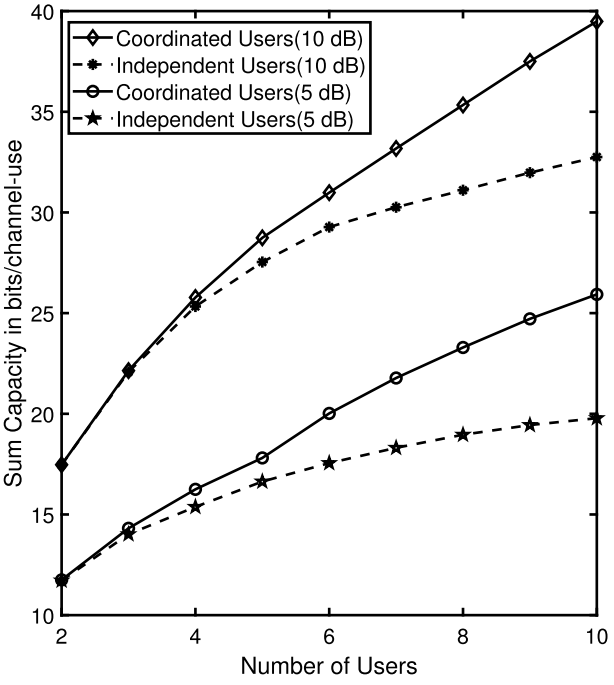

FIGURE 18. Sum capacity vs users for MU MIMO MAC.

Further the capacity region of a 2 users MIMO MAC scenario is presented in Figure 19 with transmit power budgets $P_1 = P_2 = 5$ dB. We consider $K = 2$ where the base station has $N_R = 8$ antennas and 2 different settings for transmit antennas $N_T = 4$ and 8. The capacity region obtained from

the tensor formulation for two users can be written as:

$$\mathcal{C}_{\mathcal{R}}= \bigcup_{\substack{\mathrm{tr}(\mathrm{Q}^{(1)})\le P_1,\\ \mathrm{tr}(\mathrm{Q}^{(2)})\le P_2,\\ \sum_{n=1}^{N_T} \mathcal{Q}_{n,k,n,k} \\ \le P_k, k=1,2.}} \left\{ \begin{array}{c} R_1 \le \log\det\left(\mathrm{I}+\mathrm{H}^{(1)}\mathrm{Q}^{(1)}\mathrm{H}^{(1)H}\right) \\ R_2 \le \log\det\left(\mathrm{I}+\mathrm{H}^{(2)}\mathrm{Q}^{(2)}\mathrm{H}^{(2)H}\right) \\ R_1+R_2 \le \log\det\left(\mathrm{I}+\mathcal{H}*_2\mathcal{Q}*_2\mathcal{H}^H\right) \end{array} \right\} \tag{114}$$

For the two users case, finding the optimum covariance that maximizes the achievable rate of user 1 assuming user 2 is silent reduces to a single user MIMO scenario. Hence the bounds on individual rates R_1 and R_2 in (114) and (102) are the same. Note however that the bound on the sum rate $R_1 + R_2$ differs. With the additional constraint on covariance from (109), the bound on the sum rate in (114) translates to:

$$R_1 + R_2 \le \log\det\left(\mathrm{I}+\mathcal{H}*_2\mathcal{Q}*_2\mathcal{H}^H\right) \tag{115}$$

$$= \log\det\left(\mathrm{I}+\mathrm{H}^{(1)}\mathrm{Q}^{(1)}\mathrm{H}^{(1)H}+\mathrm{H}^{(2)}\mathrm{Q}^{(2)}\mathrm{H}^{(2)H}\right) \tag{116}$$

where (115) to (116) follow from (113). Hence, with additional constraint on covariance $\mathcal{Q}$ as defined by (109), the sum rate bound in (114) reduces to the sum rate bound in (102), and thus all three bounds in (114) depend only on $\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)}$. The additional constraint corresponds to the transmit scheme where each user acts independent of the other user. In such cases, the capacity region is characterized by a pair of covariance matrices $(\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)})$ which satisfies the power constraints. In general, the capacity region of the two users case corresponding to (114) is characterized by a triplet $(\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)}, \mathcal{Q})$ where $\mathcal{Q}$ is the transmit covariance tensor which prescribes a joint transmission scheme. The transmit covariance matrix of individual users $\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)}$ form the sub-tensors of $\mathcal{Q}$. However the optimal $\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)}$ which maximizes R_1, R_2 respectively may not be the sub-tensors of the optimal $\mathcal{Q}$ which maximizes $R_1 + R_2$. The optimal $\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)}$ are found assuming the other user to be silent for transmission, whereas the optimal $\mathcal{Q}$ is found assuming joint transmission by both the users.

The rate regions in (114) and (102) forms a pentagon on a two dimensional R_1, R_2 plane. The capacity region is determined by the convex hull of the union of all such pentagons obtained through different choices of covariances which satisfy the constraints. In Figure 19, the solid line (case 1) represents the pentagon corresponding to (102) where $(\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)})$ are obtained via iterative water-filling from [105, Algorithm 1] to maximize the sum rate with independent transmissions. The dotted line (case 2) represents the pentagon corresponding to (102) where $(\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)})$ are obtained via conventional water-filling for single user MIMO to maximize R_1 and R_2 individually. Similarly, different $(\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)})$ will correspond to different pentagons based on (102). The capacity region with independent users is obtained as a convex hull of the union of all such pentagons

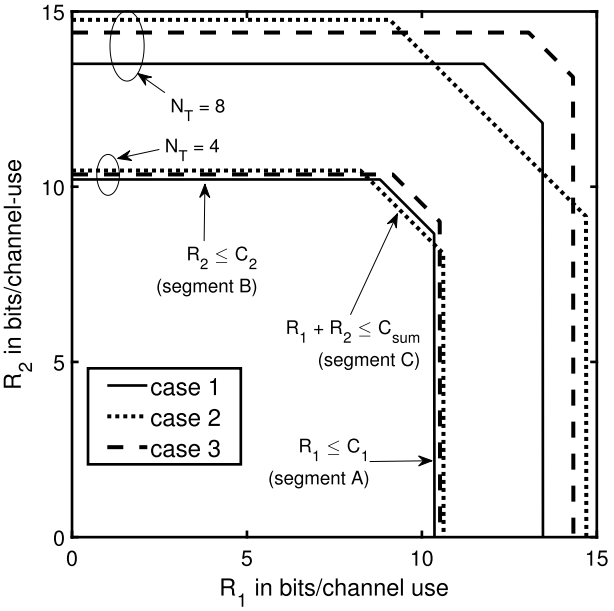

FIGURE 19. Capacity region of 2-users MIMO MAC.

associated with different covariances which satisfy the constraints in (102). Thus the convex hull of the regions of case 1 and 2 gives an inner bound to the capacity region with independent user transmission. The capacity region signifies the bounds on the rate of transmission by each user.

The dashed line (case 3) represents the pentagon corresponding to (114) where $\mathcal{Q}$ is approximated using Algorithm 1 to maximize the sum rate, and $\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)}$ are the sub-tensors of the $\mathcal{Q}$ obtained from Algorithm 1. However such a choice of $\mathrm{Q}^{(1)}$ and $\mathrm{Q}^{(2)}$ may not maximize the individual rates of user 1 and 2. This is reflected in Figure 19 as the dotted line shows a larger bound on individual rates as compared to other cases in both the vertical and horizontal segments, A and B. Case 2 can be seen as another scenario for (114) where $\mathrm{Q}^{(1)}, \mathrm{Q}^{(2)}$ are chosen to maximize the individual rates and the joint covariance tensor $\mathcal{Q}$ has structure $\mathcal{Q}_{:,i,:,j} = \mathrm{Q}^{(i)}$ for $i = j$ and 0 for $i \ne j$ with $i, j = 1, 2$. Such a covariance tensor does not maximize the sum rate but only individual rates and still satisfies all the constraints in (114). The capacity region with user coordination is given by the convex hull of the union of pentagons associated with case 3 and case 2 along with every pentagon associated with different covariances which satisfy the constraints in (114). Thus, a convex hull of regions of case 2 and 3 gives an inner bound to the capacity region with user cooperation. The bound on the sum achievable rate indicated by segment C in Figure 19 is lowest in case 2, followed by the case when transmit covariances are chosen via iterative water-filling (case 1), and is largest when covariance is obtained using the tensor approach assuming user coordination (case 3). Moreover, this difference increases as N_T increases from 4 to 8, as observed in the figure. Hence it is seen that the boundary of the

capacity region expands in segment C with user cooperation as opposed to the independent user transmissions.

B. MIMO INTERFERENCE CHANNEL (IC)

Let us now consider the general multi-user MIMO interference channel (IC) where both transmit and receive side have user separation. Consider K transmit devices having N_T antennas each that are communicating with their respective K receive devices having N_R antennas each. Such a channel model assumes that all the transmitting devices are communicating with their respective receivers while generating interference to all the other receivers. Finding the exact capacity region of a general K user interference channel is still an ongoing effort [106], [107]. For the two users scenario, capacity bounds have been discussed in [108], [109] and references within, under various assumptions regarding interference such as Z interference channels (where one of the two receive users does not experience interference), strong interference, and noisy interference. The term 'noisy interference' refers to conditions where the sum capacity can be achieved by treating interference as noise [110]. For a two users MIMO IC, [110, Theorem 2] presents a set of conditions involving the channel and transmit covariance matrices which are sufficient for interference to be treated as noise. Such conditions ensure that the indirect links are much weaker than the direct links, and power allocation is such that the received power of the message received via the indirect link at each user is very low compared to the received power of the message via the direct link. Further, it is shown in [111] that under strong interference, each receiver can jointly decode the signal and the interference to achieve the sum capacity. As an extension of the two users case, [112] derives the conditions under which treating noise as interference can achieve the sum capacity for a K users IC. Most of these works assume no coordination among the source transmitters or the destinations. If different transmitting users coordinate for transmission, and receivers coordinate for reception then interference can be treated as information bearing entity which can be extracted using the tensor framework.

The system model for a MIMO interference network is given by [113]:

$$\underline{\mathbf{y}}^{(k)} = \mathrm{H}^{(k,k)}\underline{\mathbf{x}}^{(k)} + \sum_{\substack{u \\ u\neq k}} \mathrm{H}^{(k,u)}\underline{\mathbf{x}}^{(u)} + \underline{\mathbf{n}}^{(k)} \tag{117}$$

for $k = 1, \ldots, K$ and $\underline{\mathbf{x}}^{(k)} \in \mathbb{C}^{N_T \times 1}$ is the vector transmitted by source k. Also, $\underline{\mathbf{y}}^{(k)}, \underline{\mathbf{n}}^{(k)} \in \mathbb{C}^{N_R \times 1}$ are the received signal and noise vectors at destination k. Matrix $\mathrm{H}^{(k,k)} \in \mathbb{C}^{N_R \times N_T}$ is the direct channel between source k and destination k and $\mathrm{H}^{(k,u)} \in \mathbb{C}^{N_R \times N_T}$ is the cross-channel matrix between source u and destination k. For each transmitting source, there is an individual power constraint defined as $\mathrm{tr}(\mathrm{Q}^{(k)}) \leq P_k$, where $\mathrm{Q}^{(k)} \in \mathbb{C}^{N_T \times N_T}$ is the covariance matrix of vector $\underline{\mathbf{x}}^{(k)}$ and P_k denotes the power budget. Such interference networks can be thought of as a tensor communication link. The input can be represented as a matrix $\mathbf{X} \in \mathbb{C}^{N_T \times K}$ where each $\underline{\mathbf{x}}^{(k)}$ forms a column of the matrix $\mathbf{X}$. Similarly the received signal and noise can be represented using matrices $\mathbf{Y}, \mathbf{N} \in \mathbb{C}^{N_R \times K}$ where each $\underline{\mathbf{y}}^{(k)}$ and $\underline{\mathbf{n}}^{(k)}$ form columns of the matrices $\mathbf{Y}$ and $\mathbf{N}$. The overall channel between such an input and output can be represented as a fourth order tensor $\mathcal{H} \in \mathbb{C}^{N_R \times K \times N_T \times K}$ where $\mathcal{H}_{:,k,:,u} = \mathrm{H}^{(k,u)}$. Subsequently the interference network system model can be represented in tensor form as:

$$\mathbf{Y} = \mathcal{H} *_2 \mathbf{X} + \mathbf{N}. \tag{118}$$

Note that (118) is different from MIMO MAC specified by (103) in the sense that in (103) the channel is a third order tensor and thus the output is a vector. On the other hand in (118) the channel is a fourth order tensor to account for user separation at the receiver side as well and thus the output is a matrix or an order-2 tensor. The power constraints on the input can be defined in a similar way as for (103). Assuming the noise covariance is an identity tensor $\mathcal{I}$ of size $N_R \times K \times N_R \times K$, the tensor formulation can be used to specify the channel capacity as

$$\max_{\mathcal{Q}} \log\det\left(\mathcal{H} *_2 \mathcal{Q} *_2 \mathcal{H}^H + \mathcal{I}\right) \tag{119}$$

$$s.t. \sum_{n=1}^{N_T} \mathcal{Q}_{n,k,n,k} \leq P_k \quad \forall k, \quad \mathcal{Q} \succeq 0. \tag{120}$$

Note that capacity obtained from such a tensor formulation assumes that all the sources cooperate for transmission and all the destinations cooperate for reception.

Now we consider an example with two users interference channel and compare the sum rate achieved using the tensor framework which assumes user cooperation, with the upper bound on rate suggested in [108] while treating interference as noise. We consider the same example from [108] consisting of a system composed of two transmitters and two receivers, equipped with N_T antennas each. The model introduces a positive scalar $a \geq 0$ to control the interference power:

$$\underline{\mathbf{y}}^{(1)} = \mathrm{H}^{(1,1)}\underline{\mathbf{x}}^{(1)} + a \cdot \mathrm{H}^{(1,2)}\underline{\mathbf{x}}^{(2)} + \underline{\mathbf{n}}^{(1)}, \tag{121}$$

$$\underline{\mathbf{y}}^{(2)} = a \cdot \mathrm{H}^{(2,1)}\underline{\mathbf{x}}^{(1)} + \mathrm{H}^{(2,2)}\underline{\mathbf{x}}^{(2)} + \underline{\mathbf{n}}^{(2)}. \tag{122}$$

Such a system of equations can be equivalently represented using (118) where the channel $\mathcal{H}$ is a tensor of size $N_T \times 2 \times N_T \times 2$ and the input, output and noise are matrices of size $N_T \times 2$ each. We find the capacity of such MIMO interference channels with user coordination using the tensor framework where the optimal input covariance is approximated using Algorithm 1.

For our example, in (121), (122) we take $a = 1/\sqrt{3}$ and the channel entries are i.i.d. zero mean unit variance Gaussian random variables as in [108]. The results are averaged over 100 different channel realizations. These channel realizations are known at the transmitters and the receivers. We find the sum rate achieved via the tensor framework assuming user cooperation, and denote it using R_T. We compare R_T with the capacity of K parallel non-interfering channels found using standard water-filling approach and denote it as R_U. In [108], R_U has been used as an upper bound on

the achievable sum rate while treating interference as noise. Figures 20-24 presents a comparison between R_T and R_U. To calculate R_U, we find the transmit covariance matrix $\mathrm{Q}^{(k)}$ for each user input $\underline{\mathbf{x}}^{(k)}$ based on MIMO water-filling corresponding to the channel $\mathrm{H}^{(k,k)}$, and set $R_U = \sum_k \log\det(\mathrm{I} + \mathrm{H}^{(k,k)}\mathrm{Q}^{(k)}\mathrm{H}^{(k,k)H})$. Also, R_T is calculated using the objective function in (119) where the optimal covariance tensor $\mathcal{Q}$ is approximated using Algorithm 1.

Figure 20 compares R_T and R_U for a 2 users case with different values of power constraints $P_1 = P_2 = P$. The achievable sum rates with the two approaches are plotted against the number of antennas N_T. The sum rates for both the cases increase as N_T increases. The achievable rate with the tensor approach, R_T is higher than the upper bound on the sum rate, R_U from [108]. It can also be observed in Figure 20 that as P increases, the gap between R_T and R_U increases as well. The tensor approach shows that the presence of interference can in fact give higher achievable sum rates if the transmit and receive operations are performed jointly by all the transmitting and receiving users respectively. Hence the sum rate achieved via the tensor approach, that allows cooperation at the transmitter and receiver sides, can be higher than the sum rate achieved in the absence of interference.

This feature is further observed in Figure 21 which presents the sum rate against the number of users K for a multi-user MIMO interference scenario from (117) with $N_T = 2$. Each user's receive signal contains a desired signal and information from $K-1$ interfering links whose power is controlled by a scalar factor a as described in the two users case. The result presented is for $a = 1/\sqrt{3}$ and for two different total power budgets $P = 5, 10$ dB with individual power constraints as $P_k = P/K$. It can be seen that R_T is always larger than R_U. Furthermore, as the number of users grows, the difference between R_T and R_U also widens, which shows the advantage of considering interference as information bearing term rather than noise, through the tensor framework.

To further elaborate on the role of interference, Figure 22 presents the sum rate R_T of the two users scenario achieved using the tensor framework for different interference power and compares it with R_U. The solid lines represent R_T and dashed lines represent R_U. The result is plotted against the interference coupling power gain defined as $G_1 = 10\log_{10} a^2$ dB [108] for different number of antennas N_T at each device, and with power constraints $P_k = P/K$ for $k = 1, 2$ where $K = 2$ and P is set to 10 dB. It is seen clearly that higher interference leads to higher sum rate R_T using the tensor approach. Also, R_T increases with N_T and the gap between sum rate for different number of antennas widens with increasing interference. However, R_U is always lower than R_T and does not change with G_1. Since R_U is calculated by assuming zero interference, it does not vary with changing the interference power. At very low interference power, we see that R_T and R_U are almost same. The difference between R_T and R_U starts to be significant (more than 1 bit/channel-use) at an interference power G_1 of approximately -2 dB when $N_T = 2$, and around -6 dB when $N_T = 8$.

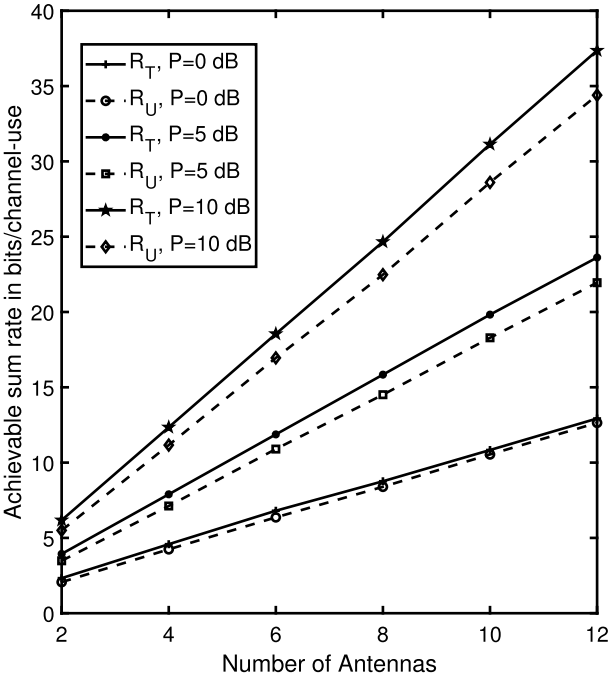

FIGURE 20. Achievable sum rate vs number of antennas for two users MIMO IC.

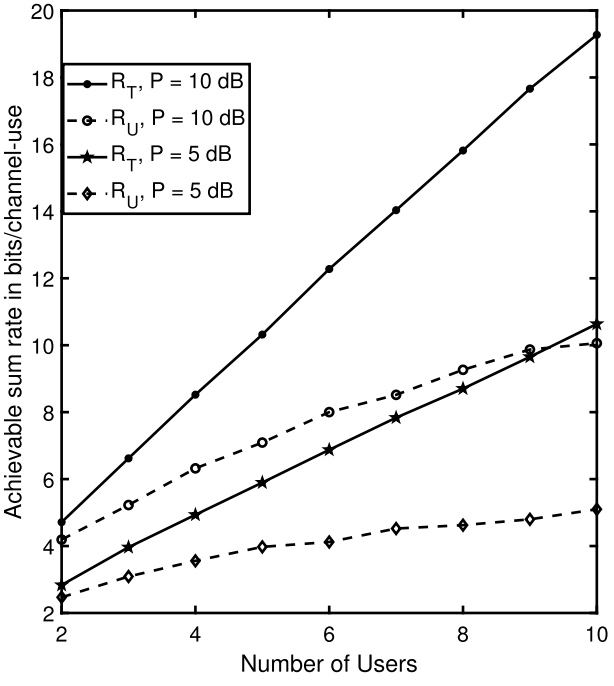

FIGURE 21. Achievable sum rate vs number of users for K-users MIMO IC with $N_T = 2$.

Next we consider a three users system specified by:

$$\underline{\mathbf{y}}^{(1)} = \mathrm{H}^{(1,1)}\underline{\mathbf{x}}^{(1)} + a \cdot \mathrm{H}^{(1,2)}\underline{\mathbf{x}}^{(2)} + b \cdot \mathrm{H}^{(1,3)}\underline{\mathbf{x}}^{(3)} + \underline{\mathbf{n}}^{(1)} \tag{123}$$

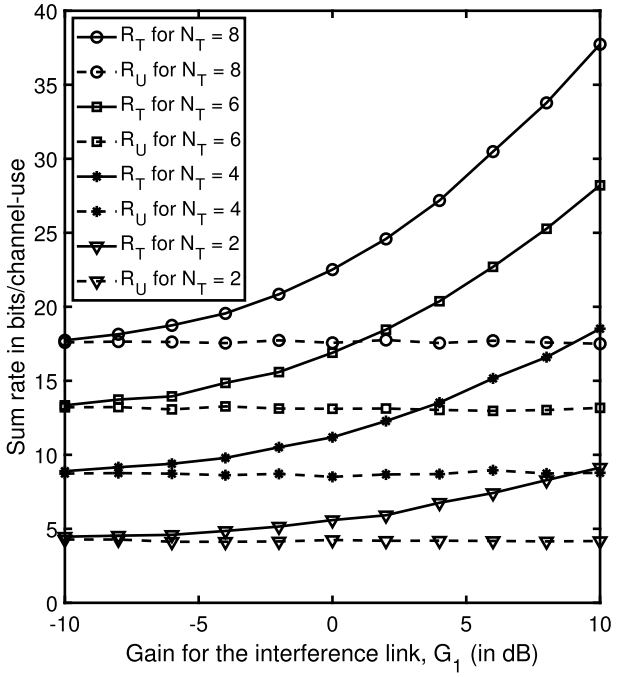

FIGURE 22. Achievable sum rate vs interference power for two users MIMO IC.

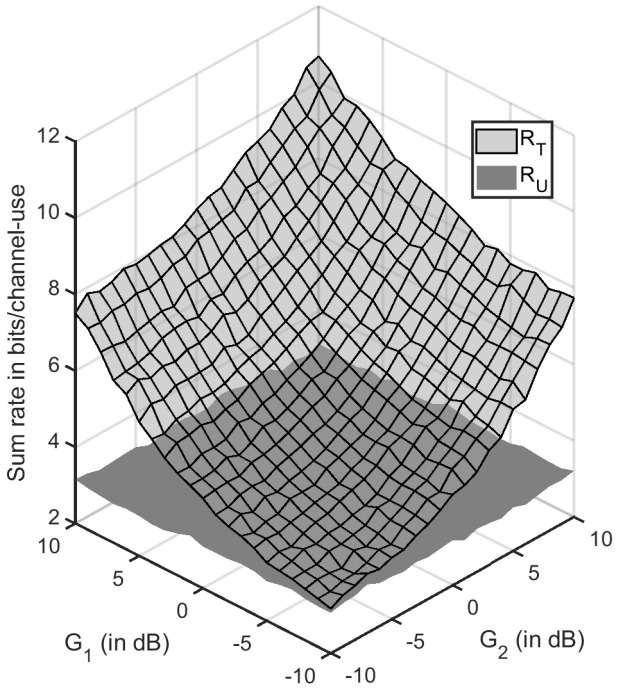

FIGURE 23. Achievable sum rate vs interference power for three users MIMO IC with $N_T = 2$.

$$\underline{\mathbf{y}}^{(2)} = a \cdot \mathrm{H}^{(2,1)}\underline{\mathbf{x}}^{(1)} + \mathrm{H}^{(2,2)}\underline{\mathbf{x}}^{(2)} + b \cdot \mathrm{H}^{(2,3)}\underline{\mathbf{x}}^{(3)} + \underline{\mathbf{n}}^{(2)} \tag{124}$$

$$\underline{\mathbf{y}}^{(3)} = a \cdot \mathrm{H}^{(3,1)}\underline{\mathbf{x}}^{(1)} + b \cdot \mathrm{H}^{(3,2)}\underline{\mathbf{x}}^{(2)} + \mathrm{H}^{(3,3)}\underline{\mathbf{x}}^{(3)} + \underline{\mathbf{n}}^{(3)} \tag{125}$$

Such a system model can be represented using (118) with $\mathbf{Y}, \mathbf{X}$ and $\mathbf{N}$ as $N_T \times 3$ matrices each and $\mathcal{H}$ as an $N_T \times 3 \times N_T \times 3$ tensor. The number of antennas at each device is denoted by N_T. Each destination user receives signals from a direct link and 2 interfering links whose power is controlled by scalar factors a and b respectively. In Figure 23 we present R_T and R_U against interference power with three users for $N_T = 2$. The power constraints for each user is same, i.e. $P_k = P/K$ where $K = 3$, and P is set to 5 dB. In Figure 23 we have $G_1 = 10\log_{10} a^2$ dB and $G_2 = 10\log_{10} b^2$ dB, where G_1, G_2 represent the strength of the 2 interfering links for each user. It can be seen that R_T is low when the interference power of both links is weak. With increasing strength of the interfering links, we get higher sum rate R_T. The curve for R_U does not vary with change in G_1 and G_2 and is always lower than R_T. Similar observation can be made in Figure 24 which represents R_T and R_U for $N_T = 4$. On comparing Figures 23 and 24, we see that R_T and R_U increase as N_T increases from 2 to 4. Notice that difference in R_T for $N_T = 2$ and 4 is wider for large values of G_1, G_2, i.e. when the strengths of the interfering links increase. The tensor framework allows to treat interfering terms as information bearing components, resulting in higher rates with increasing G_1, G_2 and N_T.

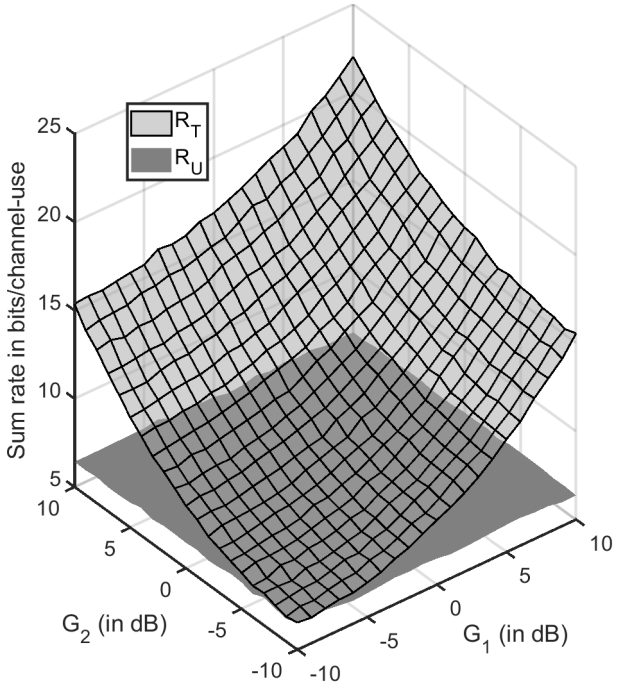

FIGURE 24. Achievable sum rate vs interference power for three users MIMO IC with $N_T = 4$.

C. COMPLEXITY COMPARISON

The matrix based methods used for estimating the sum capacity for MIMO MAC and IC often treats interference as noise. In most cases, the objective of such simplification is a reduction in the computational complexity of the problem. Hence

TABLE 3. Computational complexity for MIMO MAC.

Approach	Complexity
Matrix based iterative water-filling	$\mathcal{O}(K \cdot I \cdot N_R \cdot N_T^2)$
Tensor based Algorithm 1	$\mathcal{O}((K \cdot N_T)^3)$

it is important to compare the complexity of the matrix based methods with the tensor method presented in this paper to highlight the relevance of both the approaches.

In section V-C, we presented the complexity analysis of employing Algorithm 1 for any given tensor channel of size $J_1 \times J_2 \times \ldots J_M \times I_1 \times I_2 \times \cdots \times I_N$. In this section, we will use the results from section V-C and compare the computational complexity of finding capacity for the specific cases of MIMO MAC and IC channels using the tensor method with the matrix based methods.

1) COMPARISON FOR MIMO MAC

Consider the system model for K users MIMO MAC from (100) where $\mathrm{H}^{(k)}$ denotes the uplink channel matrix of size $N_R \times N_T$ between the BS equipped with N_R antennas and the kth user which has N_T transmit antennas. The iterative water-filling approach of finding the sum capacity computes a separate covariance matrix for each user while treating the interference from all other users as noise. A standard water-filling approach is used in finding the individual covariance matrices. Furthermore, all the K covariance matrices are iteratively updated until a convergence criteria is met. The dominant step in a standard water-filling approach is the singular value decomposition of the channel matrix. The computational complexity of the SVD of an $m \times n$ matrix is given by $\mathcal{O}(m \cdot n \cdot \min(m, n))$ [114], [115]. Thus, the computational complexity of computing the water-filling corresponding to the matrix $\mathrm{H}^{(k)} \in \mathbb{C}^{N_R \times N_T}$ would be given as $\mathcal{O}(N_R \cdot N_T^2)$ (assuming $N_R > N_T$). For the MIMO MAC, since the water-filling is computed for each of the K users iteratively, the complexity scales linearly with the number of users K, and the number of required iterations per user, denoted by I. Thus the overall complexity is given as $\mathcal{O}(K \cdot I \cdot N_R \cdot N_T^2)$.

On the contrary, the tensor approach in Algorithm 1 is a non-iterative method to approximate the joint covariance of all the users, which also takes into account the interference as information bearing entities. Consider the tensor model for MIMO MAC from (103) where the channel is defined as a third order tensor of size $N_R \times N_T \times K$. Thus based on the discussion in section V-C, the complexity of employing Algorithm 1 for the MIMO MAC channel is given by $\mathcal{O}((K \cdot N_T)^2 \cdot N_R + (K \cdot N_T)^3)$. Assuming that $K \cdot N_T \geq N_R$, we can write the computational complexity as $\mathcal{O}((K \cdot N_T)^3)$. The complexity of the matrix and tensor approaches are listed in Table 3 for a clear comparison.

It can be seen that the complexity of tensor approach is cubic in the number of users as opposed to the matrix based iterative water-filling where the complexity is linear in the number of users. Although it is to be noted that the matrix based approach also depends on the number of iterations. Within each such iteration, a standard water-filling algorithm for a matrix channel is executed. In contrast, the tensor approach presented in this paper is non-iterative, i.e. a single iteration is required irrespective of the number of users and antennas.

However, clearly the advantage of the matrix approach should not be dismissed as it offers a low complexity solution. Despite the iterative nature of the matrix approach, the complexity is linear in the number of users as opposed to the tensor approach where it is cubic. However, as the number of users grows the advantage of tensor based approach becomes significant in terms of the achievable sum rate possible due to user coordination. This is evident in Figure 18 where the sum capacity of the coordinated users shows a substantial increase as compared with the independent users, when the number of users grow. For instance, at a total transmit power of 10 dB, the sum capacity for 10 users is 40 bits/channel-use as opposed to around 32 bits/channel-use for the independent users as seen in Figure 18. To better quantify this gap, for the example of Figure 18 we plot the difference between the sum rates achievable with tensor framework and the matrix approach in Figure 25 against the number of users and different values of N_T at $P = 10$ dB. We can see that the difference increases substantially with increasing the number of users, and also increasing the number of transmit antennas. To summarize, Table 3 shows the advantage of the matrix approach in terms of lower complexity, and Figure 25 shows the advantage of the tensor approach in terms of higher rates due to user coordination.

2) COMPARISON FOR MIMO IC

Now let us consider the case of MIMO IC for K users from (117). To calculate R_U, which is the upper bound on the achievable sum rate while treating interference as noise, K separate transmit covariance matrices have to be calculated corresponding to each user. Thus, the complexity in this case is also linear in the number of users. Each user solves a water-filling solution corresponding to a matrix channel $\mathrm{H}^{(k,k)}$ of size $N_R \times N_T$. Assuming $N_R = N_T = N$, the complexity is given as $\mathcal{O}(K \cdot N^3)$. The corresponding tensor model for K users MIMO IC is given by (118) where the channel is a fourth order tensor of size $N \times K \times N \times K$. The process of calculating R_T, which denotes the achievable sum rate using user coordination, requires Algorithm 1 to be executed for the fourth order tensor channel. Thus, based on the discussion in section V-C, its computational complexity is given by $\mathcal{O}((K \cdot N)^3)$. The complexity of the matrix and the tensor models are listed in Table 4 for a comparison.

Ignoring the interference leads to a degenerate situation where a higher order channel is reduced to K non-interfering matrix channels. Thus, the computational complexity of calculating R_U is lower when compared with the tensor method used to find R_T, as seen in Table 4. However, as shown in Figure 21, the difference between R_T and R_U can be very large with increasing the number of users. For instance, for

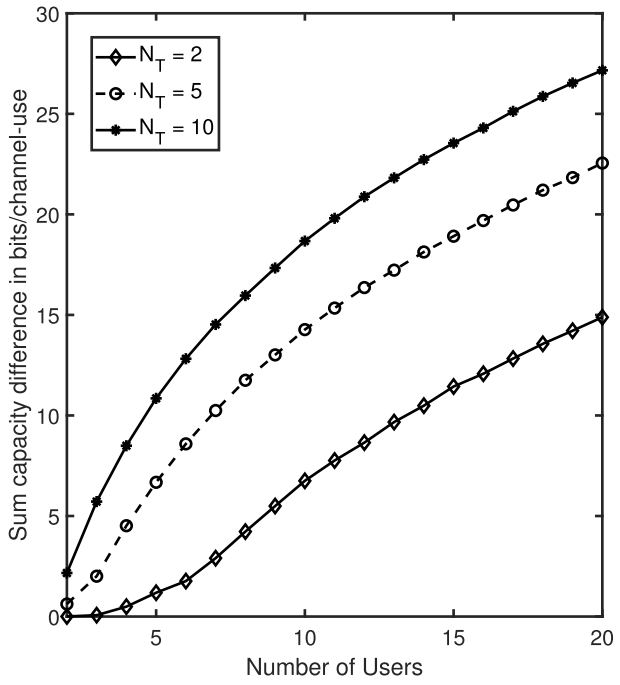

FIGURE 25. Difference in sum capacity achieved via tensor approach (coordinated users) and matrix approach (independent users) vs number of users for MIMO MAC.

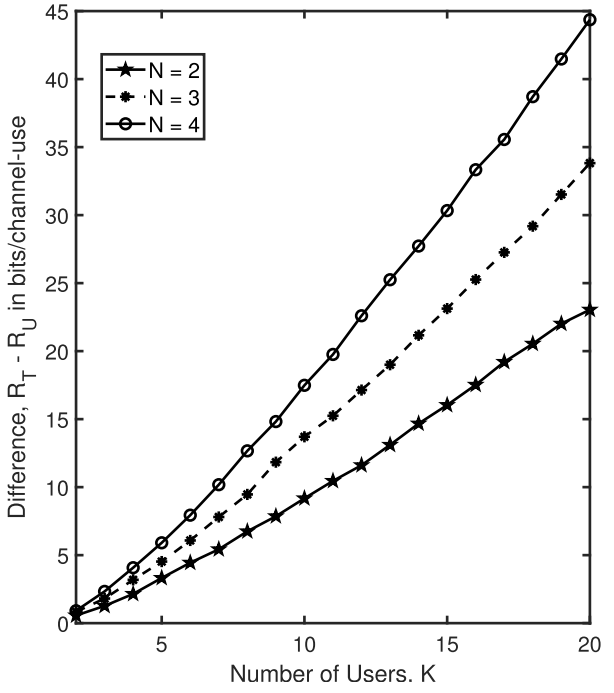

FIGURE 26. Difference in achievable sum rate, $R_T - R_U$ at $P = 10$ dB vs number of users for K users MIMO IC.

TABLE 4. Computational complexity for MIMO IC.

Approach	Complexity
Calculating R_U using water-filling	$\mathcal{O}(K \cdot N^3)$
Calculating R_T using Algorithm 1	$\mathcal{O}((K \cdot N)^3)$

10 users at 10 dB total power, R_T is around 19 bits/channel-use as opposed to R_U which is only around 10 bits/channel-use as observed in Figure 21. To better quantify the difference between R_T and R_U, we consider the example from Figure 21 for $P = 10$ dB, and plot the difference $R_T - R_U$ against number of users for different values of N in Figure 26. It can be seen in Figure 26 that this difference increases significantly as the number of users, or the number of antennas increases. Thus, Figure 26 clearly shows the advantage of the tensor framework in providing higher achievable sum rates.

3) ADVANTAGE OF THE TENSOR APPROACH

In both the MIMO MAC and MIMO IC, the achievable sum rates are interference limited using the matrix approaches. Hence as the number of users or antennas grows which leads to higher interference, the performance degrades in terms of achievable sum rates. It is clear that the tensor method should be preferred when user coordination is allowed leading to a joint transmission across all users, whereas the matrix based methods can be used when users are constrained to act independently and transmit on a per user basis. The tensor method while admitting higher complexity, as seen in Tables 3 and 4, leads to much higher achievable sum rates also, as seen in Figures 25 and 26. Thus there is an inherent performance/complexity trade-off where at the cost of higher complexity, much significant gain in sum rates can be achieved especially in the presence of strong inter-domain interference.

The representation of such multi-domain systems using matrix channels and vector signals often gets more complex as the number of domains or dimensions within a domain increases. It is important to note that the tensor method does not necessarily yields lower complexity for such high complexity situations. Its main feature is that it provides a structured mathematically accurate framework to treat high complexity systems, while having the capability to naturally scale back to lower complexity when the problem is of reduced complexity. For example, for a single user MIMO, the tensor approach scales back to a conventional matrix approach. For the MU MIMO MAC with additional constraint in (109), the tensor framework reduces to the conventional MU MIMO MAC model. Hence the tensor method can be seen as an umbrella set up wherein several other low complexity approaches act as specific cases with additional constraints imposed for simplifying the problem.

VIII. CONCLUSIONS

In this paper, we introduced a unified framework using tensors to represent a multi-domain communication system. The proposed framework can be used to model many of the currently used multi-domain approaches in communications by leveraging the multi-linear structure of the system. Several examples of such systems were included in this paper to demonstrate the usefulness of the proposed tensor framework.

The information-theoretic analysis for a fixed tensor channel presented in this paper suggests an exponential increase in Shannon capacity with increase in the number of domains, if the channel is known at both transmit and receive side. In particular, the tensor framework's ability to mathematically represent a family of power constraints is utilized. This allows to characterize the capacity of multi-user MIMO systems having per user power constraints for any number of users. A tensor representation of the channel and its information theoretic analysis using the proposed framework leads to a joint transmission scheme across all the domains. An example of MIMO GFDM system was presented to illustrate that such a joint domain processing can lead to better BER performance as compared to per domain processing. It was also shown that the capacity of tensor channels decreases as the correlation among the channel components increases. Furthermore, for the same total power as the number of domain elements under individual power constraints increases, the capacity decreases. In case of multi-user MAC and IC channels, it was shown that the tensor approach leads to higher achievable sum rates with user-coordination as compared to the sum rates achieved with independent users. The tensor framework allows to capture the user cooperation in the form of a joint covariance tensor across all the domains. Our results show that such an improvement in achievable rates through the tensor approach becomes even more significant as the number of users grows or the power of the interfering links increases. With independent user transmissions, the interference is treated as noise, as opposed to the tensor approach where interference is treated as information bearing entity. The performance advantages resulting from using the tensor framework come very often at the expense of a complexity increase. However, tensors provide structured and mathematically accurate tools of handling such complexity increase.

.

APPENDIX A SOLVING THE EQUATIONS DERIVED FROM KKT CONDITIONS FOR THE OPTIMAL COVARIANCE TENSOR

In this appendix, we present the solution to the equations obtained through KKT conditions for finding the optimal transmit covariance tensor $\mathcal{Q}$. The results presented here are a generalization of Theorem 1 from [73] to a tensor setting. In this appendix, $\mathcal{H}$ represents the channel tensor, $\mathcal{M}$ represents the Lagrange multiplier corresponding to the semi-definite constraint on covariance and $\mathcal{B}$ represents the tensor whose pseudo-diagonal entries are the Lagrange multipliers corresponding to the other constraints on transmit covariance (such as power). For the sum power constraint, $\mathcal{B}$ is an identity tensor and the Lagrange multiplier is a scalar $\mu > 0$. So for $\mathcal{M} \succeq 0$, $\mathcal{B} \succ 0$, our objective is to find $\mathcal{Q} \succeq 0$, that satisfies (126) and (127):

$$\mathcal{H}^H *_M (\mathcal{H} *_N \mathcal{Q} *_N \mathcal{H}^H + \mathcal{I}_M)^{-1} *_M \mathcal{H} = \mu\mathcal{B} - \mathcal{M} \tag{126}$$

$$\mathcal{Q}^{1/2} *_N \mathcal{M} *_N \mathcal{Q}^{1/2} = 0_{\mathcal{T}} \tag{127}$$

Contracting along the N consecutive modes in (126) on left and right by $\mathcal{Q}^{1/2}$ results in

$$\begin{aligned}\mathcal{Q}^{1/2} *_N \mathcal{H}^H *_M (\mathcal{H} *_N \mathcal{Q} *_N \mathcal{H}^H + \mathcal{I}_M)^{-1} *_M \mathcal{H} *_N \mathcal{Q}^{1/2} \\ = \mu\mathcal{Q}^{1/2} *_N \mathcal{B}^{1/2} *_N \mathcal{B}^{1/2} *_N \mathcal{Q}^{1/2} - \underbrace{\mathcal{Q}^{1/2} *_N \mathcal{M} *_N \mathcal{Q}^{1/2}}_{=0_{\mathcal{T}} \text{ (from (127))}}\end{aligned} \tag{128}$$

$$\mathcal{A} \triangleq \mathcal{B}^{1/2} *_N \mathcal{Q}^{1/2} \Rightarrow \mathcal{Q}^{1/2} = \mathcal{B}^{-1/2} *_N \mathcal{A} \tag{129}$$

Since $\mathcal{Q}$ is Hermitian, we have $\mathcal{Q}^{1/2} = (\mathcal{Q}^{1/2})^H = \mathcal{A}^H *_N \mathcal{B}^{-1/2}$, which gives

$$\mathcal{Q} = \mathcal{B}^{-1/2} *_N \mathcal{A} *_N \mathcal{A}^H *_N \mathcal{B}^{-1/2} \tag{130}$$

Substituting (130) in (126) and contracting both sides along N modes of $\mathcal{B}^{-1/2}$ results in

$$\begin{aligned}&\underbrace{\mathcal{B}^{-1/2} *_N \mathcal{H}^H}_{\mathcal{K}^H} *_M \Big(\underbrace{\mathcal{H} *_N \mathcal{B}^{-1/2}}_{\mathcal{K}} *_N \mathcal{A} *_N \mathcal{A}^H \\ &\quad *_N \underbrace{\mathcal{B}^{-1/2} *_N \mathcal{H}^H}_{\mathcal{K}^H} + \mathcal{I}_M \Big)^{-1} *_M \underbrace{\mathcal{H} *_N \mathcal{B}^{-1/2}}_{\mathcal{K}} \\ &= \mu\mathcal{I}_N - \mathcal{B}^{-1/2} *_N \mathcal{M} *_N \mathcal{B}^{-1/2}\end{aligned} \tag{131}$$

Proposition 1.1: Let $\mathcal{A} = \mathcal{U}_{\mathcal{A}} *_N \mathcal{D}_{\mathcal{A}} *_N \mathcal{V}_{\mathcal{A}}^H$ and $\mathcal{K} \triangleq \mathcal{H} *_N \mathcal{B}^{-1/2} = \mathcal{U}_{\mathcal{K}} *_M \mathcal{D}_{\mathcal{K}} *_N \mathcal{V}_{\mathcal{K}}^H$ represent the tensor SVD of $\mathcal{A}$ and $\mathcal{K}$, then $\mathcal{U}_{\mathcal{A}} = \mathcal{V}_{\mathcal{K}}$.

Proof: Substituting (129) and (130) into (128), we get

$$\begin{aligned}&\underbrace{\mathcal{A}^H *_N \mathcal{B}^{-1/2} *_N \mathcal{H}^H}_{\mathcal{P}^H} *_M (\underbrace{\mathcal{H} *_N \mathcal{B}^{-1/2} *_N \mathcal{A}}_{\mathcal{P}} \\ &\quad *_N \underbrace{\mathcal{A}^H *_N \mathcal{B}^{-1/2} *_N \mathcal{H}^H}_{\mathcal{P}^H} + \mathcal{I}_M)^{-1} \\ &\quad *_M \underbrace{\mathcal{H} *_N \mathcal{B}^{-1/2} *_N \mathcal{A}}_{\mathcal{P}} \\ &= \mu\mathcal{A}^H *_N \mathcal{A}\end{aligned} \tag{132}$$

Substituting tensor SVD of $\mathcal{A}$ and $\mathcal{P} \triangleq \mathcal{H} *_N \mathcal{B}^{-1/2} *_N \mathcal{A} = \mathcal{U}_{\mathcal{P}} *_M \mathcal{D}_{\mathcal{P}} *_N \mathcal{V}_{\mathcal{P}}^H$ into (132),

$$\begin{aligned}&\mathcal{V}_{\mathcal{P}} *_N \mathcal{D}_{\mathcal{P}}^H *_M \mathcal{U}_{\mathcal{P}}^H *_M (\mathcal{U}_{\mathcal{P}} *_M (\mathcal{D}_{\mathcal{P}} *_N \mathcal{D}_{\mathcal{P}}^H + \mathcal{I}_M) *_M \mathcal{U}_{\mathcal{P}}^H)^{-1} \\ &\quad *_M \mathcal{U}_{\mathcal{P}} *_M \mathcal{D}_{\mathcal{P}} *_N \mathcal{V}_{\mathcal{P}}^H = \mu(\mathcal{V}_{\mathcal{A}} *_N \mathcal{D}_{\mathcal{A}}^H *_N \mathcal{D}_{\mathcal{A}} *_N \mathcal{V}_{\mathcal{A}}^H) \\ &\Rightarrow \mathcal{V}_{\mathcal{P}} *_N \overbrace{\mathcal{D}_{\mathcal{P}}^H *_M (\mathcal{D}_{\mathcal{P}} *_N \mathcal{D}_{\mathcal{P}}^H + \mathcal{I}_M)^{-1} *_M \mathcal{D}_{\mathcal{P}}}^{\text{pseudo-diagonal}} *_N \mathcal{V}_{\mathcal{P}}^H \\ &= (\mathcal{V}_{\mathcal{A}} *_N \overbrace{\mu\mathcal{D}_{\mathcal{A}}^H}^{\text{pseudo-diagonal}} *_N \mathcal{D}_{\mathcal{A}} *_N \mathcal{V}_{\mathcal{A}}^H)\end{aligned} \tag{133}$$

Since $\mathcal{V}_{\mathcal{P}}$ and $\mathcal{V}_{\mathcal{A}}$ are unitary tensors, and the middle quantities on both sides of (133) are pseudo-diagonal, both the right and left side represent the tensor EVD of two equal tensors. From the uniqueness of tensor EVD, (133) implies that $\mathcal{V}_{\mathcal{P}} = \mathcal{V}_{\mathcal{A}}$. Also,

$$\mathcal{P}^H *_M \mathcal{P} = \mathcal{A}^H *_N \underbrace{\mathcal{B}^{-1/2} *_N \mathcal{H}^H}_{\mathcal{K}^H} *_M \underbrace{\mathcal{H} *_N \mathcal{B}^{-1/2}}_{\mathcal{K}} *_N \mathcal{A} \tag{134}$$

$$\mathcal{V}_{\mathcal{K}} *_N \mathcal{D}_{\mathcal{K}}^H *_M \mathcal{U}_{\mathcal{K}}^H *_M (\mathcal{U}_{\mathcal{K}} *_M \mathcal{D}_{\mathcal{K}} *_N \mathcal{D}_{\mathcal{A}} *_N \mathcal{D}_{\mathcal{A}}^H *_N \mathcal{D}_{\mathcal{K}}^H *_M \mathcal{U}_{\mathcal{K}}^H + \mathcal{I}_M)^{-1} *_M \mathcal{U}_{\mathcal{K}} *_M \mathcal{D}_{\mathcal{K}} *_N \mathcal{V}_{\mathcal{K}}^H$$
$$= \mathcal{V}_{\mathcal{K}} *_N \underbrace{\mathcal{D}_{\mathcal{K}}^H *_M (\mathcal{D}_{\mathcal{K}} *_N \mathcal{D}_{\mathcal{A}} *_N \mathcal{D}_{\mathcal{A}}^H *_N \mathcal{D}_{\mathcal{K}}^H + \mathcal{I}_M)^{-1} *_M \mathcal{D}_{\mathcal{K}}}_{\text{pseudo-diagonal}} *_N \mathcal{V}_{\mathcal{K}}^H \tag{137}$$

$$= \mathcal{V}_{\mathcal{A}} *_N \underbrace{\mathcal{D}_{\mathcal{A}}^H *_N \mathcal{U}_{\mathcal{A}}^H *_N \mathcal{V}_{\mathcal{K}} *_N \mathcal{D}_{\mathcal{K}}^H *_M \mathcal{D}_{\mathcal{K}} *_N}_{\text{middle term}}$$
$$\underbrace{\mathcal{V}_{\mathcal{K}}^H *_N \mathcal{U}_{\mathcal{A}} *_N \mathcal{D}_{\mathcal{A}}}_{\text{middle term}} *_N \mathcal{V}_{\mathcal{A}}^H \tag{135}$$

Since $\mathcal{V}_{\mathcal{A}} = \mathcal{V}_{\mathcal{P}}$ we see that (135) represents the tensor EVD of $\mathcal{P}^H *_M \mathcal{P}$. Hence the middle term in (135) is pseudo-diagonal. So $\mathcal{U}_{\mathcal{A}}^H *_N \mathcal{V}_{\mathcal{K}}$ must also be a pseudo-diagonal tensor. Let $\mathcal{S} \triangleq \mathcal{U}_{\mathcal{A}}^H *_N \mathcal{V}_{\mathcal{K}}$, then since $\mathcal{U}_{\mathcal{A}}$ and $\mathcal{V}_{\mathcal{K}}$ are unitary, we get

$$\mathcal{S}^H *_N \mathcal{S} = (\mathcal{U}_{\mathcal{A}}^H *_N \mathcal{V}_{\mathcal{K}})^H *_N (\mathcal{U}_{\mathcal{A}}^H *_N \mathcal{V}_{\mathcal{K}})$$
$$= (\mathcal{V}_{\mathcal{K}}^H *_N \underbrace{\mathcal{U}_{\mathcal{A}}) *_N (\mathcal{U}_{\mathcal{A}}^H}_{\mathcal{I}_N} *_N \mathcal{V}_{\mathcal{K}}) = \mathcal{I}_N \tag{136}$$

Since $\mathcal{S}$ is pseudo-diagonal, (136) implies $\mathcal{S} = \mathcal{I}_N \Rightarrow \mathcal{U}_{\mathcal{A}} = \mathcal{V}_{\mathcal{K}}$, proving the proposition. □

From the tensor SVD of $\mathcal{A}$ and $\mathcal{K}$, and Proposition 1.1, the left-hand side of (131) can be written as (137), as shown at the top of the page.

From the EVD, $\mathcal{B}^{-1/2} *_N \mathcal{M} *_N \mathcal{B}^{-1/2} = \mathcal{U} *_N \mathcal{D}_{\mathcal{B}\mathcal{M}} *_N \mathcal{U}^H$, the right-hand side in (131) becomes

$$\mu \mathcal{I}_N - \mathcal{B}^{-1/2} *_N \mathcal{M} *_N \mathcal{B}^{-1/2}$$
$$= \mathcal{U} *_N (\mu \mathcal{I}_N - \mathcal{D}_{\mathcal{B}\mathcal{M}}) *_N \mathcal{U}^H. \tag{138}$$

Equations (137) and (138) represent the tensor EVD of the left and right hand side of (131), hence from uniqueness of tensor EVD we get $\mathcal{U} = \mathcal{V}_{\mathcal{K}}$ and:

$$\mathcal{D}_{\mathcal{K}}^H *_M (\mathcal{D}_{\mathcal{K}} *_N \mathcal{D}_{\mathcal{A}} *_N \mathcal{D}_{\mathcal{A}}^H *_N \mathcal{D}_{\mathcal{K}}^H + \mathcal{I}_M)^{-1} *_M \mathcal{D}_{\mathcal{K}}$$
$$= \mu \mathcal{I}_N - \mathcal{D}_{\mathcal{B}\mathcal{M}} \tag{139}$$

Let the pseudo-diagonal elements of $\mathcal{D}_{\mathcal{A}}$, $\mathcal{D}_{\mathcal{K}}$ and $\mathcal{D}_{\mathcal{B}\mathcal{M}}$ be $a_{i_1,\ldots,i_N}$, $k_{i_1,\ldots,i_N}$ and $m_{i_1,\ldots,i_N}$ respectively. Pseudo-diagonal elements of $\mathcal{D}_{\mathcal{A}}^H$ and $\mathcal{D}_{\mathcal{K}}^H$ will also be $a_{i_1,\ldots,i_N}$ and $k_{i_1,\ldots,i_N}$ respectively as these are real values. Since both sides of (139) are pseudo-diagonal, hence (139) can be written component-wise as:

$$\frac{k_{i_1,\ldots,i_N}^2}{1 + a_{i_1,\ldots,i_N}^2 k_{i_1,\ldots,i_N}^2} = \mu - m_{i_1,\ldots,i_N} \tag{140}$$
$$\Rightarrow a_{i_1,\ldots,i_N}^2 = \frac{1}{\mu - m_{i_1,\ldots,i_N}} - \frac{1}{k_{i_1,\ldots,i_N}^2} \tag{141}$$

Substituting (129) into (127) gives $\mathcal{A}^H *_N \mathcal{B}^{-1/2} *_N \mathcal{M} *_N \mathcal{B}^{-1/2} *_N \mathcal{A} = 0_{\mathcal{T}}$ where using tensor SVD of $\mathcal{A}$ and tensor EVD of $\mathcal{B}^{-1/2} *_N \mathcal{M} *_N \mathcal{B}^{-1/2}$ we can write:

$$\mathcal{V}_{\mathcal{A}} *_N \mathcal{D}_{\mathcal{A}}^H *_N \mathcal{D}_{\mathcal{B}\mathcal{M}} *_N \mathcal{D}_{\mathcal{A}} *_N \mathcal{V}_{\mathcal{A}}^H$$
$$= 0_{\mathcal{T}} \quad (\text{as } \mathcal{U} = \mathcal{V}_{\mathcal{K}} = \mathcal{U}_{\mathcal{A}}) \tag{142}$$

This implies that $\mathcal{D}_{\mathcal{A}}^H *_N \mathcal{D}_{\mathcal{B}\mathcal{M}} *_N \mathcal{D}_{\mathcal{A}} = 0_{\mathcal{T}}$ which can be written element-wise as $a_{i_1,\ldots,i_N}^2 \cdot m_{i_1,\ldots,i_N} = 0$. Since $\mathcal{B} \succ 0$ and $\mathcal{M} \succeq 0$, we know $m_{i_1,\ldots,i_N} \geq 0$. So $a_{i_1,\ldots,i_N}^2 = 0$ when $m_{i_1,\ldots,i_N} > 0$, otherwise it is given by (141) with $m_{i_1,\ldots,i_N} = 0$. Together it can be written as

$$a_{i_1,\ldots,i_N}^2 = \left(\frac{1}{\mu} - \frac{1}{k_{i_1,\ldots,i_N}^2}\right)^+ \tag{143}$$

where $(z)^+ = \max\{0, z\}$. From (143) and Proposition 1.1 we get

$$\mathcal{A} *_N \mathcal{A}^H = \mathcal{U}_{\mathcal{A}} *_N \left(\mu^{-1}\mathcal{I}_N - \bar{\mathcal{D}}^{-1}\right)^+ *_N \mathcal{U}_{\mathcal{A}}^H$$
$$= \mathcal{V}_{\mathcal{K}} *_N \left(\mu^{-1}\mathcal{I}_N - \bar{\mathcal{D}}^{-1}\right)^+ *_N \mathcal{V}_{\mathcal{K}}^H \tag{144}$$

where $\mathcal{V}_{\mathcal{K}}$ and $\bar{\mathcal{D}}$ are obtained from tensor EVD of $\mathcal{K}^H *_M \mathcal{K} = \mathcal{V}_{\mathcal{K}} *_N \bar{\mathcal{D}} *_N \mathcal{V}_{\mathcal{K}}^H$. Based on the tensor SVD of $\mathcal{K}$ in Proposition 1.1, we have $\bar{\mathcal{D}} = \mathcal{D}_{\mathcal{K}}^H *_M \mathcal{D}_{\mathcal{K}}$. Substituting (144) into (130), we can conclude that

$$\mathcal{Q} = \mathcal{B}^{-1/2} *_N \mathcal{V}_{\mathcal{K}} *_N \left(\mu^{-1}\mathcal{I}_N - \bar{\mathcal{D}}^{-1}\right)^+ *_N \mathcal{V}_{\mathcal{K}}^H *_N \mathcal{B}^{-1/2} \tag{145}$$

REFERENCES

[1] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Aug. 2009, doi: 10.1137/07070111X.

[2] P. Comon, "Tensors : A brief introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 44–53, May 2014.

[3] L. R. Tucker, "The extension of factor analysis to three-dimensional matrices," in *Contributions to Mathematical Psychology*, H. Gulliksen and N. Frederiksen, Eds. New York, NY, USA: Holt, Rinehart and Winston, 1964, pp. 110–127.

[4] C. J. Appellof and E. R. Davidson, "Strategies for analyzing data from video fluorometric monitoring of liquid chromatographic effluents," *Anal. Chem.*, vol. 53, no. 13, pp. 2053–2056, 1981.

[5] R. Bro, "Review on multiway analysis in chemistry—2000–2005," *Crit. Rev. Anal. Chem.*, vol. 36, nos. 3–4, pp. 279–293, Dec. 2006.

[6] L. Yu, D. Zhang, N. Liu, and W. Zhou, "A multi-view fusion method via tensor learning and gradient descent for image features," *IEEE Access*, vol. 9, pp. 79389–79399, 2021.

[7] J. J. Guerrero, A. C. Murillo, and C. Sagues, "Localization and matching using the planar trifocal tensor with bearing-only data," *IEEE Trans. Robot.*, vol. 24, no. 2, pp. 494–501, Apr. 2008.

[8] X. Li, M. K. Ng, and Y. Ye, "MultiComm: Finding community structure in multi-dimensional networks," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 4, pp. 929–941, Apr. 2014.

[9] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: Models, applications, and scalable algorithms," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 2, pp. 1–44, Jan. 2017.

[10] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.

[11] C. Chatzichristos, E. Kofidis, M. Morante, and S. Theodoridis, "Blind fMRI source unmixing via higher-order tensor decompositions," *J. Neurosci. Methods*, vol. 315, pp. 17–47, Mar. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027018303996

[12] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2014.
[13] P. R. B. Gomes, J. P. C. L. da Costa, A. L. F. de Almeida, and R. T. de Sousa, "Tensor-based multiple denoising via successive spatial smoothing, low-rank approximation and reconstruction for R-D sensor array processing," *Digit. Signal Process.*, vol. 89, pp. 1–7, Jun. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1051200418302847
[14] D. Pandey and H. Leib, "A tensor framework for multi-linear complex MMSE estimation," *IEEE Open J. Signal Process.*, vol. 2, pp. 336–358, 2021.
[15] C. Chen, A. Surana, A. Bloch, and I. Rajapakse, "Multilinear time invariant system theory," in *Proc. Conf. Control Appl.*, 2019, pp. 118–125, doi: 10.1137/1.9781611975758.18.
[16] C. Chen, A. Surana, A. M. Bloch, and I. Rajapakse, "Multilinear control systems theory," *SIAM J. Control Optim.*, vol. 59, no. 1, pp. 749–776, Jan. 2021.
[17] B. W. Bader and T. G. Kolda, "Algorithm 862: MATLAB tensor classes for fast algorithm prototyping," *ACM Trans. Math. Softw.*, vol. 32, no. 4, pp. 635–653, Dec. 2006.
[18] M. Brazell, N. Li, C. Navasca, and C. Tamon, "Solving multilinear systems via tensor inversion," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 2, pp. 542–570, Jan. 2013, doi: 10.1137/100804577.
[19] I. Kisil, G. G. Calvi, B. S. Dees, and D. P. Mandic, "Tensor decompositions and practical applications: A hands-on tutorial," in *Recent Trends in Learning From Data*. Cham, Switzerland: Springer, 2020, pp. 69–97.
[20] H. Chen, F. Ahmad, S. Vorobyov, and F. Porikli, "Tensor decompositions in wireless communications and MIMO radar," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 438–453, Apr. 2021.
[21] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.
[22] J. Gao, O. C. Ozdural, S. H. Ardalan, and H. Liu, "Performance modeling of MIMO OFDM systems via channel analysis," *IEEE Trans. Wireless Commun.*, vol. 5, no. 9, pp. 2358–2362, Sep. 2006.
[23] A. Stamoulis, S. N. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2451–2464, Oct. 2002.
[24] D. Zhang, L. L. Mendes, M. Matthé, I. S. Gaspar, N. Michailow, and G. P. Fettweis, "Expectation propagation for near-optimum detection of MIMO-GFDM signals," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1045–1062, Feb. 2016.
[25] M. Matthe, I. Gaspar, D. Zhang, and G. Fettweis, "Near-ML detection for MIMO-GFDM," in *Proc. IEEE 82nd Veh. Technol. Conf. (VTC-Fall)*, Sep. 2015, pp. 1–2.
[26] M. Caus and A. I. Perez-Neira, "Multi-stream transmission in MIMO-FBMC systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 5041–5045.
[27] W. Su, Z. Safar, and K. J. R. Liu, "Towards maximum achievable diversity in space, time, and frequency: Performance analysis and code design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1847–1857, Jul. 2005.
[28] A. L. F. D. Almeida, G. Favier, and L. R. Ximenes, "Space-time-frequency (STF) MIMO communication systems with blind receiver based on a generalized PARATUCK2 model," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 1895–1909, Apr. 2013.
[29] G. Favier and A. L. F. de Almeida, "Tensor space-time-frequency coding with semi-blind receivers for MIMO wireless communication systems," *IEEE Trans. Signal Process.*, vol. 62, no. 22, pp. 5987–6002, Nov. 2014.
[30] M. N. da Costa, G. Favier, and J. M. T. Romano, "Tensor modelling of MIMO communication systems with performance analysis and Kronecker receivers," *Signal Process.*, vol. 145, pp. 304–316, Apr. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168417304280
[31] G. T. de Araujo, A. L. F. de Almeida, and R. Boyer, "Channel estimation for intelligent reflecting surface assisted MIMO systems: A tensor modeling approach," *IEEE J. Sel. Topics Signal Process.*, vol. 15, no. 3, pp. 789–802, Apr. 2021.
[32] C. Qian, X. Fu, N. D. Sidiropoulos, and Y. Yang, "Tensor-based channel estimation for dual-polarized massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6390–6403, Dec. 2018.
[33] D. C. Araújo, A. L. De Almeida, J. P. Da Costa, and R. T. de Sousa, "Tensor-based channel estimation for massive MIMO-OFDM systems," *IEEE Access*, vol. 7, pp. 42133–42147, 2019.
[34] Z. Lin, T. Lv, J. A. Zhang, and R. P. Liu, "Tensor-based high-accuracy position estimation for 5G mmWave massive MIMO systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
[35] Y. Lin, S. Jin, M. Matthaiou, and X. You, "Tensor-based channel estimation for millimeter wave MIMO-OFDM with dual-wideband effects," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4218–4232, Jul. 2020.
[36] J. Du, M. Han, L. Jin, Y. Hua, and X. Li, "Semi-blind receivers for multi-user massive MIMO relay systems based on block tucker2-PARAFAC tensor model," *IEEE Access*, vol. 8, pp. 32170–32186, 2020.
[37] A. Venugopal and H. Leib, "A tensor based framework for multi-domain communication systems," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 606–633, 2020.
[38] H. Gacanin, A. Ligata, and Y. Chahibi, "On multiple-domain cooperative diversity for communications with distributed content in G.Hn networks," *Int. J. Commun. Syst.*, vol. 30, no. 1, Jan. 2017, Art. no. e2956.
[39] K. Cain, V. Vakilian, and R. Abdolee, "Low-complexity universal-filtered multi-carrier for beyond 5G wireless systems," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Mar. 2018, pp. 254–258.
[40] E. C. Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez, D. Ktenas, N. Cassiau, L. Maret, and C. Dehos, "6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 42–50, Sep. 2019.
[41] P. Comon, "Tensors versus matrices usefulness and unexpected properties," in *Proc. IEEE/SP 15th Workshop Stat. Signal Process.*, Aug. 2009, pp. 781–788.
[42] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.
[43] H. Bölcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, Feb. 2002.
[44] L.-B. Cui, C. Chen, W. Li, and M. K. Ng, "An eigenvalue problem for even order tensors with its applications," *Linear Multilinear Algebra*, vol. 64, no. 4, pp. 602–621, Apr. 2016, doi: 10.1080/03081087.2015.1071311.
[45] D. Pandey and H. Leib, "A tensor based precoder and receiver for MIMO GFDM systems," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6.
[46] D. Pandey, A. Venugopal, and H. Leib, "Multi-domain communication systems and networks: A tensor-based approach," *Network*, vol. 1, no. 2, pp. 50–74, Jul. 2021. [Online]. Available: https://www.mdpi.com/2673-8732/1/2/5
[47] D. Pandey and H. Leib, "Shannon capacity of tensor channels under a family of power constraints," in *Proc. 30th Biennial Symp. Commun. (BSC)*, Jun. 2021, pp. 1–6.
[48] L. De Lathauwer, J. Castaing, and J.-F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2965–2973, Jun. 2007.
[49] M.-L. Liang, B. Zheng, and R.-J. Zhao, "Tensor inversion and its application to the tensor equations with Einstein product," *Linear Multilinear Algebra*, vol. 67, no. 4, pp. 843–870, Apr. 2019, doi: 10.1080/03081087.2018.1500993.
[50] L. de Lathauwer, N. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, Jul. 2000, doi: 10.1137/S0895479896305696.
[51] Q.-W. Wang and X. Xu, "Iterative algorithms for solving some tensor equations," *Linear Multilinear Algebra*, vol. 67, no. 7, pp. 1325–1349, Jul. 2019, doi: 10.1080/03081087.2018.1452889.
[52] W. Nicholson, *Linear Algebra with Applications*. New York, NY, USA: McGraw-Hill, 2009.
[53] L. Sun, B. Zheng, C. Bu, and Y. Wei, "Moore–Penrose inverse of tensors via Einstein product," *Linear Multilinear Algebra*, vol. 64, no. 4, pp. 686–698, 2016, doi: 10.1080/03081087.2015.1083933.
[54] A. Zhang and D. Xia, "Tensor SVD: Statistical and computational limits," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7311–7338, Nov. 2018.
[55] L. Qi, "Eigenvalues of a real supersymmetric tensor," *J. Symbolic Comput.*, vol. 40, no. 6, pp. 1302–1324, 2005.
[56] L. Qi, H. Chen, and Y. Chen, *Tensor Eigenvalues Their Application*, vol. 39. Singapore: Springer, 2018.
[57] L.-H. Lim, "Singular values and eigenvalues of tensors: A variational approach," in *Proc. 1st Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, 2005, pp. 129–132.

[58] K. B. Petersen and M. S. Pedersen. (Nov. 2012). *The Matrix Cookbook*. [Online]. Available: http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html

[59] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 141–154, Jan. 2006.

[60] A. Feiten, S. Hanly, and R. Mathar, "Derivatives of mutual information in Gaussian vector channels with applications," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2007, pp. 2296–2300.

[61] H. L. Van Trees, *Optimum Array Processing: Part IV Detection, Estimation, Modulation Theory*. Hoboken, NJ, USA: Wiley, 2004.

[62] F. Hlawatsch and G. Matz, *Wireless Communications Over Rapidly Time-Varying Channels*. New York, NY, USA: Academic, 2011.

[63] N. Costa and S. Haykin, "A novel wideband MIMO channel model and experimental validation," *IEEE Trans. Antennas Propag.*, vol. 56, no. 2, pp. 550–562, Feb. 2008.

[64] L. Hanzo, Y. Akhtman, J. Akhtman, L. Wang, and M. Jiang, *MIMO-OFDM for LTE, WiFi WiMAX: Coherent Versus Non-Coherent Cooperative Turbo Transceivers*. Hoboken, NJ, USA: Wiley, 2010.

[65] Y. J. Zhang and K. B. Letaief, "An efficient resource-allocation scheme for spatial multiuser access in MIMO/OFDM systems," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 107–116, Jan. 2005.

[66] Y. Tan and Q. Chang, "Multi-user MIMO-OFDM with adaptive resource allocation over frequency selective fading channel," in *Proc. 4th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Oct. 2008, pp. 1–5.

[67] A. Tolli and M. Juntti, "Efficient user, bit and power allocation for adaptive multiuser MIMO-OFDM with low signalling overhead," in *Proc. IEEE Int. Conf. Commun.*, 2006, pp. 5360–5365.

[68] C. Leung, S. Huberman, K. Ho-Van, and T. Le-Ngoc, "Vectored DSL: Potential, implementation issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1907–1923, 4th Quart., 2013.

[69] S. Serbetli and A. Yener, "MIMO-CDMA systems: Signature and beamformer design with various levels of feedback," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2758–2772, Jul. 2006.

[70] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.

[71] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[72] M. R. D. Rodrigues, F. Perez-Cruz, and S. Verduy, "Multiple-input multiple-output Gaussian channels: Optimal covariance for non-Gaussian inputs," in *Proc. IEEE Inf. Theory Workshop*, May 2008, pp. 445–449.

[73] C. Xing, Y. Jing, S. Wang, J. Wang, S. Xin Ng, S. Chen, and L. Hanzo, "Unified framework of KKT conditions based matrix optimizations for MIMO communications," 2017, *arXiv:1711.04449*.

[74] C. Xing, Z. Fei, Y. Zhou, and Z. Pan, "Matrix-field water-filling architecture for MIMO transceiver designs with mixed power constraints," in *Proc. IEEE 26th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Aug. 2015, pp. 392–396.

[75] M. Grant and S. Boyd. (Mar. 2014). *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.1*. [Online]. Available: http://cvxr.com/cvx

[76] A. E. Youssef, "Exploring cloud computing services and applications," *J. Emerg. Trends Comput. Inf. Sci.*, vol. 3, no. 6, pp. 838–847, Jul. 2012.

[77] S. Garg, S. Versteeg, and R. Buyya, "A framework for ranking of cloud computing services," *Future Gener. Comput. Syst.*, vol. 29, no. 4, pp. 1012–1023, 2013.

[78] A. Fang, L. Cui, Z. Zhang, C. Chen, and Z. Sheng, "A parallel computing framework for cloud services," in *Proc. IEEE Int. Conf. Adv. Electr. Eng. Comput. Appl. (AEECA)*, Aug. 2020, pp. 832–835.

[79] A. M. Mahmood, A. Al-Yasiri, and O. Y. K. Alani, "A new processing approach for reducing computational complexity in cloud-RAN mobile networks," *IEEE Access*, vol. 6, pp. 6927–6946, 2018.

[80] C. Fan, Y. J. Zhang, and X. Yuan, "Dynamic nested clustering for parallel PHY-layer processing in cloud-RANs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1881–1894, Mar. 2016.

[81] D. Pandey and H. Leib, "Tensor multi-linear MMSE estimation using the Einstein product," in *Advances in Information and Communication. FICC 2021*. Cham, Switzerland: Springer, 2021, pp. 47–64.

[82] B. Huang and W. Li, "Numerical subspace algorithms for solving the tensor equations involving Einstein product," *Numer. Linear Algebra Appl.*, vol. 28, no. 2, p. e2351, Mar. 2021.

[83] R. Behera and D. Mishra, "Further results on generalized inverses of tensors via the Einstein product," *Linear Multilinear Algebra*, vol. 65, no. 8, pp. 1662–1682, Aug. 2017.

[84] K. Panigrahy and D. Mishra, "Extension of Moore–Penrose inverse of tensor via Einstein product," *Linear Multilinear Algebra*, vol. 3, pp. 1–24, Apr. 2020, doi: 10.1080/03081087.2020.1748848.

[85] B. Huang, "Numerical study on Moore-penrose inverse of tensors via Einstein product," *Numer. Algorithms*, vol. 87, no. 4, pp. 1767–1797, Aug. 2021.

[86] J. K. Sahoo and R. Behera, "Reverse-order law for core inverse of tensors," *Comput. Appl. Math.*, vol. 39, no. 2, pp. 1–22, May 2020.

[87] H. Liu, L. T. Yang, J. Ding, Y. Guo, and S. S. Yau, "Tensor-train-based high-order dominant eigen decomposition for multimodal prediction services," *IEEE Trans. Eng. Manag.*, vol. 68, no. 1, pp. 197–211, Feb. 2021.

[88] M. Vu, "MIMO capacity with per-antenna power constraint," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2011, pp. 1–5.

[89] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[90] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[91] A. G. Burr, "Capacity bounds and estimates for the finite scatterers MIMO wireless channel," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 812–818, Jun. 2003.

[92] S. Loyka and G. Levin, "On physically-based normalization of MIMO channel matrices," *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1107–1112, Mar. 2009.

[93] G. Favier, C. A. R. Fernandes, and A. L. F. de Almeida, "Nested tucker tensor decomposition with application to MIMO relay systems using tensor space–time coding (TSTC)," *Signal Process.*, vol. 128, pp. 318–331, Nov. 2016.

[94] C. F. Caiafa and A. Cichocki, "Block sparse representations of tensors using Kronecker bases," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 2709–2712.

[95] H. Zhang and F. Ding, "On the Kronecker products and their applications," *J. Appl. Math.*, vol. 2013, pp. 1–8, Jan. 2013, doi: 10.1155/2013/296185.

[96] P. D. Hoff, "Separable covariance arrays via the tucker product, with applications to multivariate relational data," *Bayesian Anal.*, vol. 6, no. 2, pp. 179–196, Jun. 2011.

[97] N. Costa and S. Haykin, *Multiple-Input Multiple-Output Channel Models: Theory Practice*. Hoboken, NJ, USA: Wiley, 2010.

[98] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[99] L. Hanlen and A. Grant, "Capacity analysis of correlated MIMO channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6773–6787, Nov. 2012.

[100] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A model-driven deep learning network for MIMO detection," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 584–588.

[101] G. Yang, H. Zhang, Z. Shi, S. Ma, and H. Wang, "Asymptotic outage analysis of spatially correlated Rayleigh MIMO channels," *IEEE Trans. Broadcast.*, vol. 67, no. 1, pp. 263–278, Mar. 2021.

[102] H. Shin, M. Z. Win, J. H. Lee, and M. Chiani, "On the capacity of doubly correlated MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2253–2265, Aug. 2006.

[103] N. Michailow, M. Matthe, I. S. Gaspar, A. N. Caldevilla, L. L. Mendes, A. Festag, and G. Fettweis, "Generalized frequency division multiplexing for 5th generation cellular networks," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3045–3061, Sep. 2014.

[104] S. A. Jafar and A. Goldsmith, "Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990–997, May 2005.

[105] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for Gaussian vector multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[106] M. Kiamari and A. S. Avestimehr, "Capacity region of the symmetric injective K-user deterministic interference channel," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, pp. 4010–4022, Mar. 2019.

[107] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "Capacity regions and sum-rate capacities of vector Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5030–5044, Oct. 2010.

[108] X. Shang, B. Chen, and M. J. Gans, "On the achievable sum rate for MIMO interference channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4313–4320, Sep. 2006.
[109] X. Shang, B. Chen, G. Kramer, and H. V. Poor, "On the capacity of MIMO interference channels," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2008, pp. 700–707.
[110] X. Shang and H. V. Poor, "Noisy-interference sum-rate capacity for vector Gaussian interference channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 132–153, Jan. 2013.
[111] X. Shang and H. V. Poor, "Capacity region of vector Gaussian interference channels with generally strong interference," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3472–3496, Jun. 2012.
[112] C. Geng, N. Naderializadeh, S. Avestimehr, and S. A. Jafar, "On the optimality of treating interference as noise," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1753–1767, Apr. 2015.
[113] A. Ghasemi, A. S. Motahari, and A. K. Khandani, "Interference alignment for the k user MIMO interference channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2010, pp. 360–364.
[114] V. Vasudevan and M. Ramakrishna, "A hierarchical singular value decomposition algorithm for low rank matrices," 2017, *arXiv:1710.02812*.
[115] J. W. Demmel, *Applied Numerical Linear Algebra*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 1997. [Online]. Available: https://epubs.siam.org/doi/abs/10.1137/1.9781611971446

DIVYANSHU PANDEY (Graduate Student Member, IEEE) received the B.Tech. degree in communication and computer engineering from The LNM Institute of Information Technology, Jaipur, India, in 2011, and the M.S. degree in electrical engineering from the University of Minnesota, Twin Cities, USA, in 2014. He is currently pursuing the Ph.D. degree in electrical engineering with McGill University, Montreal, QC, Canada. He worked as a Systems Engineer at WLAN PHY Research and Development Team, Marvell Semiconductors Inc., Santa Clara, CA, USA, from February 2015 to August 2017. His research interests include information theory, wireless communications, and tensor algebra with applications to communications and signal processing.

HARRY LEIB (Life Senior Member, IEEE) received the B.Sc. *(cum laude)* and M.Sc. degrees in electrical engineering from the Technion—Israel Institute of Technology, Israel, in 1977 and 1984, respectively, and the Ph.D. degree in electrical engineering from the University of Toronto, Canada, in 1987.

From 1977 to 1984, he was with the Israel Ministry of Defense, working in communication systems. After completing his Ph.D. studies, he was with the University of Toronto as a Postdoctoral Research Associate and an Assistant Professor. Since September 1989, he has been with the Department of Electrical and Computer Engineering, McGill University, where he is now a Full Professor. His current research interests include digital communications, wireless communication systems, global navigation satellite systems, detection, estimation, and information theory.

Dr. Leib was an Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS (2000–2013) and an Associate Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (2001–2007). He was a Guest Co-Editor for special issues of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on "Differential and Noncoherent Wireless Communication" (2003–2005) and on "Spectrum and Energy Efficient Design of Wireless Communication Networks" (2012–2013). Since 2017, he has been the founding Editor-in-Chief of *AIMS Electronics and Electrical Engineering* journal.