Statistical Eigenmode Transmission Over Jointly Correlated MIMO Channels

Xiqi Gao, *Senior Member, IEEE*, Bin Jiang, *Student Member, IEEE*, Xiao Li, *Student Member, IEEE*, Alex B. Gershman, *Fellow, IEEE*, and Matthew R. McKay, *Member, IEEE*

Abstract—We investigate multiple-input multiple-output (MIMO) eigenmode transmission using statistical channel state information at the transmitter. We consider a general jointly correlated MIMO channel model, which does not require separable spatial correlations at the transmitter and receiver. For this model, we first derive a closed-form tight upper bound for the ergodic capacity, which reveals a simple and interesting relationship in terms of the matrix permanent of the eigenmode channel coupling matrix and embraces many existing results in the literature as special cases. Based on this closed-form and tractable upper bound expression, we then employ convex optimization techniques to develop low-complexity power allocation solutions involving only the channel statistics. Necessary and sufficient optimality conditions are derived, from which we develop an iterative water-filling algorithm with guaranteed convergence. Simulations demonstrate the tightness of the capacity upper bound and the near-optimal performance of the proposed low-complexity transmitter optimization approach.

Index Terms—Capacity bound, convex optimization, ergodic capacity, multiple-input multiple-output (MIMO) channel, permanents, power allocation, statistical eigenmode transmission.

I. INTRODUCTION

MULTIPLE-input multiple-output (MIMO) wireless systems, equipped with multiple antennas at both the transmitter and the receiver, have attracted tremendous interest in recent years as a means of enabling substantially increased link capacity and reliability compared with conventional systems [1]–[6]. The performance of practical MIMO systems is characterized by various system parameters, such as the average transmit power and the transmit–receive antenna configurations, as well as various channel phenomena such as spatial correlation, line-of-sight components, thermal noise, interference, and Doppler effects due to mobility. Each of these factors has an impact on the MIMO channel capacity.

In realistic environments, where the channel characteristics may vary significantly over time, substantial MIMO capacity benefits can be obtained by tracking the states of the fading channels, and using this channel state information (CSI) to optimally adapt the MIMO transceiver parameters. However, such closed-loop MIMO strategies require both the transmitter and receiver to acquire some form of CSI. Whilst it is reasonable to assume that the instantaneous CSI can be obtained accurately at the receiver through channel estimation, whether or not this information can be obtained at the transmitter depends highly on the application scenario. For example, for fixed or low mobility applications, the channel conditions vary relatively slowly, in which case instantaneous CSI can be fed to the transmitter via well-designed feedback channels in frequency division duplex (FDD) systems, or using the reciprocity of uplink and downlink in time division duplex (TDD) systems. However, as the mobility and hence the fading rate increases, obtaining accurate instantaneous CSI at the transmitter becomes much more difficult. For such a scenario, it is reasonable to exploit *statistical* CSI at the transmitter. The motivation for this approach stems from the fact that the channel statistics vary over much larger time scales than the instantaneous channel gains, and the uplink and downlink statistics are usually reciprocal in both FDD and TDD systems[7], [8]. Therefore, the statistical information can be easily obtained by exploiting reciprocity, or by employing feedback channels with significantly lower bandwidth compared with instantaneous CSI feedback systems. In addition, transceiver designs based on statistical information are typically more robust to imperfections, such as delays, in the feedback channel.

Capacity analysis and transceiver designs using the statistical CSI at the transmitter are highly dependant on the channel modeling. The most common approach has been to adopt the popular Kronecker model [9]–[19], where the correlation between the fading of two distinct antenna pairs is the product of the corresponding transmit and receive correlations [9], [20]. The primary advantage of this separable model is that it is analytically friendly, however various measurement campaigns have demonstrated that it can have deficiencies in practice [21], [22]. To overcome these deficiencies, more generalized channel models have been proposed, including the virtual channel representation [23], [24], the unitary-independent-unitary (UIU) model [25], [26], and Weichselberger's model [21]. In contrast to the Kronecker model, these are *jointly correlated* channel models

which not only account for the correlation at both link ends, but also characterize their mutual dependence.

Under various assumptions on the system configuration and channel model, several important works have been reported on the transmitter optimization problem using the statistical CSI at the transmitter in recent years. In particular, for multi-input and single-output (MISO) wireless channels with correlated Rayleigh or uncorrelated Rician entries, it was shown in [27] that the capacity-achieving strategy is to send independent data streams in the directions defined by the dominant eigenvectors of the transmit correlation matrix. This result was extended to Rayleigh fading MIMO channels with Kronecker correlation structure in [15]–[18], to uncorrelated Rician MIMO channels in [28], [29], to the UIU model in [25], [26], and to the virtual channel representation in [24]. These prior contributions have also considered the task of optimally allocating power across the transmit eigendirections (i.e., defining the eigenvalues of the optimal transmit covariance matrix), for maximizing capacity. However, in most cases, the power allocation problem has been tackled by optimizing the exact ergodic capacity expression, and this approach has led to computationally involved numerical optimization procedures. For example, see [30], [31] for Kronecker channels, and [26] for jointly correlated UIU channels. In these contributions, iterative power allocation approaches were presented which involved numerical averaging over channel samples for each iteration of the algorithm.

In this paper, we investigate the statistical eigenmode transmission (SET) over a general jointly correlated MIMO channel, using the statistical CSI at the transmitter. Our idea is to first derive a *closed-form* tight upper bound for the ergodic capacity of the general jointly correlated MIMO channel model. This upper bound expression reveals a simple and interesting relationship in terms of matrix permanents, and embraces many existing results in the literature as special cases, such as those presented for Kronecker channels in [10], [13], [14], [17], [18]. Based on this closed-form and tractable upper bound expression, we then employ convex optimization techniques to develop low-complexity power allocation solutions in terms of only the channel statistics. We derive necessary and sufficient optimality conditions, and propose a simple computation algorithm, inspired by the iterative water-filling techniques presented previously for transmitter optimization of multiuser systems [32], [33], which is shown to converge within only a few iterations. Numerical simulations demonstrate the tightness of the capacity upper bound and the near-optimal performance of the proposed low-complexity transmitter optimization approach, i.e., suffering negligible loss with respect to the ergodic capacity of the jointly correlated MIMO channel.

A. Notation

The following notation is adopted throughout the paper: Upper (lower) bold-face letters are used to denote matrices (column vectors); in some cases, where it is not clear, we will employ subscripts to emphasize dimensionality. The $N \times N$ identity matrix is denoted by $\mathbf{I}_N$, the all-zero matrix is denoted by $\mathbf{0}$, and the all-one matrix is denoted by $\mathbf{1}$. The superscripts $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^*$ stand for the conjugate-transpose, transpose, and conjugate operations, respectively. We employ $\mathbb{E}\{\cdot\}$

to denote expectation with respect to all random variables within the brackets, and use $\mathbf{A} \odot \mathbf{B}$ to denote the Hadamard product of the two matrices $\mathbf{A}$ and $\mathbf{B}$. We use $[\mathbf{A}]_{kl}$ or the lower-case representation $a_{k,l}$ to denote the (k, l)th entry of the matrix $\mathbf{A}$, and a_k denotes the kth entry of the column vector $\mathbf{a}$. The operators $\mathrm{tr}(\cdot)$, $\det(\cdot)$, and $\mathrm{Per}(\cdot)$ represent the matrix trace, determinant, and permanent, respectively, and $\mathrm{diag}(\mathbf{x})$ denotes a diagonal matrix with $\mathbf{x}$ along its main diagonal.

We will use $\mathcal{S}_N^k$ to denote the set of all size-k permutations of the numbers $\{1, 2, \ldots, N\}$, where $k \leq N$. By using the notation $\hat{\alpha}_k \in \mathcal{S}_N^k$, we mean that $\hat{\alpha}_k = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, $\alpha_i \in \{1, 2, \ldots, N\}$ for $1 \leq i \leq k$, and $\alpha_i \neq \alpha_j$ for $1 \leq i, j \leq k$ and $i \neq j$. We will use $\mathcal{S}_N^{(k)}$ to denote the set of all ordered length-k subsets of the numbers $\{1, 2, \ldots, N\}$. By the notation $\hat{\alpha}_k \in \mathcal{S}_N^{(k)}$, we mean that $\hat{\alpha}_k = (\alpha_1, \alpha_2, \ldots, \alpha_k)$, $\alpha_i \in \{1, 2, \ldots, N\}$ for $1 \leq i \leq k$, and $\alpha_1 < \alpha_2 < \cdots < \alpha_k$. The cardinalities of the sets $\mathcal{S}_N^k$ and $\mathcal{S}_N^{(k)}$ are $\frac{N!}{(N-k)!}$ and $\frac{N!}{k!(N-k)!}$ respectively.

With $\hat{\alpha}_k$ and $\hat{\beta}_k$ defined as above, we will use $\mathbf{A}_{\hat{\beta}_k}^{\hat{\alpha}_k}$ to denote the sub-matrix of an $M \times N$ matrix $\mathbf{A}$ obtained by selecting the rows and columns indexed by $\hat{\alpha}_k$ and $\hat{\beta}_k$ respectively. $\mathbf{A}^{\hat{\alpha}_N}$ will denote the sub-matrix of $\mathbf{A}$ obtained by selecting the rows indexed by $\hat{\alpha}_N$ when $M \geq N$, and $\mathbf{A}_{\hat{\beta}_M}$ the sub-matrix of $\mathbf{A}$ obtained by selecting the columns indexed by $\hat{\beta}_M$ when $M \leq N$. Also, we will use $\hat{\alpha}_k'$ and $\hat{\beta}_k'$ to denote the sequences complementary to $\hat{\alpha}_k$ and $\hat{\beta}_k$ in $\{1, 2, \ldots, M\}$ and $\{1, 2, \ldots, N\}$, respectively. As such, $\mathbf{A}_{\hat{\beta}_k'}^{\hat{\alpha}_k'}$ will represent the sub-matrix of $\mathbf{A}$ obtained by deleting the rows and columns indexed by $\hat{\alpha}_k$ and $\hat{\beta}_k$, respectively. Finally, for notational convenience, we will use $\mathbf{A}_{i,j}$ to represent the sub-matrix of $\mathbf{A}$ obtained by deleting its ith row and jth column.

II. CHANNEL MODEL AND STATISTICAL EIGENMODE TRANSMISSION

A. Channel Model

We consider a single-user MIMO link with N_t transmit and N_r receive antennas, operating over a frequency-flat fading channel. The N_r-dimensional complex baseband received signal vector for a single symbol interval can be written as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$ is the $N_t \times 1$ transmitted signal vector, $\mathbf{H}$ is the $N_r \times N_t$ channel matrix with (i, j)th element representing the complex fading coefficient between the jth transmit and ith receive antenna, and $\mathbf{n}$ is the $N_r \times 1$ zero-mean additive complex Gaussian noise vector with $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = \sigma_n^2 \mathbf{I}_{N_r}$. It is assumed that $\mathbf{x}$ and $\mathbf{H}$ satisfy the following power constraints:

$$\mathbb{E}\{\mathrm{tr}(\mathbf{x}\mathbf{x}^H)\} = P \tag{2}$$

$$\mathbb{E}\{\mathrm{tr}(\mathbf{H}\mathbf{H}^H)\} = N_t N_r. \tag{3}$$

We define the transmit signal to noise ratio (SNR) as $\rho = P/\sigma_n^2$. If the total transmitted power P is equally distributed across all transmit antennas, so that $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = (P/N_t)\mathbf{I}_{N_t}$, then ρ also corresponds to the average SNR per receive antenna.

For the jointly correlated MIMO channel which we consider throughout this paper, the channel matrix $\mathbf{H}$ is given by

$$\mathbf{H} = \mathbf{U}_r \tilde{\mathbf{H}} \mathbf{U}_t^H$$

$$= \mathbf{U}_r \left(\mathbf{D} + \mathbf{M} \odot \mathbf{H}_{iid} \right) \mathbf{U}_t^H \qquad (4)$$

where $\tilde{\mathbf{H}} = \mathbf{D} + \mathbf{M} \odot \mathbf{H}_{iid}$, $\mathbf{U}_t$ and $\mathbf{U}_r$ are $N_t \times N_t$ and $N_r \times N_r$ deterministic unitary matrices, $\mathbf{D}$ is an $N_r \times N_t$ deterministic matrix with at most one nonzero element in each row and each column, $\mathbf{M}$ is an $N_r \times N_t$ deterministic matrix with nonnegative elements, and $\mathbf{H}_{\text{i.i.d.}}$ is an $N_r \times N_t$ random matrix with elements having zero mean and independent identical distributions (i.i.d.). Note that we do not constrain the elements of $\mathbf{H}_{iid}$ to be Gaussian. Without any loss of generality, we can assume that the nonzero elements of $\mathbf{D}$ are real, with indices (i, i) for $1 \leq i \leq \min(N_t, N_r)$. Let us define

$$\mathbf{\Omega} = \mathbb{E}\{\tilde{\mathbf{H}} \odot \tilde{\mathbf{H}}^*\}. \qquad (5)$$

It is easy to show that

$$\mathbf{\Omega} = \mathbf{D} \odot \mathbf{D} + \mathbf{M} \odot \mathbf{M} \qquad (6)$$

and the power constraint (3) can be rewritten as

$$\sum_{i=1}^{N_r} \sum_{j=1}^{N_t} [\mathbf{\Omega}]_{ij} = N_t N_r. \qquad (7)$$

From (4), the transmit and receive correlation matrices can be expressed as

$$\mathbf{R}_t = \mathbb{E}\{\mathbf{H}^H \mathbf{H}\} = \mathbf{U}_t \mathbf{\Lambda}_t \mathbf{U}_t^H \qquad (8)$$

$$\mathbf{R}_r = \mathbb{E}\{\mathbf{H} \mathbf{H}^H\} = \mathbf{U}_r \mathbf{\Lambda}_r \mathbf{U}_r^H \qquad (9)$$

where $\mathbf{\Lambda}_t$ and $\mathbf{\Lambda}_r$ are diagonal matrices with $[\mathbf{\Lambda}_t]_{ii} = \sum_{j=1}^{N_r} [\mathbf{\Omega}]_{ji}$ and $[\mathbf{\Lambda}_r]_{ii} = \sum_{j=1}^{N_t} [\mathbf{\Omega}]_{ij}$. This implies that in the channel model defined in (4), $\mathbf{U}_t$ and $\mathbf{U}_r$ are the eigenvector matrices of the transmit and receive correlation matrices, respectively. These matrices are characterized by the transmit and receive antenna configurations. For example, when uniform linear arrays (ULA) are employed at both the transmitter and receiver, it is shown in [23] that the eigenvector matrices can be set to discrete Fourier transform (DFT) matrices.

The statistics of $\tilde{\mathbf{H}} = \mathbf{U}_r^H \mathbf{H} \mathbf{U}_t$ characterize realistic propagation environments. From (4) and (6), we have

$$\mathbf{D} = \mathbb{E}\{\tilde{\mathbf{H}}\} \qquad (10)$$

$$[\mathbf{M}]_{ij} = \sqrt{\text{var}\{[\tilde{\mathbf{H}}]_{ij}\}}$$

$$= \sqrt{[\mathbf{\Omega}]_{ij} - [\mathbf{D}]_{ij}^2} \qquad (11)$$

where $\text{var}\{\cdot\}$ denotes variance. The matrices $\mathbf{D}$ and $\mathbf{M}$ reflect the line-of-sight (LOS) and scattering components of the channel, respectively. The (i, j)th element of $\mathbf{\Omega}$, i.e., $[\mathbf{\Omega}]_{ij}$, corresponds to the average power of $[\tilde{\mathbf{H}}]_{ij}$ and captures the average coupling between the ith receive eigenmode and jth transmit eigenmode. For this reason, we refer to $\mathbf{\Omega}$ as the eigenmode channel coupling matrix. It can be seen that the eigenvalues of the transmit and receive correlation matrices are summations of the elements of the matrix $\mathbf{\Omega}$ in each column and each row, respectively. These eigenvalues are *nonseparable*, which reflects the joint correlation feature of the channel.

The channel model described by (4) provides a general formula which embraces many existing channel models [9], [20]–[26]. For example, if $\mathbf{D} = 0$, $\mathbf{M}$ is a rank-one matrix, and $\mathbf{H}_{iid}$ has Rayleigh-faded elements, then (4) reduces to the popular separable-correlation Kronecker model [9], [20]. By allowing $\mathbf{M}$ to have arbitrary rank and fixing $\mathbf{U}_t$ and $\mathbf{U}_r$ to be DFT matrices, one can achieve the virtual channel representation for ULAs [23]. If we further allow $\mathbf{U}_t$ and $\mathbf{U}_r$ to be arbitrary unitary matrices, one can obtain Weichselberger's channel model[21]. Moreover, by setting $\mathbf{D} = 0$ we arrive at the unitary-independent-unitary (UIU) model introduced in [25], [26]. Our model is also related to the model in [24], where one LOS component was included in the virtual channel representation for the ULA MIMO channels. Here, we allow multiple LOS components in eigenmode to cover more general transmission links, such as those in distributed radio networks [34].

B. Statistical Eigenmode Transmission

Throughout the paper, we assume that the receiver knows the channel perfectly, whilst the transmitter only has access to the statistical parameters, including $\mathbf{U}_t$, $\mathbf{U}_r$, $\mathbf{D}$ and $\mathbf{M}$ (and thus $\mathbf{\Omega}$). Under these assumptions, the ergodic capacity of the MIMO channel is achieved by selecting the transmitted signal vector $\mathbf{x}$ to have zero mean and to follow a proper Gaussian distribution [1]. Let the covariance matrix of $\mathbf{x}$ be $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = (P/N_t) \mathbf{Q}$. Then the power constraint on $\mathbf{x}$ can be rewritten as $\text{tr}(\mathbf{Q}) = N_t$, and the ergodic capacity is given by

$$C = \max_{\mathbf{Q} \succeq \mathbf{0}, \text{tr}(\mathbf{Q})=N_t} \mathbb{E}\left\{ \log \det \left(\mathbf{I}_{N_r} + \gamma \mathbf{H} \mathbf{Q} \mathbf{H}^H \right) \right\} \qquad (12)$$

where $\gamma = \frac{\rho}{N_t}$. Substituting (4) into (12) yields

$$C = \max_{\tilde{\mathbf{Q}} \succeq \mathbf{0}, \text{tr}(\tilde{\mathbf{Q}})=N_t} \mathbb{E}\left\{ \log \det \left(\mathbf{I}_{N_r} + \gamma \tilde{\mathbf{H}} \tilde{\mathbf{Q}} \tilde{\mathbf{H}}^H \right) \right\} \qquad (13)$$

where $\tilde{\mathbf{Q}} = \mathbf{U}_t^H \mathbf{Q} \mathbf{U}_t$. Let $\mathbf{Q} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$, where $\mathbf{U}$ is the eigenvector matrix, and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_{N_t})$ is a diagonal matrix of the corresponding eigenvalues. When $\tilde{\mathbf{H}}$ has independent and symmetrically distributed elements, it has been shown in [26] and [24] that the optimal eigenvector matrix $\mathbf{U}$ for achieving the capacity is $\mathbf{U} = \mathbf{U}_t$, and thus $\tilde{\mathbf{Q}}$ is diagonal. In [24], it has been pointed out that this solution also applies when one element of $\tilde{\mathbf{H}}$ contains a LOS component. We note however, that the channel model given by (4) allows for multiple possible LOS components. For this more general case, one can arrive at the following result.

Theorem 1: The eigenvector matrix of the capacity-achieving matrix $\mathbf{Q}$ for the jointly correlated channel (4) is given by $\mathbf{U} = \mathbf{U}_t$. The ergodic capacity can therefore be expressed as

$$C = \max_{\boldsymbol{\lambda} \geq \mathbf{0}, \sum_{i=1}^{N_t} \lambda_i = N_t} \mathbb{E}\left\{ \log \det \left(\mathbf{I}_{N_r} + \gamma \tilde{\mathbf{H}} \, \text{diag}(\boldsymbol{\lambda}) \tilde{\mathbf{H}}^H \right) \right\} \qquad (14)$$

where $\boldsymbol{\lambda}$ is an $N_t \times 1$ vector containing the eigenvalues λ_i, $i = 1, 2, \ldots, N_t$.

The proof follows similar approaches to those used in [24], [26], [27], [29] and is therefore omitted. Theorem 1 demonstrates that the optimal signaling directions are defined by the eigenvectors of the transmit correlation matrix of the MIMO

channel. This agrees with and extends prior results in the literature to the more general channel model given by (4). For the transmitter optimization problem, the major remaining challenge is to determine the eigenvalues of the capacity-achieving input covariance matrix $\mathbf{Q}$. This is equivalent to the task of optimally allocating the available transmit power budget over the optimized transmit eigen-directions, determined in Theorem 1.

In general, it is very difficult to derive an exact closed-form solution for the power allocation problem. A major source of this difficulty is due to the complexity in evaluating tractable closed-form solutions for the expectation in (14). This is also the case for many other less general MIMO channel scenarios, such as the popular Kronecker correlation model [15], [16]. As such, the standard approach has been to develop numerical optimization techniques (see e.g., [30] and [31]).

In this paper, considering the general jointly correlated MIMO channel model, we develop a new approach which leads to the design of simple, robust and practical power allocation solutions. In particular, our approach is based on deriving a tight closed-form upper bound on the expectation in (14) which can then serve as an approximation to the capacity. Based on this expression, we are then able to derive new optimized power allocation solutions which are simple and fast to compute. These power allocation solutions will be shown to serve as very accurate approximations to the optimal capacity-achieving solution, with low computational complexity requirements.

We note that the power-allocation problem for jointly correlated channel scenarios has also been considered in [26], where necessary and sufficient conditions as well as an iterative numerical algorithm were proposed. One drawback of that algorithm is that for each iteration it requires numerically averaging certain random matrix structures involving the inverse of instantaneous realizations of the MIMO channel. Moreover, since the computation algorithm requires access to instantaneous MIMO CSI, then under the statistical-feedback assumption, such power-allocation computations must be typically performed at the receiver. In contrast, in this paper we develop more practically appealing power-allocation algorithms which involve only the channel statistics. As such, they are simpler and more efficient to compute, since they do not require random matrix averaging during the power-allocation computation. Moreover, our new power-allocation algorithm has the additional advantage of permitting computation at either the receiver or the transmitter. This extra flexibility is particularly important for various practical applications, such as downlink transmission where it is often desirable or necessary to restrict computations to the base station.

We start by rewriting the ergodic capacity (14) as

$$C = \max_{\boldsymbol{\lambda} \geq \mathbf{0}, \sum_{i=1}^{N_t} \lambda_i = N_t} \tilde{C}(\boldsymbol{\lambda}) \qquad (15)$$

where

$$\tilde{C}(\boldsymbol{\lambda}) = \mathbb{E}\left\{ \log \det \left(\mathbf{I}_{N_r} + \gamma \tilde{\mathbf{H}} \operatorname{diag}(\boldsymbol{\lambda}) \tilde{\mathbf{H}}^H \right) \right\} \qquad (16)$$

is the expected mutual information between the transmitted signal $\mathbf{x}$ and the received signal $\mathbf{y}$ under SET. Due to the

concavity of the $\log(\cdot)$ function, the mutual information $\tilde{C}(\boldsymbol{\lambda})$ is upper bounded by

$$\tilde{C}(\boldsymbol{\lambda}) \leq \tilde{C}_u(\boldsymbol{\lambda}) = \log \mathbb{E}\left\{ \det \left(\mathbf{I}_{N_r} + \gamma \tilde{\mathbf{H}} \operatorname{diag}(\boldsymbol{\lambda}) \tilde{\mathbf{H}}^H \right) \right\}. \qquad (17)$$

Thus, the ergodic capacity is upper bounded by

$$C \leq C_u = \max_{\boldsymbol{\lambda} \geq \mathbf{0}, \sum_{i=1}^{N_t} \lambda_i = N_t} \tilde{C}_u(\boldsymbol{\lambda}). \qquad (18)$$

For the case of Kronecker MIMO channels, it has been shown in [10], [13], [14], [17], [18] that such bounds are very tight and admit closed-form expressions by using the expansion of the determinant.

III. CLOSED-FORM CAPACITY UPPER BOUND USING PERMANENTS

In this section, we derive a closed-form expression for the capacity upper bound (18) for the jointly correlated MIMO channel model in (4). We also develop algorithms for its efficient computation. The upper bound derivation is based heavily on exploiting linear-algebraic concepts and properties of *matrix permanents*, which we introduce and develop in the sequel.

A. Matrix Permanents: Definitions and Properties

The permanent of a matrix is defined in a similar fashion to the determinant. The primary difference is that when taking the expansion over minors, all signs are positive [35]–[38]. The permanents of square matrices have been thoroughly investigated in linear algebra and various applied fields. The permanents of $M \times N$ matrices with $M \leq N$ have also been defined and investigated[35]. In this paper, to facilitate our capacity upper bound derivation we find it necessary to extend the definition of permanents to allow for arbitrary M and N, and provide their useful properties.

Definition 1: For an $M \times N$ matrix $\mathbf{A}$, the permanent is defined as

$$\operatorname{Per}(\mathbf{A}) = \begin{cases} \displaystyle\sum_{\hat{\alpha}_M \in \mathcal{S}_N^M} \prod_{i=1}^{M} a_{i,\alpha_i}, & M \leq N \\ \displaystyle\sum_{\hat{\beta}_N \in \mathcal{S}_M^N} \prod_{i=1}^{N} a_{\beta_i,i}, & M > N \end{cases} \qquad (19)$$

where $a_{i,j}$ denotes the (i,j)th element of $\mathbf{A}$.

From this definition, one can easily establish a number of important properties of the matrix permanent, as given in the following lemma. These properties will be useful in subsequent derivations.

Lemma 1: Let $\mathbf{A}$ be an $M \times N$ matrix, $\mathbf{a}$ an $M \times 1$ vector, $\mathbf{b}$ an $N \times 1$ vector, and μ a scale constant. Then

$$\operatorname{Per}(\mathbf{A}) = \operatorname{Per}(\mathbf{A}^T) \qquad (20)$$

$$\operatorname{Per}(\mathbf{a}) = \sum_{i=1}^{M} a_i \qquad (21)$$

$$\operatorname{Per}(\operatorname{diag}(\mathbf{a})) = \det(\operatorname{diag}(\mathbf{a})) \qquad (22)$$

$$\operatorname{Per}(\mu\mathbf{A}) = \mu^{\min(M,N)}\operatorname{Per}(\mathbf{A}) \qquad (23)$$

$$\text{Per}(\text{diag}(\mathbf{a})\mathbf{A}) = \det(\text{diag}(\mathbf{a}))\text{Per}(\mathbf{A}), \ M \leq N \quad (24)$$

$$\text{Per}(\mathbf{A}\text{diag}(\mathbf{b})) = \det(\text{diag}(\mathbf{b}))\text{Per}(\mathbf{A}), \ M \geq N. \quad (25)$$

For an $M \times N$ matrix with $M \leq N$, there exists an analogy between the matrix permanent and the Laplace expansion of the determinant [35], [39]. The following lemma gives the straight-forward extension of this result for arbitrary[1] M and N.

Lemma 2: Let $\mathbf{A}$ be an $M \times N$ matrix. Then

$$\text{Per}(\mathbf{A}) = \begin{cases} \displaystyle\sum_{\hat{\alpha}_k \in \mathcal{S}_N^{(k)}} \text{Per}\left(\mathbf{A}_{\hat{\sigma}_k}^{\hat{\alpha}_k}\right)\text{Per}\left(\mathbf{A}_{\hat{\sigma}_k'}^{\hat{\alpha}_k'}\right), & M \leq N \\ \displaystyle\sum_{\hat{\sigma}_k \in \mathcal{S}_M^{(k)}} \text{Per}\left(\mathbf{A}_{\hat{\beta}_k}^{\hat{\sigma}_k}\right)\text{Per}\left(\mathbf{A}_{\hat{\beta}_k'}^{\hat{\sigma}_k'}\right), & M > N \end{cases} \quad (26)$$

where $\hat{\alpha}_k \in \mathcal{S}_M^{(k)}$ and $\hat{\beta}_k \in \mathcal{S}_N^{(k)}$ with $1 \leq k \leq \min(M,N)$. Note that for the case $k = \min(M,N)$, $\text{Per}(\mathbf{A}_{\hat{\sigma}_k'}^{\hat{\alpha}_k'}) = 1$ and $\text{Per}(\mathbf{A}_{\hat{\beta}_k'}^{\hat{\sigma}_k'}) = 1$.

For the special case $k = 1$, (26) can be reexpressed as follows:

$$\text{Per}(\mathbf{A}) = \begin{cases} \displaystyle\sum_{j=1}^{N} a_{i,j}\text{Per}\left(\mathbf{A}_{i,j}\right), & M \leq N \\ \displaystyle\sum_{j=1}^{M} a_{j,i}\text{Per}\left(\mathbf{A}_{j,i}\right), & M > N \end{cases} \quad (27)$$

where $1 \leq i \leq \min(M,N)$. This is analogous to the cofactor expansion of the determinant [39]. With $k = \min(M,N)$, (26) simplifies to

$$\text{Per}(\mathbf{A}) = \begin{cases} \displaystyle\sum_{\hat{\sigma}_M \in \mathcal{S}_N^{(M)}} \text{Per}\left(\mathbf{A}_{\hat{\sigma}_M}\right), & M \leq N \\ \displaystyle\sum_{\hat{\sigma}_N \in \mathcal{S}_M^{(N)}} \text{Per}\left(\mathbf{A}^{\hat{\sigma}_N}\right), & M > N. \end{cases} \quad (28)$$

The following two key lemmas are particularly important for deriving the closed-form capacity upper bound in the sequel.

Lemma 3: Let $\mathbf{A}$ be an $M \times N$ matrix. Then

$$\begin{aligned} \text{Per}([\mathbf{I}_M \ \mathbf{A}]) &= \text{Per}([\mathbf{I}_N \ \mathbf{A}^T] \\ &= \sum_{k=0}^{\min(N,M)} \sum_{\hat{\alpha}_k \in \mathcal{S}_M^{(k)}} \text{Per}(\mathbf{A}^{\hat{\alpha}_k}) \\ &= \sum_{k=0}^{\min(N,M)} \sum_{\hat{\beta}_k \in \mathcal{S}_N^{(k)}} \text{Per}(\mathbf{A}_{\hat{\beta}_k}) \end{aligned} \quad (29)$$

where $\text{Per}(\mathbf{A}^{\hat{\alpha}_k}) = 1$ and $\text{Per}(\mathbf{A}_{\hat{\beta}_k}) = 1$ when $k = 0$.

A proof is provided in Appendix A. The values of $\text{Per}([\mathbf{I}_M \ \mathbf{A}])$ and $\text{Per}([\mathbf{I}_N \ \mathbf{A}^T])$ in Lemma 3 will be called *extended permanents* of $\mathbf{A}$, which we denote as

$$\underline{\text{Per}}(\mathbf{A}) = \text{Per}([\mathbf{I}_M \ \mathbf{A}]) = \text{Per}([\mathbf{I}_N \ \mathbf{A}^T]) . \quad (30)$$

Lemma 4: For an $N \times N$ random matrix $\mathbf{X}$ with independent elements, suppose that there exists at most one non-zero element in each row of $\bar{\mathbf{X}} = \mathbb{E}\{\mathbf{X}\}$. Then we have

$$\mathbb{E}\left\{\det\left(\mathbf{X}\right)\det\left(\mathbf{X}^H\right)\right\} = \text{Per}\left(\boldsymbol{\Xi}\right), \quad (31)$$

where $\boldsymbol{\Xi} = \mathbb{E}\{\mathbf{X} \odot \mathbf{X}^*\}$.

A proof is provided in Appendix B. For the special case where all elements of $\mathbf{X}$ are independent and identically distributed with zero mean and unit variance, we have that $\boldsymbol{\Xi} = \mathbb{E}\{\mathbf{X} \odot \mathbf{X}^*\} = \mathbf{1}_{N \times N}$ and $\mathbb{E}\left\{\det\left(\mathbf{X}\right)\det\left(\mathbf{X}^H\right)\right\} = \text{Per}\left(\mathbf{1}_{N \times N}\right) = N!$. This agrees with prior results in [10], [17], [18], [40].

The following conjecture is useful when dealing with the optimal power allocation problem in Section IV.

Conjecture 1: Let $\mathbf{A}$ be an $M \times N$ matrix of nonnegative elements. Then $f(\boldsymbol{\lambda}) = \log \text{Per}(\mathbf{A}\text{diag}(\boldsymbol{\lambda}))$ is concave on $\mathcal{D}^N = \{\boldsymbol{\lambda} \,|\, \text{Per}(\mathbf{A}\text{diag}(\boldsymbol{\lambda})) > 0, \text{ and } \lambda_i \geq 0, \ 1 \leq i \leq N\}$.

For the general case with arbitrary M and N, the formal proof of this result is not available at this stage. In Appendix C, we provide proofs for several special cases, which lend support to the validity of this conjecture.

B. Capacity Upper Bound

Armed with the general results of the preceding subsection, we can now derive a closed-form expression for the upper bound on the ergodic capacity.

Theorem 2: The ergodic capacity in (14) is upper bounded by

$$C \leq C_u = \max_{\boldsymbol{\lambda} \geq \mathbf{0}, \mathbf{1}^T\boldsymbol{\lambda} = N_t} \log \tilde{C}_u(\boldsymbol{\lambda}) \quad (32)$$

where

$$\tilde{C}_u(\boldsymbol{\lambda}) = \log \underline{\text{Per}}\left(\gamma \boldsymbol{\Omega} \text{diag}(\boldsymbol{\lambda})\right). \quad (33)$$

Proof: We start by writing the upper bound for the expected mutual information under SET in (17) as

$$\tilde{C}_u(\boldsymbol{\lambda}) = \log E(\boldsymbol{\lambda}) \quad (34)$$

where

$$E(\boldsymbol{\lambda}) = \mathbb{E}\left\{\det\left(\mathbf{I}_{N_r} + \gamma \tilde{\mathbf{H}} \text{diag}(\boldsymbol{\lambda})\tilde{\mathbf{H}}^H\right)\right\}. \quad (35)$$

By using the characteristic polynomial expansion of the determinant, as well as the Cauchy-Binet formula for the determinant of a product matrix [39], $E(\boldsymbol{\lambda})$ can be expressed in (36), shown at the top of the following page. Let us denote $\mathbf{X} = (\tilde{\mathbf{H}})_{\hat{\beta}_k}^{\hat{\alpha}_k}$. Then, $\mathbf{X}^H = (\tilde{\mathbf{H}}^H)_{\hat{\alpha}_k}^{\hat{\beta}_k}$, and it is easily found that $\mathbb{E}\{\mathbf{X} \odot \mathbf{X}^*\} = \boldsymbol{\Omega}_{\hat{\beta}_k}^{\hat{\alpha}_k}$. The matrix $\mathbf{X}$ satisfies the conditions in Lemma 4. Thus, we have

$$\mathbb{E}\left\{\det\left(\tilde{\mathbf{H}}_{\hat{\beta}_k}^{\hat{\alpha}_k}\right)\det\left(\left(\tilde{\mathbf{H}}^H\right)_{\hat{\alpha}_k}^{\hat{\beta}_k}\right)\right\} = \text{Per}\left(\boldsymbol{\Omega}_{\hat{\beta}_k}^{\hat{\alpha}_k}\right). \quad (37)$$

$$
\begin{aligned}
E(\boldsymbol{\lambda}) &= \mathbb{E}\left\{ \sum_{k=0}^{\min(N_t,N_r)} \gamma^k \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_r}^{(k)}} \det\left(\left(\tilde{\mathbf{H}} \operatorname{diag}(\boldsymbol{\lambda}) \tilde{\mathbf{H}}^H \right)_{\hat{\alpha}_k}^{\hat{\alpha}_k} \right) \right\} \\
&= \mathbb{E}\left\{ \sum_{k=0}^{\min(N_t,N_r)} \gamma^k \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_r}^{(k)}} \sum_{\hat{\beta}_k \in \mathcal{S}_{N_t}^{(k)}} \sum_{\hat{\xi}_k \in \mathcal{S}_{N_t}^{(k)}} \det\left(\tilde{\mathbf{H}}_{\hat{\beta}_k}^{\hat{\alpha}_k} \right) \det\left(\operatorname{diag}(\boldsymbol{\lambda})_{\hat{\xi}_k}^{\hat{\beta}_k} \right) \det\left(\left(\tilde{\mathbf{H}}^H \right)_{\hat{\alpha}_k}^{\hat{\xi}_k} \right) \right\} \\
&= \sum_{k=0}^{\min(N_t,N_r)} \gamma^k \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_r}^{(k)}} \sum_{\hat{\beta}_k \in \mathcal{S}_{N_t}^{(k)}} \det\left(\operatorname{diag}(\boldsymbol{\lambda})_{\hat{\beta}_k}^{\hat{\beta}_k} \right) \mathbb{E}\left\{ \det\left(\tilde{\mathbf{H}}_{\hat{\beta}_k}^{\hat{\alpha}_k} \right) \det\left(\left(\tilde{\mathbf{H}}^H \right)_{\hat{\alpha}_k}^{\hat{\beta}_k} \right) \right\}.
\end{aligned}
\tag{36}
$$

Substituting (37) into (36) and using the properties of the permanents in Lemma 1, as well as (28) and Lemma 3, we find that

$$
\begin{aligned}
E(\boldsymbol{\lambda}) &= \sum_{k=0}^{\min(N_t,N_r)} \gamma^k \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_r}^{(k)}} \sum_{\hat{\beta}_k \in \mathcal{S}_{N_t}^{(k)}} \det\left(\operatorname{diag}(\boldsymbol{\lambda})_{\hat{\beta}_k}^{\hat{\beta}_k} \right) \operatorname{Per}\left(\boldsymbol{\Omega}_{\hat{\beta}_k}^{\hat{\alpha}_k} \right) \\
&= \sum_{k=0}^{\min(N_t,N_r)} \gamma^k \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_r}^{(k)}} \sum_{\hat{\beta}_k \in \mathcal{S}_{N_t}^{(k)}} \operatorname{Per}\left((\boldsymbol{\Omega} \operatorname{diag}(\boldsymbol{\lambda}))_{\hat{\beta}_k}^{\hat{\alpha}_k} \right) \\
&= \sum_{k=0}^{\min(N_t,N_r)} \gamma^k \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_r}^{(k)}} \operatorname{Per}\left((\boldsymbol{\Omega} \operatorname{diag}(\boldsymbol{\lambda}))^{\hat{\alpha}_k} \right) \\
&= \sum_{k=0}^{\min(N_t,N_r)} \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_r}^{(k)}} \operatorname{Per}\left((\gamma \boldsymbol{\Omega} \operatorname{diag}(\boldsymbol{\lambda}))^{\hat{\alpha}_k} \right) \\
&= \underline{\operatorname{Per}\left(\gamma \boldsymbol{\Omega} \operatorname{diag}(\boldsymbol{\lambda}) \right)}.
\end{aligned}
\tag{38}
$$

Substituting (38) into (34) and using (18) complete the proof. $\square$

From the above theorem, we see that the upper bound on capacity is completely determined by the average SNR $\rho\ (= \gamma N_t)$ and the eigenmode channel coupling matrix $\boldsymbol{\Omega}$. This bound is particularly useful, since we may now apply (33) to maximize $\tilde{C}_u(\boldsymbol{\lambda})$ with the respect to $\boldsymbol{\lambda}$ (i.e., address the power allocation problem), without the need for performing Monte-Carlo averaging over random realizations of the MIMO channel matrix.

It is interesting to consider the special case when $\mathbf{D} = 0$ and $\mathbf{M} = \mathbf{a}\mathbf{b}^T$, where $\mathbf{a}$ and $\mathbf{b}$ are $N_r \times 1$ and $N_t \times 1$ real vectors. In this case, the jointly correlated channel model considered in this paper reduces to the popular Kronecker correlation

model. Defining $\boldsymbol{\lambda}_r = \mathbf{a} \odot \mathbf{a}$ and $\boldsymbol{\lambda}_t = \mathbf{b} \odot \mathbf{b}$, the eigenmode channel coupling matrix can then be expressed as $\boldsymbol{\Omega} = \boldsymbol{\lambda}_r \boldsymbol{\lambda}_t^T = \operatorname{diag}(\boldsymbol{\lambda}_r)\ \mathbf{1}_{N_r \times N_t}\ \operatorname{diag}(\boldsymbol{\lambda}_t)$, and (33) reduces to (39), shown at the bottom of the page where (39) is equivalent to the upper bounds presented previously for Kronecker-correlated channels in [17], [18]. Moreover, for the special case $\boldsymbol{\lambda} = \mathbf{1}$ (i.e., the case of equal-power allocation), (39) reduces further to the capacity upper bound presented in [10].

C. Efficient Computation Algorithms

To evaluate the closed-form capacity upper bound expression given by (32) and (33), we must evaluate the extended permanent of the matrix $\gamma \boldsymbol{\Omega} \operatorname{diag}(\boldsymbol{\lambda})$. Clearly, when the size of the matrix is small, this can be done by simply expressing the extended permanent as a conventional permanent via (30), and then either directly applying Definition 1, or using the Laplace expansion in Lemma 2. However, in both cases, as the size of the matrix grows, the computational complexity increases significantly, and more efficient methods are needed. To see this, consider the task of evaluating the permanent of a general $M \times N$ matrix $\mathbf{A}$. The complexity associated with computing matrix permanents is conventionally measured in terms of the number of the required multiplications. Adopting this measure, the number of multiplications required for evaluating the matrix permanent using Definition 1 and the Laplace expansion (e.g., via (27)) are $\frac{(m-1)n!}{(n-m)!}$ and $\sum_{k=1}^{m-1} \frac{n!}{(n-k)!}$, respectively, where $m = \min(M, N)$ and $n = \max(M, N)$. Clearly, as the matrix dimensions increase, the computational complexity increases exponentially. For this reason, it is necessary to investigate more efficient computational algorithms.

$$
\begin{aligned}
\tilde{C}_u(\boldsymbol{\lambda}) &= \log \underline{\operatorname{Per}}\left(\gamma \boldsymbol{\Omega} \operatorname{diag}(\boldsymbol{\lambda}) \right) \\
&= \log \sum_{k=0}^{\min(N_t,N_r)} \gamma^k \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_r}^{(k)}} \sum_{\hat{\beta}_k \in \mathcal{S}_{N_t}^{(k)}} \det\left(\operatorname{diag}(\boldsymbol{\lambda})_{\hat{\beta}_k}^{\hat{\beta}_k} \right) \operatorname{Per}\left((\operatorname{diag}(\boldsymbol{\lambda}_r)\ \mathbf{1}_{N_r \times N_t}\ \operatorname{diag}(\boldsymbol{\lambda}_t))_{\hat{\beta}_k}^{\hat{\alpha}_k} \right) \\
&= \log \sum_{k=0}^{\min(N_t,N_r)} \gamma^k\, k! \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_r}^{(k)}} \sum_{\hat{\beta}_k \in \mathcal{S}_{N_t}^{(k)}} \det\left(\operatorname{diag}(\boldsymbol{\lambda}_r)_{\hat{\alpha}_k}^{\hat{\alpha}_k} \right) \det\left(\operatorname{diag}(\boldsymbol{\lambda} \odot \boldsymbol{\lambda}_t)_{\hat{\beta}_k}^{\hat{\beta}_k} \right).
\end{aligned}
\tag{39}
$$

The best-known algorithm for computing the matrix permanent of arbitrary dimensions is due to Ryser [36],[2] who showed that the permanent of the $M \times N$ matrix $\mathbf{A}$ (with $M \leq N$) can be evaluated via

$$\text{Per}(\mathbf{A}) = \sum_{k=0}^{M} (-1)^{M-k} C_{N-k}^{M-k} \sum_{\hat{\alpha}_k \in \mathcal{S}_N^{(k)}} \prod_{i=1}^{M} r_i(\mathbf{A}_{\alpha_k}) \quad (40)$$

where $C_j^i = \frac{j!}{i!(j-i)!}$, and $r_i(\cdot)$ represents the sum of the elements in the ith row of the matrix argument. A similar formula also exists for the case $M > N$. This algorithm requires $m + (m-1) \sum_{k=1}^{m} C_n^k$ multiplications, with m and n defined as above.

In our case, we are interested in computing the extended permanent $\underline{\text{Per}}(\hat{\mathbf{\Omega}})$ in (33), i.e., the permanent of $[\mathbf{I}_{N_r} \ \hat{\mathbf{\Omega}}]$ or $[\mathbf{I}_{N_t} \ \hat{\mathbf{\Omega}}^T]$, where $\hat{\mathbf{\Omega}} = \gamma \mathbf{\Omega} \text{diag}(\boldsymbol{\lambda})$. By directly computing this quantity based on Definition 1, the Laplace expansion, or Ryser's method, the number of required multiplications is $\frac{(N_{\min}-1)(N_{\min}+N_{\max})!}{N_{\max}!}$, $\sum_{k=1}^{N_{\min}-1} \frac{(N_{\min}+N_{\max})!}{(N_{\min}+N_{\max}-k)!}$ and $N_{\min} + (N_{\min}-1) \sum_{k=1}^{N_{\min}} C_{N_{\min}+N_{\max}}^k$, respectively, where $N_{\min} = \min(N_t, N_r)$ and $N_{\max} = \max(N_t, N_r)$. For practical values of N_r and N_t, these complexities can be quite high. As such, we are motivated to establish new and more efficient methods for computing the extended permanent, which we now consider.

Let us define the following auxiliary matrix :

$$\hat{\mathbf{\Omega}}(z) = \mathbf{1}_{N_r \times N_t} + z \hat{\mathbf{\Omega}}. \quad (41)$$

We will establish new efficient computation algorithms for $\underline{\text{Per}}(\hat{\mathbf{\Omega}})$ based on the following key result.

Lemma 5: Let $\text{Per}(\hat{\mathbf{\Omega}}(z)) = \sum_{k=0}^{\min(N_r, N_t)} \mu_k z^k$. Then

$$\underline{\text{Per}}(\hat{\mathbf{\Omega}}) = \sum_{k=0}^{\min(N_r, N_t)} \mu_k c_k \quad (42)$$

where $c_k = |N_t - N_r|!/(\max(N_r, N_t) - k)!$.

A proof is presented in Appendix D. This result shows that the extended permanent $\underline{\text{Per}}(\hat{\mathbf{\Omega}})$ can be calculated directly from the polynomial expansion of $\text{Per}(\hat{\mathbf{\Omega}}(z))$. Considering the case $N_r \leq N_t$, from Definition 1 in (19), Laplace expansion (27) and Ryser's expression (40), we have the following three formulas for $\text{Per}(\hat{\mathbf{\Omega}}(z))$:

$$\text{Per}(\hat{\mathbf{\Omega}}(z)) = \sum_{\hat{\alpha}_{N_r} \in \mathcal{S}_{N_t}^{N_r}} \prod_{i=1}^{N_r} (1 + \hat{\omega}_{i,\alpha_i} z) \quad (43)$$

$$\text{Per}(\hat{\mathbf{\Omega}}(z)) = \sum_{j=1}^{N_t} (1 + \hat{\omega}_{i,j} z) \text{Per}\left(\hat{\mathbf{\Omega}}(z)_{i,j}\right) \quad (44)$$

$$\text{Per}(\hat{\mathbf{\Omega}}(z)) = \sum_{k=0}^{N_r} (-1)^{N_r-k} C_{N_t-k}^{N_r-k} \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_t}^{(k)}} \prod_{i=1}^{N_r} r_i(\hat{\mathbf{\Omega}}(z)_{\alpha_k}). \quad (45)$$

It is convenient to reexpress (45) by letting

$$\prod_{i=1}^{N_r} (1 + r_i(\hat{\mathbf{\Omega}}_{\alpha_k}) z) = \sum_{i=0}^{N_r} a_{i,\alpha_k} z^i \quad (46)$$

with $a_{0,\alpha_k} = 1$, such that

$$\prod_{i=1}^{N_r} r_i(\hat{\mathbf{\Omega}}(z)_{\alpha_k}) = \sum_{i=0}^{N_r} k^{N_r-i} a_{i,\alpha_k} z^i. \quad (47)$$

This yields

$$\text{Per}(\hat{\mathbf{\Omega}}(z)) = \sum_{k=0}^{N_r} \sum_{i=0}^{N_r} z^i (-1)^{N_r-k} C_{N_r-k}^{N_t-k} k^{N_r-i} \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_t}^{(k)}} a_{i,\alpha_k}. \quad (48)$$

Importantly, we find that each of the equivalent expressions (43), (44) and (48) admit simple and efficient recursive algorithms for calculating the coefficients of z. To demonstrate this, consider (43). Let $\tilde{b}_k(z) = \prod_{i=1}^{k} (1 + \hat{\omega}_{i,\alpha_i} z) = 1 + \sum_{n=1}^{k} b_{k,n} z^n$, where $k = 1, 2, \ldots N_r$. Then, $\tilde{b}_{k+1}(z) = \tilde{b}_k(z) \left(1 + \hat{\omega}_{k+1,\alpha_{k+1}} z\right)$ for $1 \leq k \leq N_r - 1$, and therefore the coefficients of z can be evaluated recursively via

$$b_{k+1,n} = \begin{cases} \hat{\omega}_{k+1,\alpha_{k+1}} + b_{k,1}, & n = 1 \\ \hat{\omega}_{k+1,\alpha_{k+1}} b_{k,n-1} + b_{k,n}, & 2 \leq n \leq k \\ \hat{\omega}_{k+1,\alpha_{k+1}} b_{k,k}, & n = k+1. \end{cases} \quad (49)$$

This result, combined with Lemma 5, presents an efficient algorithm for computing the extended permanent $\underline{\text{Per}}(\hat{\mathbf{\Omega}})$. In a similar manner, efficient computational algorithms can also be easily obtained based on (44) and (48). We omit the specific details of these. For arbitrary N_t and N_r, with $N_{\min}$ and $N_{\max}$ defined as above, the number of required multiplications for the three polynomial-based computation algorithms are $\frac{N_{\min}(N_{\min}-1)N_{\max}!}{2(N_{\max}-N_{\min})!}$, $\sum_{k=1}^{N_{\min}-1} \frac{(N_{\min}-k)N_{\max}!}{(N_{\max}-k)!}$ and $N_{\min}^2 + \frac{N_{\min}(N_{\min}-1)}{2} \sum_{k=1}^{N_{\min}} C_{N_{\max}}^k$, respectively.

Fig. 1 presents the number of required multiplications for evaluating $\tilde{C}_u(\boldsymbol{\lambda})$ based on the three polynomial-based computation algorithms, for various antenna configurations of the form $N = N_t = N_r$. The number of required multiplications for calculating $\tilde{C}_u(\boldsymbol{\lambda})$ by *directly* using Definition 1, Laplace expansion and Ryser's formula are also shown for comparison. We clearly see that the polynomial-based algorithms have significantly reduced computational complexity compared with the direct methods; in many cases yielding orders of magnitude improvements. Of the polynomial-based algorithms, the Laplace expansion gives the least complexity for $N \leq 5$, whereas the Ryser-based formula is most efficient for $N > 5$.

IV. OPTIMAL POWER ALLOCATION WITH THE CAPACITY BOUND

A. Asymptotic Optimality at Low and High SNR

Based on the tight closed-form capacity upper bound in Theorem 2, we can now address the transmitter power allocation optimization problem by dealing with only the eigenmode channel coupling matrix $\mathbf{\Omega}$ and the transmit SNR $\rho (= \gamma N_t)$. The optimal solution for maximizing the upper bound will then serve

---

[2]For the case of square matrices, further improvements have also been proposed [37].
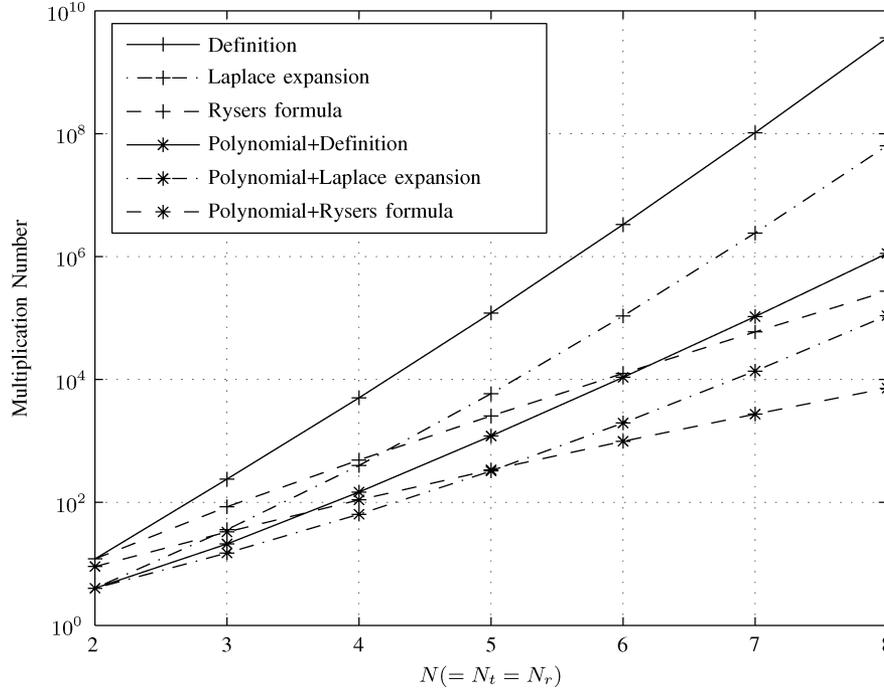
Fig. 1.   Comparison of the number of required multiplications for calculating $\bar{C}_u(\boldsymbol{\lambda})$ with the polynomial-based algorithms and the direct computation algorithms.

as an approximation to the optimal capacity-achieving power allocation solution. Our numerical results will confirm the accuracy of this approximation.

The power allocation optimization problem can be formulated as follows:

$$\max_{\boldsymbol{\lambda}} \quad \tilde{C}_u(\boldsymbol{\lambda}) \tag{50}$$

$$\text{subject to} \quad \boldsymbol{\lambda} \geq \mathbf{0}, \ \mathbf{1}^T\boldsymbol{\lambda} = N_t. \tag{51}$$

Before dealing with this problem in its most generality, we briefly check the asymptotic optimality of our approach at low and high SNR. For arbitrary SNRs, we will then develop optimality conditions and an iterative numerical algorithm in the framework of convex optimization.

For low SNRs, $\tilde{C}_u(\boldsymbol{\lambda})$ can be expressed as

$$\tilde{C}_u(\boldsymbol{\lambda}) = \log\left(1 + \gamma\sum_{i=1}^{N_t}\tau_i\lambda_i + O(\gamma^2)\right)$$

$$= \gamma\sum_{i=1}^{N_t}\tau_i\lambda_i + O(\gamma^2) \tag{52}$$

where $\tau_i = \sum_{j=1}^{N_r}[\boldsymbol{\Omega}]_{ji}$. Without any loss of generality, assume that $\tau_1 = \tau_2 = \cdots = \tau_l > \tau_{l+1} \geq \cdots \geq \tau_{N_t}$. Maximizing the first-order (in γ) term in (52) subject to the constraint (51) gives the following power-allocation policy

$$\lambda_i = \begin{cases} N_t/l, & \text{for } i = 1, \ldots, l \\ 0, & \text{for } i = l+1, \ldots, N_t. \end{cases} \tag{53}$$

This means that beamforming along the strongest transmit eigenmodes (specified by the channel coupling matrix $\boldsymbol{\Omega}$) is optimal in the low SNR case.

For high SNRs, with $N_t \leq N_r$, we have

$$\tilde{C}_u(\boldsymbol{\lambda}) \rightarrow \log\left(\text{Per}(\gamma\Omega\,\text{diag}(\lambda))\right)$$

$$= \log\text{Per}(\gamma\Omega) + \log\det\left(\text{diag}(\lambda)\right) \tag{54}$$

which is maximized by the following power allocation policy

$$\lambda_i = 1, \quad i = 1, \ldots, N_t \tag{55}$$

i.e., equal-power allocation over the transmit eigenmodes. These low and high SNR power allocation policies, derived based on the capacity upper bound, coincide exactly with the optimal capacity-achieving power allocation policies for the low and high SNR regimes, considered previously in [24], [26].

B. Optimality Conditions for Arbitrary SNRs

We now address the general case with arbitrary SNRs. To this end, let $\boldsymbol{\lambda}_1 \geq \mathbf{0}$, $\boldsymbol{\lambda}_2 \geq \mathbf{0}$, and $0 \leq \theta \leq 1$. Then, using Conjecture 1, we can write

$$C_u(\theta\boldsymbol{\lambda}_1 + (1-\theta)\boldsymbol{\lambda}_2)$$
$$= \log\text{Per}\Big(\theta[\mathbf{I}_{N_r}\,\gamma\boldsymbol{\Omega}\,\text{diag}(\boldsymbol{\lambda}_1)] + (1-\theta)[\mathbf{I}_{N_r}\,\gamma\boldsymbol{\Omega}\,\text{diag}(\boldsymbol{\lambda}_2)]\Big)$$
$$= \log\text{Per}\Big(\theta[\mathbf{I}_{N_r}\,\gamma\boldsymbol{\Omega}]\text{diag}(\tilde{\boldsymbol{\lambda}}_1) + (1-\theta)[\mathbf{I}_{N_r}\,\gamma\boldsymbol{\Omega}]\text{diag}(\tilde{\boldsymbol{\lambda}}_2)\Big)$$
$$\geq \theta\log\text{Per}\Big([\mathbf{I}_{N_r}\,\gamma\boldsymbol{\Omega}]\text{diag}(\tilde{\boldsymbol{\lambda}}_1)\Big)$$
$$\quad + (1-\theta)\log\text{Per}\Big([\mathbf{I}_{N_r}\,\gamma\boldsymbol{\Omega}]\text{diag}(\tilde{\boldsymbol{\lambda}}_2)\Big)$$
$$= \theta\tilde{C}_u(\boldsymbol{\lambda}_1) + (1-\theta)\tilde{C}_u(\boldsymbol{\lambda}_2) \tag{56}$$

where $\tilde{\boldsymbol{\lambda}}_1 = [\mathbf{1}_{1 \times N_r}\ \boldsymbol{\lambda}_1^T]^T$ and $\tilde{\boldsymbol{\lambda}}_2 = [\mathbf{1}_{1 \times N_r}\ \boldsymbol{\lambda}_2^T]^T$. Therefore, the function $\tilde{C}_u(\boldsymbol{\lambda})$ is concave on the space of nonnegative $\boldsymbol{\lambda}$, and the optimization problem given by (50) and (51) is a concave optimization problem. As such, there exists only one local

optimal solution, which is also a global solution. This solution could be evaluated by employing standard convex optimization algorithms, such as interior point methods [41].

Since the problem is concave, we can derive necessary and sufficient conditions for the optimal solution using the Karush–Kuhn–Tucker (KKT) conditions. To this end, let $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_{N_t}]^T$ and ν be the Lagrange multipliers for the inequality constraint $\boldsymbol{\lambda} \geq 0$ and the equality constraint $\mathbf{1}^T \boldsymbol{\lambda} = N_t$ respectively. Then the KKT conditions satisfied by the optimal $\boldsymbol{\lambda}$ can be expressed as

$$\frac{\partial \tilde{C}_u(\boldsymbol{\lambda})}{\partial \lambda_i} + \mu_i + \nu = 0 \tag{57}$$

$$\boldsymbol{\lambda} \geq 0, \quad \mathbf{1}^T \boldsymbol{\lambda} = N_t, \quad \boldsymbol{\mu} \geq 0, \quad \mu_i \lambda_i = 0 \tag{58}$$

where $\frac{\partial \tilde{C}_u(\boldsymbol{\lambda})}{\partial \lambda_i}$ denotes the partial derivative of $\tilde{C}_u(\boldsymbol{\lambda})$ with respect to λ_i, for $1 \leq i \leq N_t$. From (33), these derivatives can be written as

$$\frac{\partial \tilde{C}_u(\boldsymbol{\lambda})}{\partial \lambda_i} = \frac{1}{E(\boldsymbol{\lambda})} \frac{\partial E(\boldsymbol{\lambda})}{\partial \lambda_i} \tag{59}$$

where $E(\boldsymbol{\lambda}) = \underline{\mathrm{Per}}(\gamma \boldsymbol{\Omega} \mathrm{diag}(\boldsymbol{\lambda}))$. To evaluate the remaining derivatives in (59) it is useful to apply the Laplace expansion property of permanents, given by (27), to express $E(\boldsymbol{\lambda})$ as follows:

$$E(\boldsymbol{\lambda}) = p(\boldsymbol{\lambda}_{(i)}) + \lambda_i q(\boldsymbol{\lambda}_{(i)}) \tag{60}$$

where

$$p(\boldsymbol{\lambda}_{(i)}) = \underline{\mathrm{Per}}\left(\gamma \boldsymbol{\Omega}_{(i)} \mathrm{diag}(\boldsymbol{\lambda}_{(i)})\right) \tag{61}$$

$$q(\boldsymbol{\lambda}_{(i)}) = \sum_{j=1}^{N_t} \gamma \omega_{j,i} \underline{\mathrm{Per}}\left(\gamma \boldsymbol{\Omega}_{(i)}^{(j)} \mathrm{diag}(\boldsymbol{\lambda}_{(i)})\right)$$
$$= \underline{\mathrm{Per}}(\gamma \boldsymbol{\Omega} \mathrm{diag}(\boldsymbol{\lambda}_i)) - \underline{\mathrm{Per}}\left(\gamma \boldsymbol{\Omega}_{(i)} \mathrm{diag}(\boldsymbol{\lambda}_{(i)})\right) \tag{62}$$

$\omega_{j,i}$ denotes the (j,i)th element of $\boldsymbol{\Omega}$, $\boldsymbol{\Omega}_{(i)}$ denotes the sub-matrix of $\boldsymbol{\Omega}$ obtained by deleting the ith column, $\boldsymbol{\Omega}_{(i)}^{(j)}$ denotes the sub-matrix of $\boldsymbol{\Omega}$ obtained by deleting the jth row and ith column, $\boldsymbol{\lambda}_{(i)}$ denotes the $(N_t - 1) \times 1$ vector obtained by deleting the ith element of $\boldsymbol{\lambda}$, and $\boldsymbol{\lambda}_i$ denotes the $N_t \times 1$ vector obtained by replacing the ith element of $\boldsymbol{\lambda}$ by unity. Therefore, (59) becomes

$$\frac{\partial \tilde{C}_u(\boldsymbol{\lambda})}{\partial \lambda_i} = \frac{q(\boldsymbol{\lambda}_{(i)})}{p(\boldsymbol{\lambda}_{(i)}) + \lambda_i q(\boldsymbol{\lambda}_{(i)})}. \tag{63}$$

Substituting (63) into (57) and eliminating the slack variable $\boldsymbol{\mu}$, the KKT conditions become

$$\lambda_i = \left(\tilde{\nu} - \frac{p(\boldsymbol{\lambda}_{(i)})}{q(\boldsymbol{\lambda}_{(i)})}\right)^+ \tag{64}$$

$$\mathbf{1}^T \boldsymbol{\lambda} = N_t \tag{65}$$

where $(a)^+ = \max\{0, a\}$ and $\tilde{\nu} = 1/\nu$.

In summary, we have the following theorem.

Theorem 3: The expected mutual information upper bound $\tilde{C}_u(\boldsymbol{\lambda})$ is concave with respect to $\boldsymbol{\lambda}$, and the necessary and sufficient conditions for optimal power allocation are given by (64), where $\tilde{\nu}$ is chosen to satisfy the power constraint in (65).

Note that when the eigenmode channel coupling matrix $\boldsymbol{\Omega}$ is square and diagonal,[3] we have

$$q(\boldsymbol{\lambda}_{(i)}) = \omega_{i,i} p(\boldsymbol{\lambda}_{(i)}), \tag{66}$$

and the conditions in (64) simplify to

$$\lambda_i = \left(\tilde{\nu} - \frac{1}{\omega_{i,i}}\right)^+. \tag{67}$$

This is the same formula as the water-filling solution when the transmitter has instantaneous CSI [1], and one can easily obtain the optimal power allocation via the water-filling algorithm. However, in the general case of an arbitrary eigenmode channel coupling matrix, the solution can not be obtained as easily and numerical approaches are required.

C. Iterative Water-Filling Algorithm

In this section, we propose a simple iterative water-filling algorithm (IWFA) for evaluating the optimal power allocation policy which satisfies (64). Our algorithm is based on observing that the right-hand side of (64) is independent of λ_i, and is motivated by the IWFA methods proposed in [32], [33] for transmitter optimization of multiuser systems with instantaneous CSI known to the transmitters. Simulation results, given in Section V, show that this approach works very well and is highly efficient; typically converging after only a few iterations, with the first iteration achieving near-optimal performance. The proposed algorithm includes the following steps:

1) Initialize $\boldsymbol{\lambda}^0 = \mathbf{1}$, $\tilde{C}_u(\boldsymbol{\lambda}^0) = \log \underline{\mathrm{Per}}(\gamma \boldsymbol{\Omega})$, and $k = 0$.
2) Calculate $p(\boldsymbol{\lambda}_{(i)}^k) = \underline{\mathrm{Per}}(\gamma \boldsymbol{\Omega}_{(i)} \mathrm{diag}(\boldsymbol{\lambda}_{(i)}^k))$ and $q(\boldsymbol{\lambda}_{(i)}^k) = \underline{\mathrm{Per}}\left(\gamma \boldsymbol{\Omega} \mathrm{diag}(\boldsymbol{\lambda}_i^k)\right) - \underline{\mathrm{Per}}(\gamma \boldsymbol{\Omega}_{(i)} \mathrm{diag}(\boldsymbol{\lambda}_{(i)}^k))$, $i = 1, 2, \ldots, N_t$.
3) Calculate $\lambda_i^{k+1} = (\tilde{\nu} - \frac{p(\boldsymbol{\lambda}_{(i)}^k)}{q(\boldsymbol{\lambda}_{(i)}^k)})^+$, $i = 1, 2, \ldots, N_t$, via the conventional water-filling algorithm with power constraint $\sum_{i=1}^{N_t} \lambda_i^{k+1} = N_t$.
4) Calculate $\tilde{C}_u(\boldsymbol{\lambda}^{k+1}) = \log \underline{\mathrm{Per}}(\gamma \boldsymbol{\Omega} \mathrm{diag}(\boldsymbol{\lambda}^{k+1}))$.
5) If $\tilde{C}_u(\boldsymbol{\lambda}^{k+1}) \leq \tilde{C}_u(\boldsymbol{\lambda}^k)$, set $\boldsymbol{\lambda}^{k+1} := \frac{1}{N_t}\boldsymbol{\lambda}^{k+1} + \frac{N_t-1}{N_t}\boldsymbol{\lambda}^k$, and recalculate $\tilde{C}_u(\boldsymbol{\lambda}^{k+1})$.
6) Set $k := k + 1$ and return to Step 2 until the algorithm converges or the iteration number is equal to a predefined value.

Here, $\boldsymbol{\lambda}^k$ stands for the value of $\boldsymbol{\lambda}$ in the kth iteration. In Step 1 in the first iteration, $\boldsymbol{\lambda}$ is initialized to $\mathbf{1}$, i.e., to the equal power allocation. Note, however, that $\boldsymbol{\lambda}$ could also be initialized in a different way. For example, in practice it is reasonable to suppose that the channel statistics change smoothly from frame to frame, where a more appropriate starting point for any given frame would be the optimal value of $\boldsymbol{\lambda}$ from the previous frame. This could speed up the convergence of the IWFA. In Step 3, the conventional water-filling algorithm is performed with the required variables $p(\boldsymbol{\lambda}_{(i)})$ and $q(\boldsymbol{\lambda}_{(i)})$ calculated in Step 2. Following the calculation of the $\tilde{C}_u(\boldsymbol{\lambda})$ in Step 4, Step 5 is performed to guarantee the convergence of the iterative procedure. We discuss this issue in detail below. In Step 6, the convergence of the algorithm can be determined by checking whether $|\tilde{C}_u(\boldsymbol{\lambda}^{k+1}) - \tilde{C}_u(\boldsymbol{\lambda}^k)|$ (or $\|\boldsymbol{\lambda}^{k+1} - \boldsymbol{\lambda}^k\|$) is less than some predefined tolerance.

---

[3]In this special case, the MIMO channel is essentially reduced to a set of non-interfering scalar subchannels.

Theorem 4: The IWFA for optimal power allocation converges to the capacity upper bound C_u.

Proof: In order to verify the convergence of our proposed IWFA for optimal power allocation, we define the following function for a given $\boldsymbol{\lambda}^k$:

$$\bar{C}_u(\boldsymbol{\lambda}) = \frac{1}{N_t} \sum_{i=1}^{N_t} \log \left(p(\boldsymbol{\lambda}_{(i)}^k) + \lambda_i q(\boldsymbol{\lambda}_{(i)}^k) \right). \qquad (68)$$

It can be seen that $\bar{C}_u(\boldsymbol{\lambda})$ is a concave function with respect to $\boldsymbol{\lambda}$. The water-filling solution in Step 3 of the IWFA is exactly equal to the solution of maximizing $\bar{C}_u(\boldsymbol{\lambda})$, for a given $\boldsymbol{\lambda}^k$ subject to the power constraint $\mathbf{1}^T \boldsymbol{\lambda} = N_t$. Therefore, with the $\boldsymbol{\lambda}^{k+1}$ resulting from Step 3 of the IWFA, we have

$$\bar{C}_u(\boldsymbol{\lambda}^{k+1}) \geq \bar{C}_u(\boldsymbol{\lambda}^k) = \tilde{C}_u(\boldsymbol{\lambda}^k). \qquad (69)$$

From the concavity of $\tilde{C}_u(\boldsymbol{\lambda})$, it can be shown that the following relation holds:

$$\bar{C}_u(\boldsymbol{\lambda}^{k+1}) \leq \tilde{C}_u \left(\frac{1}{N_t} \boldsymbol{\lambda}^{k+1} + \frac{N_t - 1}{N_t} \boldsymbol{\lambda}^k \right). \qquad (70)$$

Combining (69) and (70) yields

$$\tilde{C}_u(\boldsymbol{\lambda}^k) \leq \tilde{C}_u \left(\frac{1}{N_t} \boldsymbol{\lambda}^{k+1} + \frac{N_t - 1}{N_t} \boldsymbol{\lambda}^k \right). \qquad (71)$$

Therefore, after Step 5 of the IWFA, we have that $\tilde{C}_u(\boldsymbol{\lambda}^{k+1}) \geq \tilde{C}_u(\boldsymbol{\lambda}^k)$. This, along with the fact that the problem (50)–(51) is convex, completes the proof. $\qquad \square$

Notice that the relation (71) suggests, mathematically, to update $\boldsymbol{\lambda}$ with $(\frac{1}{N_t} \boldsymbol{\lambda}^{k+1} + \frac{N_t-1}{N_t} \boldsymbol{\lambda}^k)$ in the kth iteration of the IWFA, whereas the KKT conditions (64) suggest a more intuitive interpretation based on the water-filling principle. In our proposed IWFA, we update $\boldsymbol{\lambda}$ with the water-filling solution if the resulting $\tilde{C}_u(\boldsymbol{\lambda})$ is increased. This allows very fast convergence, as we demonstrate through simulations in the following section. To guarantee the convergence, we use $(\frac{1}{N_t} \boldsymbol{\lambda}^{k+1} + \frac{N_t-1}{N_t} \boldsymbol{\lambda}^k)$ to replace the water-filling solution when the resulting $\tilde{C}_u(\boldsymbol{\lambda})$ is not increasing in each iteration.

V. SIMULATIONS

In this section, we present numerical results to evaluate the tightness of the capacity bound, and to demonstrate the efficiency and performance of the proposed transmitter optimization approach under SET. We consider a MIMO system with five transmit and five receive antennas, and present results for both the jointly correlated MIMO channel model and the Kronecker-correlation model. For the jointly correlated channel, we adopt the same channel parameters as used in [24], where $\mathbf{D} = 0$ and $\boldsymbol{\Omega}$ has the following structure

$$\boldsymbol{\Omega} = \frac{25}{5.7} \begin{pmatrix} 0.1 & 0 & 1 & 0 & 0 \\ 0 & 0.1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.25 & 0 \\ 0 & 0 & 1 & 0 & 0.25 \end{pmatrix}. \qquad (72)$$

For the Kronecker channel, we adopt the constant-correlation model for constructing the transmit and receive correlation matrices [10]. An $N \times N$ constant-correlation matrix is given by

$$\boldsymbol{\Theta}_N(\alpha) = \alpha \mathbf{1}_{N \times N} + (1 - \alpha) \mathbf{I}_N \qquad (73)$$

where $\alpha \in [0,1]$ is the correlation coefficient. We set the transmit and receive correlation coefficients to be $\alpha_t = 0.4$ and $\alpha_r = 0.6$ respectively.

Fig. 2 compares our closed-form ergodic mutual information upper bound (18) with Monte-Carlo simulated exact curves based on (16), for the case $\boldsymbol{\lambda} = \mathbf{1}$ (equal-power allocation). Results are shown for both the jointly correlated channel and the Kronecker channel, with the above settings. We see that the upper bound is rather tight for both channel models, especially for low to moderate SNRs (e.g., < 8 dB). Moreover, we see that the bound for the Kronecker model is slightly tighter than for the more general jointly correlated model. Interestingly, we will show that, despite this difference in tightness, the low-complexity power allocation policies derived based on these bounds perform near-optimally for both the Kronecker and jointly correlated channel models.

Figs. 3 and 4 present the ergodic mutual information achieved by the SET approach employing the proposed IWFA (derived based on our closed-form upper bound), in the jointly correlated and Kronecker channel scenarios, respectively. For comparison, the exact ergodic capacity curves are also shown, which were obtained by numerically evaluating (14) using a constrained optimization function of the Matlab optimization toolbox. The ergodic mutual information achieved with equal power allocation (55) and beamforming (53) are also shown for further comparison. We clearly see that, for both channel models, the proposed SET approach performs near-optimally, suffering almost negligible loss compared with the true channel capacity. Furthermore, we see that equal power allocation and beamforming are optimal in the high and low SNR regimes, respectively, which agrees with our analytical conclusions put forth in Section IV-A. The capacity upper bound curve is also shown on the figures, and once again is seen to be tight.

Figs. 5 and 6 demonstrate the convergence of the proposed IWFA for optimal power allocation in the jointly correlated and Kronecker channel scenarios, respectively. Here, the SNR ρ was set to 10 dB, and in all cases the algorithm was initialized using $\boldsymbol{\lambda}^0 = \mathbf{1}$. These figures show the evolution of the eigenvalues λ_i, $i = 1, \ldots, 5$, and the capacity bound $\tilde{C}_u(\boldsymbol{\lambda})$ for each iteration. From these results, we see that the proposed IWFA converges after only a few iterations, with the first iteration achieving near-optimal performance in all cases.

VI. CONCLUSION

We have investigated statistical eigenmode transmission over a general jointly correlated MIMO channel. For this channel, we derived a tight closed-form upper bound for the ergodic capacity, which reveals a simple and interesting relationship in terms of matrix permanents of the eigenmode channel coupling matrix, and embraces many existing results in the literature as special cases. Based on this expression, we proposed
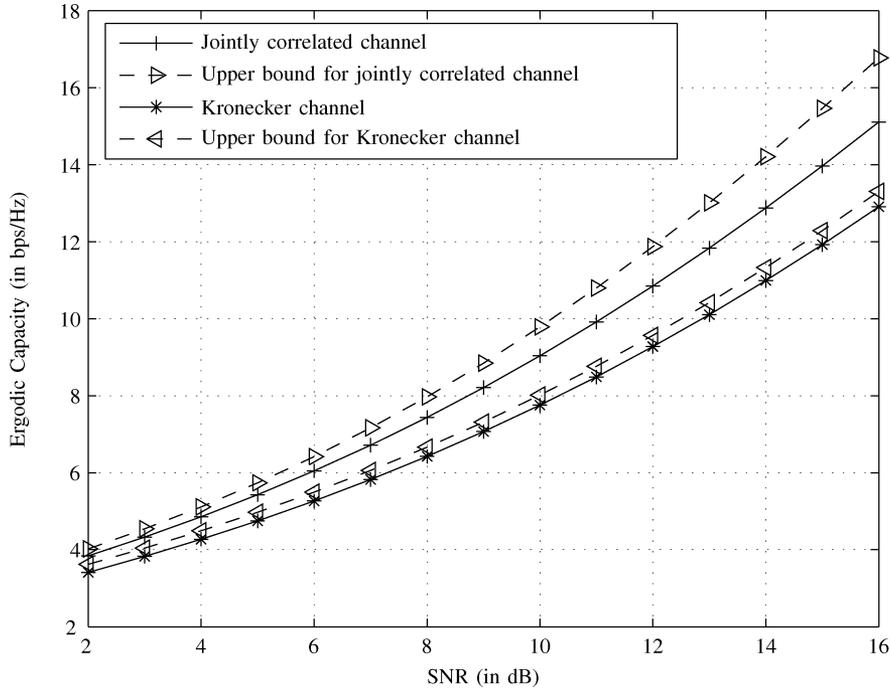
Fig. 2. Comparison of the exact ergodic mutual information and the mutual information upper bound. Results are shown for the jointly correlated channel model and the Kronecker model, with equal-power allocation (i.e., $\boldsymbol{\lambda} = \mathbf{1}$).
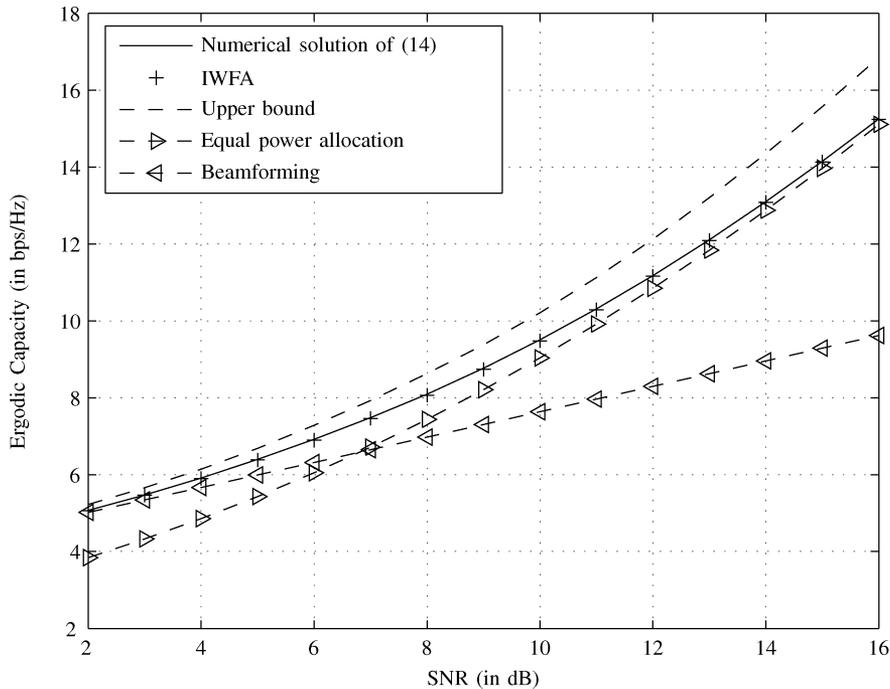
Fig. 3. Comparison of the ergodic capacity of the jointly correlated MIMO channel achieved by numerically solving (14), and our proposed iterative water-filling algorithm under SET. The capacity upper bound and the information rates achieved by equal power allocation and beamforming are also shown.

and investigated new power allocation policies in the framework of convex optimization. In particular, we obtained necessary and sufficient optimality conditions, and developed an efficient iterative water-filling algorithm with guaranteed convergence. The tightness of the capacity bound and the performance of our novel low-complexity transmitter optimization approach was confirmed through simulations. Our approach was shown to suffer near-negligible loss compared with the ergodic capacity of the jointly correlated MIMO channel.

APPENDIX I
PROOF OF LEMMA 3

Let $\underline{\mathbf{I}} = [\mathbf{I}_M \ \mathbf{0}_{M \times N}]$ and $\underline{\mathbf{A}} = [\mathbf{0}_{M \times M} \ \mathbf{A}]$. From the definition of the permanents, we have

$$\mathrm{Per}([\mathbf{I}_M \ \mathbf{A}]) = \mathrm{Per}(\underline{\mathbf{A}} + \underline{\mathbf{I}})$$

$$= \sum_{\hat{\beta}_M \in \mathcal{S}_{M+N}^M} \prod_{m=1}^{M} \left(\underline{a}_{m,\beta_m} + \underline{i}_{m,\beta_m} \right) \quad (74)$$
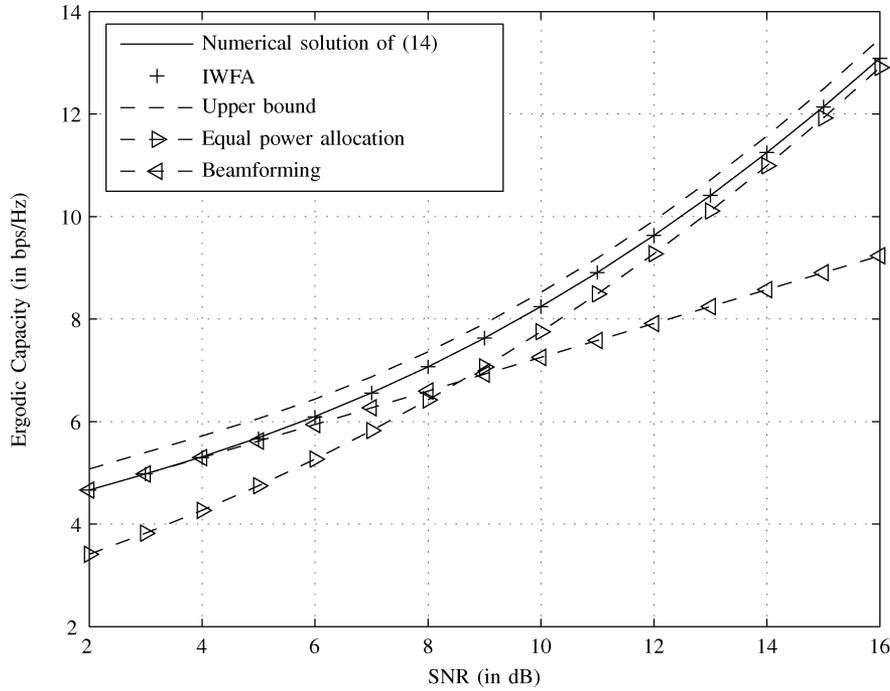
Fig. 4. Comparison of the ergodic capacity of the Kronecker MIMO channel achieved by numerically solving (14), and our proposed iterative water-filling algorithm under SET. The capacity upper bound and the information rates achieved by equal power allocation and beamforming are also shown.
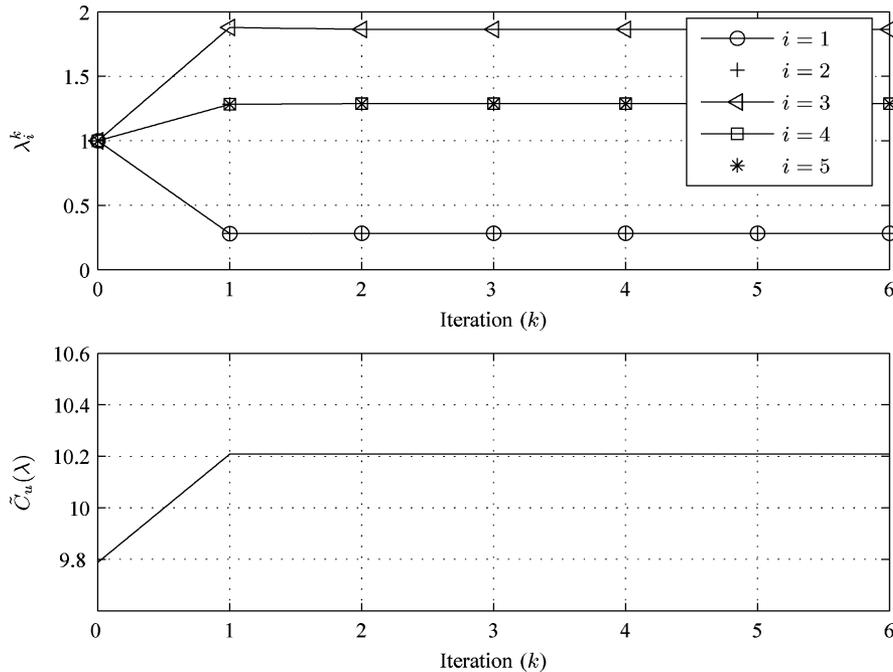
Fig. 5. Convergence of the iterative water-filling algorithm for optimal power allocation in the jointly correlated channel. Results are shown for SNR = 10 dB.

where $\underline{i}_{m,n}$ and $\underline{a}_{m,n}$ denote the (m,n)th elements of $\underline{\mathbf{I}}$ and $\underline{\mathbf{A}}$ respectively. Note that the following identity holds:

$$\prod_{m=1}^{M}(x_m+y_m)=\sum_{k=0}^{M}\sum_{\hat{\alpha}_k\in\mathcal{S}_M^{(k)}}\prod_{m=1}^{k}x_{\alpha_m}\prod_{m=1}^{M-k}y_{\alpha'_m} \quad (75)$$

where

$$(\alpha'_1,\alpha'_2,\ldots,\alpha'_{M-k})\in\mathcal{S}_M^{M-k}$$

is the sequence complementary to $\hat{\alpha}_k$ in $\{1,2,\ldots,M\}$. Hence

$$\mathrm{Per}([\mathbf{I}_M\,\mathbf{A}])$$
$$=\sum_{k=0}^{M}\sum_{\hat{\alpha}_k\in\mathcal{S}_M^{(k)}}\left(\sum_{\hat{\beta}_M\in\mathcal{S}_{M+N}^M}\prod_{m=1}^{k}\underline{a}_{\alpha_m,\beta_{\alpha_m}}\prod_{m=1}^{M-k}\underline{i}_{\alpha'_m,\beta_{\alpha'_m}}\right). \quad (76)$$

It can be seen that

$$\underline{i}_{\alpha'_m,\beta_{\alpha'_m}}=\delta(\beta_{\alpha'_m}-\alpha'_m)$$
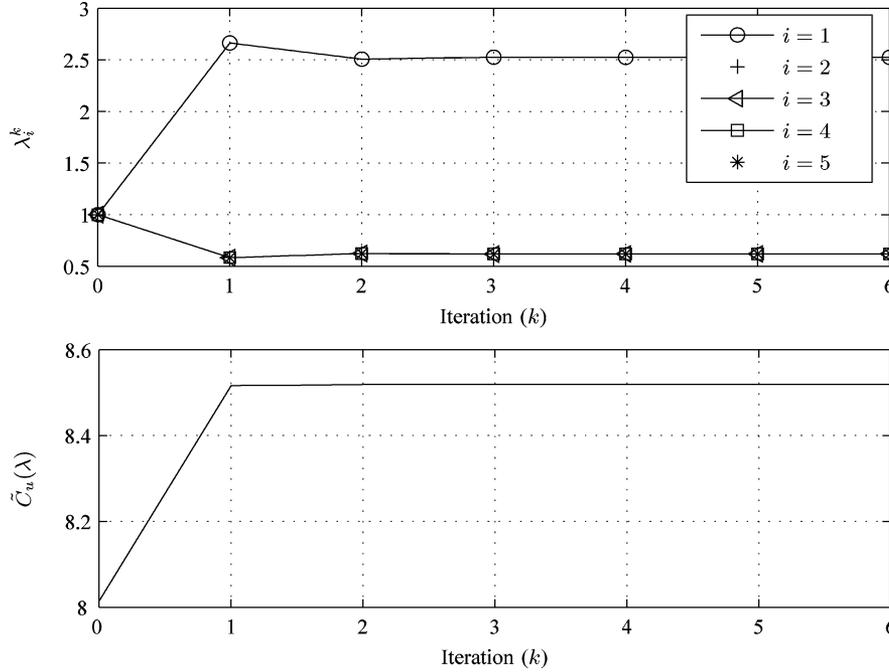
Fig. 6. Convergence of the iterative water-filling algorithm for optimal power allocation in the Kronecker channel. Results are shown for SNR = 10 dB.

where $\delta(\cdot)$ is the Kronecker delta operator, and $\prod_{m=1}^{k} \underline{a}_{\alpha_m, \beta_{\alpha_m}} \neq 0$ only if $\beta_{\alpha_m} > M$ and $k \leq \min(M, N)$. Therefore, we have

$$
\begin{aligned}
\mathrm{Per}([\mathbf{I}_M \ \mathbf{A}]) &= \sum_{k=0}^{\min(M,N)} \sum_{\hat{\alpha}_k \in \mathcal{S}_M^{(k)}} \sum_{\hat{\beta}_k \in \mathcal{S}_N^k} \prod_{m=1}^{k} \underline{a}_{\alpha_m, M+\beta_m} \\
&= \sum_{k=0}^{\min(M,N)} \sum_{\hat{\alpha}_k \in \mathcal{S}_M^{(k)}} \sum_{\hat{\beta}_k \in \mathcal{S}_N^{(k)}} \mathrm{Per}\left(\mathbf{A}_{\hat{\beta}_k}^{\hat{\alpha}_k}\right) \quad (77)
\end{aligned}
$$

where $\mathrm{Per}\left(\mathbf{A}_{\hat{\beta}_k}^{\hat{\alpha}_k}\right) = 1$ when $k = 0$. Using (28), we have

$$
\begin{aligned}
\mathrm{Per}([\mathbf{I}_M \ \mathbf{A}]) &= \sum_{k=0}^{\min(M,N)} \sum_{\hat{\alpha}_k \in \mathcal{S}_M^{(k)}} \mathrm{Per}(\mathbf{A}^{\hat{\alpha}_k}) \\
&= \sum_{k=0}^{\min(M,N)} \sum_{\hat{\beta}_k \in \mathcal{S}_N^{(k)}} \mathrm{Per}(\mathbf{A}_{\hat{\beta}_k}). \quad (78)
\end{aligned}
$$

Through a similar procedure, one can obtain that

$$
\begin{aligned}
\mathrm{Per}([\mathbf{I}_N \ \mathbf{A}^T]) &= \sum_{k=0}^{\min(M,N)} \sum_{\hat{\alpha}_k \in \mathcal{S}_M^{(k)}} \mathrm{Per}(\mathbf{A}^{\hat{\alpha}_k}) \\
&= \sum_{k=0}^{\min(M,N)} \sum_{\hat{\beta}_k \in \mathcal{S}_N^{(k)}} \mathrm{Per}(\mathbf{A}_{\hat{\beta}_k}). \quad (79)
\end{aligned}
$$

This completes the proof.

APPENDIX II
PROOF OF LEMMA 4

From the definition of the determinant, we have

$$
\begin{aligned}
&\mathbb{E}\left\{\det(\mathbf{X}) \det(\mathbf{X}^H)\right\} \\
&= \sum_{\hat{\alpha}_N \in \mathcal{S}_N^N} \sum_{\hat{\beta}_N \in \mathcal{S}_N^N} (-1)^{\sigma(\hat{\alpha}_N)} (-1)^{\sigma(\hat{\beta}_N)} \mathbb{E}\left\{\prod_{i=1}^{N} x_{i,\alpha_i} x_{i,\beta_i}^*\right\} \quad (80)
\end{aligned}
$$

where $\sigma(\hat{\alpha}_N)$ denotes the number of inversions in the permutation $\hat{\alpha}_N$ from the normal order $1, 2, \ldots, N$, and $x_{i,j}$ is the (i, j)th element of $\mathbf{X}$. Since the rows of $\mathbf{X}$ are independent, we have

$$
\mathbb{E}\left\{\prod_{i=1}^{N} x_{i,\alpha_i} x_{i,\beta_i}^*\right\} = \prod_{i=1}^{N} \mathbb{E}\left\{x_{i,\alpha_i} x_{i,\beta_i}^*\right\}. \quad (81)
$$

Since the elements in each row are independent and there is only one possible non-zero mean element in each row, we have

$$
\mathbb{E}\left\{x_{i,\alpha_i} x_{i,\beta_i}^*\right\} = \xi_{i,\alpha_i} \delta(\beta_i - \alpha_i) \quad (82)
$$

where $\xi_{i,j}$ is the (i, j)th element of $\mathbf{\Xi}$. Substituting (82) into (81) and then into (80) yield

$$
\mathbb{E}\left\{\det(\mathbf{X}) \det(\mathbf{X}^H)\right\} = \sum_{\hat{\alpha}_N \in \mathcal{S}_N^N} \prod_{i=1}^{N} \xi_{i,\alpha_i} = \mathrm{Per}(\mathbf{\Xi}). \quad (83)
$$

This completes the proof.

APPENDIX III

PROOF OF THE CONCAVITY OF $f(\boldsymbol{\lambda}) = \log \operatorname{Per}(\mathbf{A}\operatorname{diag}(\boldsymbol{\lambda}))$ IN SEVERAL CASES

Case 1: $M \geq N$*:* In this case, we have that $f(\boldsymbol{\lambda}) = \log \operatorname{Per}(\mathbf{A}) + \log \det(\operatorname{diag}(\boldsymbol{\lambda}))$. The concavity of $f(\boldsymbol{\lambda})$ comes from that of $\log \det(\operatorname{diag}(\boldsymbol{\lambda}))$[41].

Case 2: $M = 1$ *and* $N > 1$*:* In this case, $\mathbf{A}$ is a row vector, and we have that $f(\boldsymbol{\lambda}) = \log(\mathbf{A}\boldsymbol{\lambda})$. The concavity of $f(\boldsymbol{\lambda})$ comes from that of the log function.

Case 3: $M = 2$ *and* $N > 2$*:* In this case, we will first show that the following inequality holds:

$$\frac{g_2(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2)}{g_1(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2)} \geq \frac{g_2(\boldsymbol{\lambda}_1)}{g_1(\boldsymbol{\lambda}_1)} + \frac{g_2(\boldsymbol{\lambda}_2)}{g_1(\boldsymbol{\lambda}_2)} \qquad (84)$$

where $g_1(\boldsymbol{\lambda}) = \mathbf{1}_{1 \times 2}\mathbf{A}\operatorname{diag}(\boldsymbol{\lambda})\mathbf{1}_{N \times 1}$ and $g_2(\boldsymbol{\lambda}) = \operatorname{Per}(\mathbf{A}\operatorname{diag}(\boldsymbol{\lambda}))$. Then we will prove the concavity of $f(\boldsymbol{\lambda})$ from (84).

Since $g_1(\boldsymbol{\lambda})$ and $g_2(\boldsymbol{\lambda})$ are positive on $\mathcal{D}^N$, the inequality (84) holds if and only if the following inequality does:

$$\begin{aligned} g(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = & g_2(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2)g_1(\boldsymbol{\lambda}_1)g_1(\boldsymbol{\lambda}_2) \\ & - g_2(\boldsymbol{\lambda}_1)g_1(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2)g_1(\boldsymbol{\lambda}_2) \\ & - g_2(\boldsymbol{\lambda}_2)g_1(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2)g_1(\boldsymbol{\lambda}_1) \\ \geq & \, 0. \end{aligned} \qquad (85)$$

Let $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2]^T$. Then we have that $g_1(\boldsymbol{\lambda}) = \mathbf{a}_1^T\boldsymbol{\lambda} + \mathbf{a}_2^T\boldsymbol{\lambda}$ and $g_2(\boldsymbol{\lambda}) = \mathbf{a}_1^T\boldsymbol{\lambda}\mathbf{a}_2^T\boldsymbol{\lambda} - \boldsymbol{\lambda}^T\operatorname{diag}(\mathbf{a}_1 \odot \mathbf{a}_2)\boldsymbol{\lambda}$. By substituting these expressions into $g(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$, we can obtain

$$\begin{aligned} g(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = & (\mathbf{a}_1^T\boldsymbol{\lambda}_1\mathbf{a}_2^T\boldsymbol{\lambda}_2 - \mathbf{a}_1^T\boldsymbol{\lambda}_2\mathbf{a}_2^T\boldsymbol{\lambda}_1)^2 \\ & + (g_1(\boldsymbol{\lambda}_2)\boldsymbol{\lambda}_1 - g_1(\boldsymbol{\lambda}_1)\boldsymbol{\lambda}_2)^T\operatorname{diag}(\mathbf{a}_1 \odot \mathbf{a}_2) \\ & \cdot (g_1(\boldsymbol{\lambda}_2)\boldsymbol{\lambda}_1 - g_1(\boldsymbol{\lambda}_1)\boldsymbol{\lambda}_2). \end{aligned} \qquad (86)$$

Therefore we achieve (85) and then (84). From (84), we have

$$\frac{g_2(\theta\boldsymbol{\lambda}_1 + (1-\theta)\boldsymbol{\lambda}_2)}{g_1(\theta\boldsymbol{\lambda}_1 + (1-\theta)\boldsymbol{\lambda}_2)} \geq \theta\frac{g_2(\boldsymbol{\lambda}_1)}{g_1(\boldsymbol{\lambda}_1)} + (1-\theta)\frac{g_2(\boldsymbol{\lambda}_2)}{g_1(\boldsymbol{\lambda}_2)} \qquad (87)$$

where $0 \leq \theta \leq 1$. Taking logarithm on both sides and using the concavity of the log function yields

$$\begin{aligned} f(\theta\boldsymbol{\lambda}_1 &+ (1-\theta)\boldsymbol{\lambda}_2) - \theta f(\boldsymbol{\lambda}_1) - (1-\theta)f(\boldsymbol{\lambda}_2) \\ \geq & \log(g_1(\theta\boldsymbol{\lambda}_1 + (1-\theta)\boldsymbol{\lambda}_2)) + \theta\log(g_1(\boldsymbol{\lambda}_1)) \\ & - (1-\theta)\log(g_1(\boldsymbol{\lambda}_2)) \\ \geq & \, 0. \end{aligned} \qquad (88)$$

This completes the proof of the concavity of $f(\boldsymbol{\lambda})$.

Case 4: $\mathbf{A}$ *is of Rank One:* Let $\mathbf{A} = \mathbf{ab}^T$, where $\mathbf{a}$ and $\mathbf{b}$ are vectors of M and N elements, respectively. In this case, we have

$$\begin{aligned} f(\boldsymbol{\lambda}) = & \log \operatorname{Per}(\mathbf{1}_{M \times N}\operatorname{diag}(\mathbf{b} \odot \boldsymbol{\lambda})) + \log \det(\operatorname{diag}(\mathbf{a})) \\ = & \log \sum_{\hat{\alpha}_M \in \mathcal{S}_N^{(M)}} \operatorname{Per}\left((\mathbf{1}_{M \times N}\operatorname{diag}(\mathbf{b} \odot \boldsymbol{\lambda}))_{\hat{\alpha}_N}\right) \\ & + \log \det(\operatorname{diag}(\mathbf{a})) \\ = & \log E_M(\mathbf{b} \odot \boldsymbol{\lambda}) + \log(M!) + \log \det(\operatorname{diag}(\mathbf{a})) \end{aligned} \qquad (89)$$

where the function $E_M(\boldsymbol{\lambda})$ is the Mth elementary symmetric function defined as [42]

$$E_M(\boldsymbol{\lambda}) = \sum_{\hat{\alpha}_M \in \mathcal{S}_N^{(M)}} \prod_{i=1}^M \lambda_{\alpha_i}. \qquad (90)$$

Since $E_M(\boldsymbol{\lambda})$ is logarithmically concave, we obtain from (89) that $f(\boldsymbol{\lambda})$ is concave.

APPENDIX IV

PROOF OF LEMMA 5

We consider the case with $N_r \leq N_t$. The proof for the case with $N_r > N_t$ is similar. From the definition of the permanents, we have

$$\operatorname{Per}(\hat{\boldsymbol{\Omega}}(z)) = \sum_{\hat{\alpha}_{N_r} \in \mathcal{S}_{N_t}^{N_r}} \prod_{i=1}^{N_r} (1 + \hat{\omega}_{i,\alpha_i}z) \qquad (91)$$

where $\hat{\omega}_{i,j}$ represents the (i, j)th element of $\hat{\boldsymbol{\Omega}}$. For each product term in the above expression, the following relation holds:

$$\prod_{i=1}^{N_r} (1 + \hat{\omega}_{i,\alpha_i}z) = \sum_{k=0}^{N_r} z^k \sum_{\hat{\beta}_k \in \mathcal{S}_{N_r}^{(k)}} \prod_{i=1}^k \hat{\omega}_{\beta_i,\alpha_{\beta_i}}. \qquad (92)$$

Substituting (92) into (91) yields

$$\begin{aligned} \operatorname{Per}(\hat{\boldsymbol{\Omega}}(z)) = & \sum_{k=0}^{N_r} z^k \sum_{\hat{\beta}_k \in \mathcal{S}_{N_r}^{(k)}} \sum_{\hat{\alpha}_{N_r} \in \mathcal{S}_{N_t}^{N_r}} \prod_{i=1}^k \hat{\omega}_{\beta_i,\alpha_{\beta_i}} \\ = & \sum_{k=0}^{N_r} z^k \frac{(N_t - k)!}{(N_t - N_r)!} \sum_{\hat{\beta}_k \in \mathcal{S}_{N_r}^{(k)}} \sum_{\hat{\alpha}_k \in \mathcal{S}_{N_t}^k} \prod_{i=1}^k \hat{\omega}_{\beta_i,\alpha_i} \\ = & \sum_{k=0}^{N_r} z^k \frac{(N_t - k)!}{(N_t - N_r)!} \sum_{\hat{\beta}_k \in \mathcal{S}_{N_r}^{(k)}} \operatorname{Per}(\hat{\boldsymbol{\Omega}}^{\hat{\beta}_k}). \end{aligned} \qquad (93)$$

From Lemma 3, we have the expansion of $\underline{\operatorname{Per}}(\hat{\boldsymbol{\Omega}})$. By comparing the resulting expansion of $\underline{\operatorname{Per}}(\hat{\boldsymbol{\Omega}})$ with (93), we complete the proof.

REFERENCES

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 586–595, Nov./Dec. 1999.

[2] 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.

[3] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge University Press, 2003.

[4] A. B. Gershman and N. Sidiropoulos, *Space-Time Processing for MIMO Communications*. New York: Wiley, 2005.

[5] G. B. Giannakis, Z. Q. Liu, X. L. Ma, and S. L. Zhou, *Space-Time Coding for Broadband Wireless Communications*. New York: Wiley, 2007.

[6] G. L. Stuber, J. R. Barry, S. W. Mclaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, pp. 271–294, Feb. 2004.

[7] G. Barriac and U. Madhow, "Space-time communication for OFDM with implicit channel feedback," *IEEE Trans. Inform. Theory*, vol. 50, pp. 3111–3129, Dec. 2004.

[8] G. Barriac and U. Madhow, "Space-time precoding for mean and convariance feedback: Application to wideband OFDM," *IEEE Trans. Commun.*, vol. 54, pp. 96–107, Jan. 2006.

[9] D. S. Shui, 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, pp. 502–513, Mar. 2000.

[10] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering, and keyhole," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2636–2647, Oct. 2003.

[11] M. Chiani, M. Z. Win, and A. Zanella, "On the capacity of spatially correlated MIMO Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, pp. 2363–2371, Oct. 2003.

[12] X. W. Cui, Q. T. Zhang, and Z. M. Feng, "Generic procedure for tightly bounding the capacity of MIMO correlated Rician fading channels," *IEEE Trans. Commun.*, vol. 53, pp. 890–898, May 2005.

[13] S. Jin, X. Q. Gao, and X. H. You, "On the ergodic capacity of rank-1 Ricean fading MIMO channels," *IEEE Trans. Inf. Theory*, vol. 53, pp. 502–517, Feb. 2007.

[14] M. R. McKay and I. B. Collings, "General capacity bounds for spatially correlated Rician MIMO channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3121–3145, Sep. 2005.

[15] S. A. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, Jul. 2004.

[16] E. A. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1543–1553, Sep. 2004.

[17] M. Kiessling, J. Speidel, I. Viering, and M. Reinhardt, "A closed-form bound on correlated MIMO channel capacity," in *Proc. IEEE VTC 2002-Fall*, Vancouver, BC, Canada, Sep. 2002, vol. 2, pp. 859–863.

[18] M. Kiessling, "Statistical Analysis and Transmit Prefiltering for MIMO Systems in Correlated Fading Environments," Ph.D. dissertation, University of Stuttgart, Institute of Telecommunications, Stuttgart, Germany, 2004.

[19] E. A. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, pp. 3340–3453, Dec. 2004.

[20] J. P. Kermoal, L. Schumacher, K. I. Pedersen, and P. Mogensen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, pp. 1211–1226, Jun. 2002.

[21] W. Weichselberger, M. Herdin, H. Ozcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, vol. 5, no. 1, pp. 90–100, Jan. 2006.

[22] Y. Zhou, M. Herdin, A. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," *IEEE Trans. Antennas Propagat.*, 2006.

[23] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, pp. 2563–2579, Oct. 2002.

[24] V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO Rayleigh fading channels: Capacity, optimal signaling and asymptotics," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2058–2072, Jun. 2005.

[25] A. M. Tulino, A. Lozano, and S. Verdú, "Impact of antenna correlation on the capacity of multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 2491–2509, Jul. 2005.

[26] A. M. Tulino, A. Lozano, and S. Verdú, "Capacity-achieving input convariance for single-user multi-antena channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 662–671, Mar. 2006.

[27] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol. 47, pp. 2632–2639, Sep. 2001.

[28] S. Venkatesan, S. H. Simon, and R. A. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," in *Proc. IEEE VTC 2003-Fall*, Orlando, FL, Oct. 2003, vol. 3, pp. 1767–1771.

[29] D. Hösli and A. Lapidoth, "The capacity of a MIMO Rician channel is monotonic in the singular values of the mean," in *Proc. 5th Int. ITG Conf. Source and Channel Coding*, Erlangen, Germany, Jan. 2004, pp. 381–385.

[30] M. Vu and A. Paulraj, "Charactering the capacity for MIMO wireless channels with non-zero mean and transmit covariance," in *Proc. 43rd Ann. Allerton Conf. Commun., Control, and Comp.*, Monticello, IL, USA, Sep. 2005.

[31] L. W. Hanlen and A. J. Grant, "Optimal transmit covariance for MIMO channels with statistical transmitter side information," in *Proc. IEEE ISIT 2005*, Adelaide, Australia, Sep. 2005, pp. 1818–1822.

[32] 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, pp. 145–152, Jan. 2004.

[33] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1570–1580, Apr. 2005.

[34] X. H. You, G. A. Chen, M. Chen, and X. Q. Gao, "Toward beyond 3G: the FuTURE project of China," *IEEE Signal Processing Mag.*, vol. 43, no. 1, pp. 70–75, Jan. 2005.

[35] H. Minc, *Permanents*. Reading, MA: Addison-Wesley, 1978.

[36] H. J. Ryser, *Combinatorial Mathematics*. New York: Wiley, Mathematical Association of America, 1963.

[37] A. Nijenhuis and H. S. Wilf, *Combinatorial Algorithms for Computers and Calculators*, 2nd ed. New York: Academic, 1978, ch. 23.

[38] H. Liang and F. S. Bai, "A partially structure-preserving algorithm for the permanents of adjacency matrices of fullerene," *Comput. Phys. Commun.*, vol. 163, no. 2, pp. 79–84, Nov. 2004.

[39] C. Aitken, *Determinants and Matrices*. Edinburgh/London, U.K.: Oliver and Boyd, 1964.

[40] A. Grant, "Rayleigh fading multiple-antenna channels," *EURASIP J. Appl. Signal Process.*, no. 3, pp. 316–329, Mar. 2002.

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

[42] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*. New York: Academic, 1979.

Xiqi Gao (SM'07) received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 1997.

He joined the Department of Radio Engineering, Southeast University, in April 1992. Now he is a professor of information systems and communications. From September 1999 to August 2000, he was a Visiting Scholar at Massachusetts Institute of Technology, Cambridge, and Boston University, Boston, MA. From August 2007 to July 2008, he visited the Darmstadt University of Technology, Darmstadt, Germany, as a Humboldt Scholar. His current research interests include broadband multi-carrier transmission for beyond 3G mobile communications, space–time wireless communications, iterative detection/decoding, signal processing for wireless communications. He is currently serving as an editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Dr. Gao received the Science and Technology Progress Awards of the State Education Ministry of China in 1998 and 2006.

Bin Jiang (S'06) received the B.S. degree in electrical engineering from Southeast University, Nanjing, China, in 2002. He is currently working toward the Ph.D. degree in communication and signal processing with the School of Information Science and Engineering, Southeast University, Nanjing, China.

His research interests lie in the area of signal processing and wireless communication for MIMO systems.

Xiao Li (S'06) was born in Anhui, China, in 1982. She received the Bachelor's degree in Radio Engineering from Southeast University, Nanjing, Jiangsu, China, in 2003. She is currently working toward the Ph.D. degree in the National Mobile Communications Research Lab, Southeast University.

Her research interests include information theory and eigenmode transmission strategies for single-user and multi-user MIMO systems.

Alex B. Gershman (M'97–SM'98–F'06) received his Diploma and Ph.D. degrees in Radiophysics and Electronics from the Nizhny Novgorod State University, Russia, in 1984 and 1990, respectively.

From 1984 to 1999, he held several full-time and visiting research appointments in Russia, Switzerland, and Germany. In 1999, he joined the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, where he became a Professor in 2002. From April 2005, he has been with the Darmstadt University of Technology, Darmstadt, Germany, as a professor of Communication Systems. His research interests are in the area of signal processing and communications with the primary emphasis on sensor array processing, beamforming, MIMO and multiuser communications, and estimation and detection theory.

Dr. Gershman is the recipient of several awards including the 2004 IEEE Signal Processing Society (SPS) Best Paper Award; the IEEE Aerospace and Electronic Systems Society (AESS) Barry Carlton Award for the best paper published in 2004, the 2002 Young Explorers Prize from the Canadian Institute for Advanced Research (CIAR); the 2001 Wolfgang Paul Award from the Alexander von Humboldt Foundation, Germany; and the 2000 Premier's Research Excellence Award, Ontario, Canada. He also coauthored the paper that received the 2005 IEEE SPS Young Author Best Paper Award. He was Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS (2006–2008), Chair of the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE SPS (2006–2008), and Member of the IEEE SPS Publications, Conference, and Technical Directions Boards (2006–2008). He is currently a member of the IEEE SPS Board of Governors. He is on the editorial board of the IEEE SIGNAL PROCESSING MAGAZINE and several other technical journals. He was Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING (1999–2005); Technical Co-Chair of the IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Darmstadt, Germany, December 2003; Technical Co-Chair of the 4th IEEE Sensor Array and Multichannel (SAM) Signal Processing Workshop, Waltham, MA, USA, June 2006; General Co-Chair of the 1st IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Puerto Vallarta, Mexico, December 2005; Tutorial Chair of EUSIPCO, Florence, Italy, September 2006; and General Co-Chair of the 5th IEEE SAM Workshop, Darmstadt, Germany, July 2008.

Matthew R. McKay (S'03–M'07) received the combined B.E. degree in electrical engineering and B.IT. degree in computer science from the Queensland University of Technology, Australia, in 2002 and the Ph.D. degree in electrical engineering from the University of Sydney, Australia, in 2006. He then worked as a Research Scientist at the Commonwealth Science and Industrial Research Organization (CSIRO), Sydney, Australia, prior to joining the faculty at the Hong Kong University of Science and Technology (HKUST) in 2007, where he is currently an Assistant Professor. He is also a member of the Center for Wireless Information Technology at HKUST.

His research interests include communications and signal processing; in particular the analysis and design of MIMO systems, random matrix theory, information theory, and wireless ad-hoc and sensor networks.

Dr. McKay was awarded a 2006 Best Student Paper Award at IEEE ICASSP'06, and was jointly awarded the 2006 Best Student Paper Award at IEEE VTC'06-Spring. He was also awarded the University Medal upon graduating from the Queensland University of Technology.