

RESEARCH

Open Access

On base station cooperation using statistical CSI in jointly correlated MIMO downlink channels

Jun Zhang¹, Bin Jiang¹, Shi Jin^{1*}, Xiqi Gao¹ and Kai-Kit Wong²

Abstract

This article studies the transmission of a single cell-edge user's signal using statistical channel state information at cooperative base stations (BSs) with a general jointly correlated multiple-input multiple-output (MIMO) channel model. We first present an optimal scheme to maximize the ergodic sum capacity with per-BS power constraints, revealing that the transmitted signals of all BSs are mutually independent and the optimum transmit directions for each BS align with the eigenvectors of the BS's own transmit correlation matrix of the channel. Then, we employ matrix permanents to derive a closed-form tight upper bound for the ergodic sum capacity. Based on these results, we develop a low-complexity power allocation solution using convex optimization techniques and a simple iterative water-filling algorithm (IWFA) for power allocation. Finally, we derive a necessary and sufficient condition for which a beamforming approach achieves capacity for all BSs. Simulation results demonstrate that the upper bound of ergodic sum capacity is tight and the proposed cooperative transmission scheme increases the downlink system sum capacity considerably.

Keywords: base station cooperation, ergodic sum capacity, permanent, power allocation

1 Introduction

Multi-antenna systems, widely known as multiple-input multiple-output (MIMO), have shown considerable gain in spectral efficiency and attracted much attention in recent years, e.g., [1]. There have also been strong interests in utilizing MIMO to cope with multiuser scenarios. However, achieving the theoretical capacity gains in practical cellular environments is problematic because of interferences. More recently, base station (BS) cooperation [2-10] was proposed as a means to improve the performance of cell-edge user and mitigate the problem of inter-cell interference. This is greatly motivated by the fact that BSs may be connected via a wired backbone and the channel state information (CSI) can be shared among the BSs for coordinated transmission. Such BS cooperation in the downlink in particular leads to enormous throughput gains as compared to the conventional single-BS (or single-cell) signal processing where the co-channel interference (often from other cells) is treated as noise.

Coordinated BS transmission in the downlink is often analyzed using a large MIMO Gaussian broadcast channel (BC) model, with the challenge of incorporating per-antenna or per-BS power constraints. MIMO BC capacity region with a sum-power constraint has been well established in [11-14] using the uplink-downlink duality. The achievable rate region of MIMO BC under per-antenna power constraints has also been studied in [15]. Recently, this result has been extended to cope with general linear transmit power constraints in [16,17]. If the dirty-paper-coding based optimal nonlinear precoder or the minimum-mean-squared-error based optimal linear pre-coder is used, the result in [15-17] can be directly applied to BS cooperation with per-BS power constraints. However, the gain offered by BS cooperation depends greatly on the level of CSI that can be exploited in the optimization. To investigate the full diversity and multiplexing benefits, most previous studies [15-17] assumed to possess perfect CSI at the transmitter, but such results may be severely offset by expensive overheads for acquiring the CSI [18-20]. User mobility also will increase fading rate and make accurate CSI difficult to maintain. For this reason, exploiting statistical CSI at the transmitter side is often more

* Correspondence: jinshi@seu.edu.cn

¹National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China

Full list of author information is available at the end of the article

appealing due to much lower overhead. The uplink and downlink statistical CSI are also usually reciprocal in both frequency-division-duplex and time-division-duplex systems [21,22].

Capacity analysis and transceiver designs using statistical CSI at the transmitter are highly dependent on the assumption of the channel model. The conventional modeling approach has been the Kronecker model [23-25] which separates spatial correlation at the transmitter and receiver ends. Recent measurement campaigns, however, demonstrated that mutual correlation between the transmitter and receiver may be important, which makes the Kronecker model inadequate [26,27]. The jointly-correlated channel model in contrast not only accounts for the correlation at both ends, but also characterizes their mutual dependence. Recently, [28] derived a closed-form upper bound for the ergodic capacity of the jointly-correlated MIMO channel. Beamforming is a simple linear precoding strategy, in which the transmit covariance matrix is of unit-rank. The optimality of beamforming for some single-user MIMO channels have been studied for the Kronecker model [29,30], the double-scattering model [31], and the virtual representation model [32], when the transmitter has partial knowledge of the channel. These results have also been extended to the MIMO multiple access channel (MAC) in [32-36].

In this article, we aim to investigate coordinated downlink transmission with cooperative BSs assuming that the mobile user has perfect CSI but the BSs know only statistical CSI. Our main contribution is that the jointly correlated MIMO channel model in [26] is adopted to account for the spatial correlations of the antennas at the BSs and the user and between them. We first present an optimal transmission scheme to maximize the ergodic sum capacity of this channel with per-BS power constraints, from which two important results are revealed: (i) the transmit signals of all BSs are mutually independent; and (ii) the optimal transmit directions for each BS align with the eigen-directions of the BS's own transmit-side correlation matrix. We then employ matrix permanents to derive a closed-form tight upper bound for the ergodic sum capacity of the jointly correlated MIMO channel. Based on this bound, we propose an iterative power allocation algorithm using convex optimization techniques, which converges within only a few iterations. Also, we establish the beamforming optimality conditions for all the BSs. Our study for BS cooperation in the jointly-correlated MIMO downlink channel generalizes the result in [33].

The rest of this article is organized as follows. We present the system model in Section 2. In Section 3, we propose the capacity-achieving optimal transmit scheme and derive the ergodic sum capacity upper bound. Utilizing the bound, we then develop the optimal power

allocation policies. Finally, we establish the beamforming optimality conditions for all BSs. Simulation results are presented in Section 4, and we conclude the article in Section 5.

Notations

We use uppercase and lowercase boldface letters to denote matrices and vectors, respectively. \mathbf{I}_N is an $N \times N$ identity matrix and $\mathbf{0}$ denotes an all-zero matrix, while $\mathbf{1}$ is an all-one matrix. The matrix inequality \succcurlyeq shows the positive semi-definiteness. The superscripts $(\cdot)^H$, $(\cdot)^T$, and $(\cdot)^*$ represent the conjugate-transpose, transpose, and conjugate operations, respectively. We use $\mathbb{E}\{\cdot\}$ to denote expectation with respect to all random variables within the brackets, and use $\mathbf{A} \boxtimes \mathbf{B}$ to denote the Hadamard product of \mathbf{A} and \mathbf{B} . We use $[\mathbf{A}]_{kl}$ or the lower-case representation a_{kl} to denote the (k,l) th entry of \mathbf{A} , and a_k denotes the k th entry of the column vector \mathbf{a} . The operators $\text{tr}(\cdot)$, $\det(\cdot)$, and $\text{Per}(\cdot)$ represent the matrix trace, determinant, and permanent, respectively, and $\text{diag}(\mathbf{x})$ denotes a diagonal matrix with \mathbf{x} along its main diagonal.

2 System model

We consider a downlink cellular network consisting of m BSs, labeled as $\text{BS}_1, \dots, \text{BS}_m$, which are equipped with $\{N_t^{(1)}, N_t^{(2)}, \dots, N_t^{(m)}\}$ antennas, respectively, and a single cell-edge user with N_r antennas. To improve the performance of the cell-edge user, the BSs are connected by a wired backbone that allows information to be reliably exchanged among them. At the mobile user in the baseband, the received signal can be written in vector form as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where $\mathbf{x} \triangleq [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_m^T]^T$ with \mathbf{x}_i being the $N_t^{(i)} \times 1$ transmitted signal vector of BS_i ; \mathbf{n} is the $N_r \times 1$ zero-mean additive complex Gaussian noise vector with $\mathbb{E}\{\mathbf{n}\mathbf{n}^H\} = N_0\mathbf{I}_{N_r}$; $\mathbf{H} \triangleq [\sqrt{\Gamma_1}\mathbf{H}_1, \sqrt{\Gamma_2}\mathbf{H}_2, \dots, \sqrt{\Gamma_m}\mathbf{H}_m]$ with \mathbf{H}_i being the $N_r \times N_t^{(i)}$ MIMO channel matrix between BS_i and the user; and Γ_i denotes the large-scale fading between BS_i and the user. It is assumed that \mathbf{x}_i and \mathbf{H}_i satisfy the following power constraints

$$\mathbb{E}\{\text{tr}(\mathbf{x}_i\mathbf{x}_i^H)\} = P_i, \quad (2)$$

and

$$\mathbb{E}\{\text{tr}(\mathbf{H}_i\mathbf{H}_i^H)\} = N_t^{(i)}N_r, \quad \text{for } i = 1, 2, \dots, m. \quad (3)$$

We find it useful to define the total transmitted power as $P \triangleq \sum_{i=1}^m P_i$, the total number of transmit antennas

as $N_t \triangleq \sum_{i=1}^m N_t^{(i)}$, and the transmit signal-to-noise ratio (SNR) as $\rho \triangleq \frac{P}{N_0}$.

In this article, we consider the jointly-correlated MIMO channel model [28], given by

$$\mathbf{H}_i = \mathbf{U}_r \tilde{\mathbf{H}}_i \mathbf{U}_{t,i}^H \quad \text{for } i = 1, 2, \dots, m, \quad (4)$$

where

$$\tilde{\mathbf{H}}_i = \mathbf{D}_i + \mathbf{M}_i \odot \mathbf{H}_{\text{iid},i} \quad (5)$$

$\mathbf{U}_{t,i}$ and \mathbf{U}_r are $N_t^{(i)} \times N_t^{(i)}$ and $N_r \times N_r$ deterministic unitary matrices, \mathbf{D}_i is an $N_r \times N_t^{(i)}$ deterministic matrix, \mathbf{M}_i is an $N_r \times N_t^{(i)}$ deterministic matrix with nonnegative elements, and $\mathbf{H}_{\text{iid},i}$ for $i = 1, 2, \dots, m$ are statistically independent $N_r \times N_t^{(i)}$ random matrices of independent and identically distributed (i.i.d.) zero-mean unit-variance entries. Note that $\mathbf{H}_{\text{iid},i}$ is not necessarily Gaussian. The matrices \mathbf{D}_i and \mathbf{M}_i reflect the line-of-sight (LOS) and scattering components of the channel, respectively. Also, we define $\mathbf{D} \triangleq [\sqrt{\Gamma_1} \mathbf{D}_1, \sqrt{\Gamma_2} \mathbf{D}_2, \dots, \sqrt{\Gamma_m} \mathbf{D}_m]$, which has at most one nonzero element in each row and each column. Without loss of generality, we assume that the nonzero elements of \mathbf{D} are real, with indices (l, l) for $1 \leq l \leq \min(N_r, N_t)$.

From (4), the transmit correlation matrices of the BSs and the receive correlation matrix of the user can be expressed, respectively, as

$$\mathbf{R}_{t,i} = \mathbb{E} \{ \mathbf{H}_i^H \mathbf{H}_i \} = \mathbf{U}_{t,i} \Phi_{t,i} \mathbf{U}_{t,i}^H \quad \text{for } i = 1, 2, \dots, m, \quad (6)$$

and

$$\mathbf{R}_r = \mathbb{E} \{ \mathbf{H} \mathbf{H}^H \} = \mathbf{U}_r \Phi_r \mathbf{U}_r^H, \quad (7)$$

where $\Phi_{t,i}$ and Φ_r are diagonal matrices, $\mathbf{U}_{t,i}$ and \mathbf{U}_r are the eigenvector matrices of the transmit and receive correlation matrices, respectively.

For ease of exposition, the channel coupling matrix is usually defined as [26]

$$\mathbf{\Omega}_i = \mathbb{E} \{ \tilde{\mathbf{H}}_i \odot \tilde{\mathbf{H}}_i^* \}, \quad \text{for } i = 1, 2, \dots, m. \quad (8)$$

Substituting (5) into (8) yields

$$\mathbf{\Omega}_i = \mathbf{D}_i \odot \mathbf{D}_i + \mathbf{M}_i \odot \mathbf{M}_i, \quad \text{for } i = 1, 2, \dots, m, \quad (9)$$

and the power constraint (3) can be rewritten as

$$\sum_{k=1}^{N_r} \sum_{l=1}^{N_t^{(i)}} \omega_{kl}^{(i)} = N_t^{(i)} N_r, \quad \text{for } i = 1, 2, \dots, m. \quad (10)$$

In the above, the (k,l) th element of $\mathbf{\Omega}_i$, i.e., $\omega_{kl}^{(i)}$, corresponds to the average power of the (k,l) th element of $\tilde{\mathbf{H}}_i$, i.e., $\tilde{h}_{kl}^{(i)}$, which captures the average coupling between the k th receive eigenmode and the l th transmit eigenmode of BS_{*i*}.

3 Statistical CSI-aided coordinated BS transmission

Here, we first devise the optimal transmit scheme for each BS, and then derive a closed-form upper bound for the ergodic sum capacity using the matrix permanents. Based on the capacity bound, we develop low-complexity power allocation solutions using convex optimization techniques, followed by discussion of the beamforming optimality conditions for the BSs.

3.1 Optimal transmit scheme

We here assume that the mobile receiver has perfect instantaneous CSI, whereas the BSs only have the statistical CSI including $\mathbf{U}_{t,i}$, \mathbf{U}_r , \mathbf{D}_i and \mathbf{M}_i (and thus $\mathbf{\Omega}_i$) ($i = 1, 2, \dots, m$), and this information can be exchanged among the BSs via the wired backbone. Under these assumptions, the ergodic sum capacity of the downlink system is achieved by selecting the transmitted signal vector \mathbf{x} to follow a zero-mean proper Gaussian distribution [1]. Let $\mathbb{E} \{ \mathbf{x} \mathbf{x}^H \} = \frac{P}{N_t} \mathbf{Q}$, where

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} & \cdots & \mathbf{Q}_{1m} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} & \cdots & \mathbf{Q}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{Q}_{m1} & \mathbf{Q}_{m2} & \cdots & \mathbf{Q}_{mm} \end{pmatrix}, \quad (11)$$

with $\mathbf{Q}_{ij} = \frac{N_r}{P} \mathbb{E} \{ \mathbf{x}_i \mathbf{x}_j^H \}$. The power constraint (2) can be rewritten as $\text{tr}(\mathbf{Q}_{ii}) = \frac{P_i N_r}{P} = P'_i$, for $i = 1, 2, \dots, m$, and the ergodic sum capacity is given by

$$C = \max_{\substack{\text{tr}(\mathbf{Q}_{ii})=P'_i \forall i \\ \mathbf{Q}_{ii} \succeq 0 \forall i}} \mathbb{E} \{ \log \det (\mathbf{I}_{N_r} + \gamma \mathbf{H} \mathbf{Q} \mathbf{H}^H) \}, \quad (12)$$

where $\gamma = \frac{\rho}{N_t}$. Let $\mathbf{Q}_{ii} = \mathbf{U}_i \mathbf{\Lambda}_i \mathbf{U}_i^H$, for $i = 1, \dots, m$, with \mathbf{U}_i being the eigenvector matrix, and $\mathbf{\Lambda}_i = \text{diag}(\lambda_1^{(i)}, \lambda_2^{(i)}, \dots, \lambda_{N_t^{(i)}}^{(i)})$ the diagonal matrix of the corresponding eigenvalues. The following theorem addresses the optimal transmit direction of each BS.

Theorem 1 *The ergodic sum capacity is achieved if the BS transmit signals are all mutually independent (i.e., $\mathbf{Q}_{ij}^{\text{opt}} = \mathbf{0}$, for $i \neq j$), and the eigenvector matrix of $\mathbf{Q}_{ii}^{\text{opt}}$ for the jointly-correlated channel (4) is given by $\mathbf{U}_i = \mathbf{U}_{t,i}$. The ergodic sum capacity is then expressed as*

$$C = \max_{\substack{\text{tr}\{\Lambda_i\}=P_i \forall i \\ \Lambda_i \succcurlyeq 0 \forall i}} \mathbb{E} \left\{ \log \det \left(\mathbf{I}_{N_r} + \gamma \sum_{i=1}^m \Gamma_i \tilde{\mathbf{H}}_i \Lambda_i \tilde{\mathbf{H}}_i^H \right) \right\} \quad (13)$$

Proof: Form (4) and (5), the channel matrix \mathbf{H} can be expressed as

$$\mathbf{H} = \mathbf{U}_r \tilde{\mathbf{H}} \mathbf{U}_t^H, \quad (14)$$

$$\tilde{\mathbf{H}} = \mathbf{D} + \mathbf{M} \odot \mathbf{H}_{\text{iid}}, \quad (15)$$

where

$$\tilde{\mathbf{H}} = [\sqrt{\Gamma_1} \tilde{\mathbf{H}}_1, \sqrt{\Gamma_2} \tilde{\mathbf{H}}_2, \dots, \sqrt{\Gamma_m} \tilde{\mathbf{H}}_m], \mathbf{U}_t = \text{diag} \{ \mathbf{U}_{t,1}, \mathbf{U}_{t,2}, \dots, \mathbf{U}_{t,m} \},$$

$$\mathbf{M} = [\sqrt{\Gamma_1} \mathbf{M}_1, \sqrt{\Gamma_2} \mathbf{M}_2, \dots, \sqrt{\Gamma_m} \mathbf{M}_m], \mathbf{H}_{\text{iid}} = [\mathbf{H}_{\text{iid},1}, \mathbf{H}_{\text{iid},2}, \dots, \mathbf{H}_{\text{iid},m}].$$

Defining $\tilde{\mathbf{Q}} \triangleq \mathbf{U}_t^H \mathbf{Q} \mathbf{U}_t$ and substituting (14) into (12) yields

$$C = \max_{\substack{\text{tr}\{\tilde{\mathbf{Q}}_{ii}\}=P_i \forall i \\ \tilde{\mathbf{Q}}_{ii} \succcurlyeq 0 \forall i}} \mathcal{I}(\tilde{\mathbf{Q}}), \quad (16)$$

where

$$\mathcal{I}(\tilde{\mathbf{Q}}) = \mathbb{E} \left\{ \log \det \left(\mathbf{I}_{N_r} + \gamma \tilde{\mathbf{H}} \tilde{\mathbf{Q}} \tilde{\mathbf{H}}^H \right) \right\}. \quad (17)$$

Note that the optimization condition is met since $\text{tr} \{ \tilde{\mathbf{Q}}_{ii} \} = \text{tr} \{ \mathbf{Q}_{ii} \}$. Now, define $\mathbf{\Pi}_l$ for $1 \leq l \leq N_t$ as diagonal matrices all of which have their diagonal entries being all 1s except for the (l,l) th entry as -1. As $\mathbf{\Pi}_l$ is a unitary matrix, (17) can be written as

$$\mathcal{I}(\tilde{\mathbf{Q}}) = \mathbb{E} \left\{ \log \det \left(\mathbf{I}_{N_r} + \gamma \prod_l \tilde{\mathbf{H}} \mathbf{\Pi}_l \tilde{\mathbf{Q}} \mathbf{\Pi}_l^H \prod_l \tilde{\mathbf{H}}^H \right) \right\} \text{ for any } \prod_l. \quad (18)$$

Note that $\tilde{\mathbf{H}}$ is given by (15) and $\prod_l \tilde{\mathbf{H}} \mathbf{\Pi}_l$ has the same distribution as $\tilde{\mathbf{H}}$, since \mathbf{D} is a diagonal matrix plus the fact that the entries of $\mathbf{M} \boxtimes \mathbf{H}_{\text{iid}}$ are independent and their distributions being symmetric, reversing the sign of some columns does not alter the distribution. Thus, we have

$$\mathcal{I}(\tilde{\mathbf{Q}}) = \mathbb{E} \left\{ \log \det \left(\mathbf{I}_{N_r} + \gamma \tilde{\mathbf{H}} \mathbf{\Pi}_l^H \tilde{\mathbf{Q}} \mathbf{\Pi}_l \tilde{\mathbf{H}}^H \right) \right\} = \mathcal{I}(\mathbf{\Pi}_l^H \tilde{\mathbf{Q}} \mathbf{\Pi}_l). \quad (19)$$

From Jensen's inequality, it follows that [37-39]

$$\mathcal{I} \left(\frac{1}{2} (\tilde{\mathbf{Q}} + \mathbf{\Pi}_l \tilde{\mathbf{Q}} \mathbf{\Pi}_l^H) \right) \geq \frac{\mathcal{I}(\tilde{\mathbf{Q}}) + \mathcal{I}(\mathbf{\Pi}_l \tilde{\mathbf{Q}} \mathbf{\Pi}_l^H)}{2} \quad (20)$$

$$= \mathcal{I}(\tilde{\mathbf{Q}}),$$

where the matrix $\frac{1}{2} (\tilde{\mathbf{Q}} + \mathbf{\Pi}_l \tilde{\mathbf{Q}} \mathbf{\Pi}_l^H)$ has entries equal to those of $\tilde{\mathbf{Q}}$ except for the off-diagonals in the l th row

and l th column, which are zero. In particular, its trace is identical to that of $\tilde{\mathbf{Q}}$. As a result, nulling the off-diagonal entries of any column and the corresponding row of $\tilde{\mathbf{Q}}$ can only increase $\mathcal{I}(\tilde{\mathbf{Q}})$. Using the same process N_t times, (17) is maximized with a diagonal $\tilde{\mathbf{Q}}$, i.e., $\tilde{\mathbf{Q}} = \Lambda$. As a result, we have $\mathbf{Q}^{\text{opt}} = \mathbf{U}_t \Lambda \mathbf{U}_t^H$, or

$$\mathbf{Q}_{ij}^{\text{opt}} = \begin{cases} \mathbf{U}_{t,i} \Lambda_i \mathbf{U}_{t,i}^H & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \quad (21)$$

where $\Lambda = \text{diag} \{ \Lambda_1, \Lambda_2, \dots, \Lambda_m \}$. As such, (16) can be rewritten as (13).

Theorem 1 reveals that the transmitted signals of all BSs should be mutually independent and the optimal signaling directions of the i -th BS align with the eigenvectors of the transmit-side correlation matrix of the MIMO channel of the i -th BS. This results extend the prior results in [29,37,40,41] to the more general channel model given by (4).

3.2 Ergodic sum capacity upper bound

After knowing the optimal transmit directions of the BSs, the remaining challenge is to determine the eigenvalues of the capacity-achieving input covariance matrix \mathbf{Q}_{ii} for $i = 1, \dots, m$. This is equivalent to optimally allocating the available transmit power over the optimized transmit eigen-directions that are determined by *Theorem 1*.

In the most general case, it is difficult to derive exact closed-form solutions for the power allocation problem. The main obstacle lies in the complexity in evaluating the expectation in (13) which is usually done by stochastic averaging over a large number of random samples. In this section, our approach is to derive a tight upper bound for the expectation in (13) which can serve as an approximation to the ergodic capacity. Based on this, we develop closed-form power allocation solutions which will be presented in Section 3.3.

Due to the concavity of the $\log(\cdot)$ function, C is upper bounded by

$$C \leq C_u = \max_{\substack{\text{tr}\{\Lambda_i\}=P_i \forall i \\ \Lambda_i \succcurlyeq 0 \forall i}} C_u(\boldsymbol{\lambda}), \quad (22)$$

where

$$C_u(\boldsymbol{\lambda}) = \log \mathbb{E} \left\{ \det \left(\mathbf{I}_{N_r} + \gamma \sum_{i=1}^m \Gamma_i \tilde{\mathbf{H}}_i \Lambda_i \tilde{\mathbf{H}}_i^H \right) \right\}, \quad (23)$$

with $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1^T, \boldsymbol{\lambda}_2^T, \dots, \boldsymbol{\lambda}_m^T]^T$, in which $\boldsymbol{\lambda}_i$ denotes an $N_t^{(i)} \times 1$ vector containing the eigenvalues $\lambda_{ij}^{(i)}$ for $j = 1, \dots, N_t^{(i)}$ and $i = 1, \dots, m$. The upper bound (22) can be rewritten as

$$C \leq C_u = \max_{\substack{1^T \boldsymbol{\lambda}_i = P_i' \forall i \\ \lambda_i \geq 0 \forall i}} C_u(\boldsymbol{\lambda}). \quad (24)$$

The expectation derivation in (23) is based heavily on exploiting linear-algebraic concepts and the properties of matrix permanents. 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. The permanents of $M \times N$ matrices have been investigated in [28,42]. We introduce the definitions and properties of matrix permanents in Appendix 1.

From these definitions, we extend the results of [28] to the case of multiple BSs. We can derive a closed-form expression for the upper bound on the ergodic sum capacity.

Theorem 2 *The ergodic sum capacity in (13) is upper bounded by*

$$C \leq C_u = \max_{\substack{1^T \boldsymbol{\lambda}_i = P_i' \forall i \\ \lambda_i \geq 0 \forall i}} \tilde{C}_u(\boldsymbol{\lambda}), \quad (25)$$

where $\tilde{C}_u(\boldsymbol{\lambda}) = \log \underline{\text{Per}}(\gamma [\Gamma_1 \boldsymbol{\Omega}_1 \boldsymbol{\Lambda}_1, \dots, \Gamma_m \boldsymbol{\Omega}_m \boldsymbol{\Lambda}_m])$.

Proof: We start by letting

$$E(\boldsymbol{\lambda}) \triangleq \mathbb{E} \left\{ \det \left(\mathbf{I}_{N_i} + \gamma \sum_{i=1}^m \Gamma_i \tilde{\mathbf{H}}_i \mathbf{A}_i \tilde{\mathbf{H}}_i^H \right) \right\} = \mathbb{E} \left\{ \det \left(\mathbf{I}_{N_i} + \gamma \tilde{\mathbf{H}} \text{diag}(\boldsymbol{\lambda}) \tilde{\mathbf{H}}^H \right) \right\}, \quad (26)$$

where $\tilde{\mathbf{H}} = [\sqrt{\Gamma_1} \tilde{\mathbf{H}}_1, \sqrt{\Gamma_2} \tilde{\mathbf{H}}_2, \dots, \sqrt{\Gamma_m} \tilde{\mathbf{H}}_m]$. The upper bound of mutual information (23) can be rewritten as

$$C_u(\boldsymbol{\lambda}) = \log E(\boldsymbol{\lambda}). \quad (27)$$

By using the known result [28, Theorem 2], $E(\boldsymbol{\lambda})$ can be expressed as

$$E(\boldsymbol{\lambda}) = \underline{\text{Per}}(\gamma [\Gamma_1 \boldsymbol{\Omega}_1, \dots, \Gamma_m \boldsymbol{\Omega}_m] \text{diag}(\boldsymbol{\lambda})) = \underline{\text{Per}}(\gamma [\Gamma_1 \boldsymbol{\Omega}_1 \boldsymbol{\Lambda}_1, \dots, \Gamma_m \boldsymbol{\Omega}_m \boldsymbol{\Lambda}_m]). \quad (28)$$

Substituting (28) into (27) and using (24) completes the proof.

From *Theorem 2*, we can see that the upper bound of ergodic sum capacity depends on the average SNR and the eigenmode channel coupling matrices $\boldsymbol{\Omega}_i$, for $i = 1, 2, \dots, m$. Low-complexity algorithms about the computation of the matrix permanent were developed in [28].

3.3 Optimizing the power allocation policies

We now consider the transmitter power allocation optimization problem. Based on the upper bound in *Theorem 2*, we develop low-complexity power allocation solutions using convex optimization techniques and then propose a simple iterative water-filling algorithm (IWFA) for approaching the optimal power allocation policy.

From (25), the power allocation optimization problem can be formulated as

$$\max_{\lambda_i \geq 0 \forall i} \tilde{C}_u(\boldsymbol{\lambda}) \quad (29)$$

$$\text{s.t. } 1^T \boldsymbol{\lambda}_i = P_i' \forall i. \quad (30)$$

The above problem is a concave optimization problem [28] and the solution can be evaluated by employing standard convex optimization algorithms. In the following, we derive necessary and sufficient conditions for the optimal solution using the Karush-Kuhn-Tucker (KKT) conditions.

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 the optimal power allocation are given by*

$$\lambda_j^{(i)} = \left(\tilde{v}_i - \frac{p(\boldsymbol{\lambda}_{i(j)})}{q(\boldsymbol{\lambda}_{i(j)})} \right)^+, \quad (31)$$

$$1^T \boldsymbol{\lambda}_i = P_i', \text{ for } j = 1, \dots, N_t^{(i)}, \text{ and } i = 1, \dots, m, \quad (32)$$

where

$$p(\boldsymbol{\lambda}_{i(j)}) = \underline{\text{Per}}(\gamma \mathbf{B}_{i(j)}), \quad (33)$$

$$q(\boldsymbol{\lambda}_{i(j)}) = \underline{\text{Per}}(\gamma \mathbf{B}_{i[j]}) - \underline{\text{Per}}(\gamma \mathbf{B}_{i(j)}), \quad (34)$$

$\mathbf{B} = [\Gamma_1 \boldsymbol{\Omega}_1 \boldsymbol{\Lambda}_1, \dots, \Gamma_m \boldsymbol{\Omega}_m \boldsymbol{\Lambda}_m]$, $\mathbf{B}_{i(j)}$ denotes the block matrix obtained by replacing the i th sub-matrix (i.e., $\boldsymbol{\Omega}_i \boldsymbol{\Lambda}_i$) of \mathbf{B} by $\boldsymbol{\Omega}_{i(j)} \text{diag}(\boldsymbol{\lambda}_{i(j)})$, $\mathbf{B}_{i[j]}$ denotes the block matrix obtained by replacing the i th submatrix of \mathbf{B} by $\boldsymbol{\Omega}_i \text{diag}(\boldsymbol{\lambda}_{i[j]})$, $\boldsymbol{\Omega}_{i(j)}$ denotes the sub-matrix of $\boldsymbol{\Omega}_i$ obtained by deleting the j th column, $\boldsymbol{\lambda}_{i(j)}$ denotes the $(N_t^{(i)} - 1) \times 1$ vector obtained by deleting the j th element of $\boldsymbol{\lambda}_i$, and $\boldsymbol{\lambda}_{i[j]}$ denotes the $N_t^{(i)} \times 1$ vector obtained by replacing the j th element of $\boldsymbol{\lambda}_i$ by unity. In addition, $(a)^+ = \max\{0, a\}$ and \tilde{v}_i is chosen to satisfy the power constraints in (32).

Proof: See Appendix 2.

Since the right-hand-side of (31) is independent of $\lambda_j^{(i)}$, we propose a simple IWFA to evaluate the optimal power allocation policy which satisfies (31). Simulation results, to be given in Section 4, will demonstrate that this proposed 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:

Step 1 Initialize $\boldsymbol{\lambda}^0 = 1$, $\tilde{C}_u(\boldsymbol{\lambda}^k) = \log \underline{\text{Per}}(\gamma \mathbf{B}^k)$, and $k = 0$.

Step 2 Calculate $p(\boldsymbol{\lambda}_{i(j)}^k) = \underline{\text{Per}}(\gamma \mathbf{B}_{i(j)}^k)$ and $q(\boldsymbol{\lambda}_{i(j)}^k) = \underline{\text{Per}}(\gamma \mathbf{B}_{i(j)}^k) - \underline{\text{Per}}(\gamma \mathbf{B}_{i(j)}^k)$, for $j = 1, 2, \dots, N_t^{(i)}$, and $i = 1, \dots, m$.

Step 3 Calculate $\lambda_j^{(i)k+1} = \left(\tilde{\nu}_i - \frac{p(\lambda_{i(j)}^k)}{q(\lambda_{i(j)}^k)} \right)^+$, for $j = 1, \dots, N_t^{(i)}$, and $i = 1, \dots, m$, with the power constraints $\sum_{j=1}^{N_t^{(i)}} \lambda_j^{(i)k+1} = P_i'$ for $i = 1, \dots, m$.

Step 4 Calculate $\tilde{C}_u(\boldsymbol{\lambda}^{k+1}) = \log \underline{\text{Per}}(\gamma \mathbf{B}^{k+1})$.

Step 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 $\tilde{C}_u(\boldsymbol{\lambda}^{k+1})$.

Step 6 Set $k := k + 1$ and return to Step 2 until the algorithm converges or the iteration number is equal to a predefined value.

In the above, the superscript k specifies the corresponding variable in the k th iteration so that $\boldsymbol{\lambda}^k$ stands for the value of $\boldsymbol{\lambda}$ in the k th iteration. In Step 1 of the first iteration, $\boldsymbol{\lambda}$ is initialized to 1, i.e., equal-power allocation. Note, however, that $\boldsymbol{\lambda}$ could also be initialized in a different way. For example, it is expected that the channel statistics change smoothly frame by frame, where a more appropriate initialization would be the optimal value of $\boldsymbol{\lambda}$ from the previous frame. In Step 3, the conventional water-filling algorithm is performed with the required variables $p(\lambda_{i(j)})$ and $q(\lambda_{i(j)})$ calculated in Step 2. Following the calculation of $\tilde{C}_u(\boldsymbol{\lambda})$ in Step 4, Step 5 is performed to guarantee convergence [28]. 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 value for a given precision.

3.4 Optimality of beamforming

Here, we investigate the optimality of beamforming (i.e., rank-one transmission) [29-34] in the context of the multi-BS cooperation systems. We derive a necessary and sufficient condition for the optimality of beamforming in the multi-BS cooperation systems. For BS $_i$, we assume that the transmit eigenmodes satisfy the following conditions

$$\tau_1^{(i)} \geq \tau_2^{(i)} \geq \dots \geq \tau_{N_t^{(i)}}^{(i)}, \quad (35)$$

where $\tau_j^{(i)} = \sum_{k=1}^{N_r} \omega_{kj}^{(i)}$ and $\omega_{kj}^{(i)} = [\boldsymbol{\Omega}_i]_{kj}$ for $j = 1, \dots, N_t^{(i)}$ and $i = 1, \dots, m$.

Theorem 4 For multi-BS cooperation systems, the transmit covariance matrices of all the BSs that achieve the sum capacity are of unit-rank (i.e., beamforming is

optimal for all the BSs) if and only if the following inequality is fulfilled:

$$\frac{1 - \mathbb{E} \left\{ \frac{1}{1 + \rho_i \Gamma_i \tilde{\mathbf{h}}_{ij}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{ij}} \right\}}{\text{tr} [\mathbb{E} (\mathbf{A}_i^{-1} \boldsymbol{\Delta}_{ij})] - \rho_i \Gamma_i \mathbb{E} \left\{ \frac{(\tilde{\mathbf{h}}_{ij}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{ij})^2}{1 + \rho_i \Gamma_i \tilde{\mathbf{h}}_{ij}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{ij}} \right\}} > \rho_i \Gamma_i, \quad \text{for } j = 2, \dots, N_t^{(i)} \text{ and } i = 1, \dots, m. \quad (36)$$

where

$\tilde{\mathbf{h}}_{ij}$ and $\tilde{\mathbf{h}}_{ij}$ are the first and j th columns of $\tilde{\mathbf{H}}_i$, respectively, with $\tilde{\mathbf{H}}_i$ defined as in (15), and \mathbf{d}_{ij} and \mathbf{m}_{ij} are the j th columns of \mathbf{D}_i and \mathbf{M}_i , respectively.

Proof: See Appendix 3.

Note that the proof is nontrivial generalization of the techniques in [33] to the jointly correlated MIMO multi-BS cooperation systems. We can make the following observations.

- If $\tilde{\mathbf{h}}_{ij}$ for $j = 1, \dots, N_t^{(i)}$ are i.i.d., the left-hand-side of (36) remains unchanged when j varies from 2 to $N_t^{(i)}$, and the right-hand-side of (36) is maximized for $i = 2$. It means that if the condition for $i = 2$ holds, then it is also held for all other i . Thus, inserting $j = 2$ into (36) gives the following condition:

$$\frac{1 - \mathbb{E} \left\{ \frac{1}{1 + \rho_i \Gamma_i \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1}} \right\}}{\text{tr} [\mathbb{E} (\mathbf{A}_i^{-1} \boldsymbol{\Delta}_{i2})] - \rho_i \Gamma_i \mathbb{E} \left\{ \frac{(\tilde{\mathbf{h}}_{i2}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i2})^2}{1 + \rho_i \Gamma_i \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1}} \right\}} > \rho_i \Gamma_i, \quad \text{for } j = 1, \dots, m. \quad (37)$$

- If the LOS $\mathbf{D} = \mathbf{0}$ and $\mathbf{m}_{ij} = [a_1 b_{ij}, \dots, a_{N_r} b_{ij}]^T$, where a_l and b_{ij} are square-roots of the eigenvalues of the receive correlation matrix $\boldsymbol{\Phi}_r$ and transmit correlation matrix $\boldsymbol{\Phi}_{t,i}$, respectively. In this case, the channel degenerates to the Kronecker channel. It can be proved that (37) reduces to the beamforming optimality condition in [33, Theorem 2].

4 Numerical results

In this section, we present numerical results to evaluate the tightness of the capacity bound, and demonstrate the efficiency and performance of the proposed transmitter optimization approach. We consider downlink transmission for coordinated cellular networks with two BSs and single user. Assumption that all cases have the same total transmit power of BSs, i.e., $P_1 = P_2$. For the jointly correlated channel, we set the LOS $\mathbf{D} = \mathbf{0}$. The spatial channel model (SCM) [43,44] is used to generate channel matrices of two independent links, i.e., $\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2$. The simulation environment is set to be urban micro-cell, with the BS antenna spacing $d_{\text{BS}} = 0.5\lambda$, and the

user antenna spacing $d_{UE} = 0.5\lambda$. The number of distinguishable paths is set to be 1, i.e., flat fading. For each link, 1,000 time samples are generated for the calculation of the statistical CSI. The path loss model is $31.5 + 35 \log_{10}(d)$ ($d(m)$) and site to site distance is set to 1,000 m. In all of the following figures, the horizontal axis (SNR) indicates the received SNR.

Figure 1 illustrates the results of the ergodic sum capacity of the joint IWFA and the individual IWFA (where the BSs cannot cooperate) with $N_t^{(1)} = 4, N_t^{(2)} = 4, N_r = 4$. As we can see, the capacity of joint IWFA is greater than that of individual IWFA. For comparison, the results for the exact ergodic sum capacity are also shown, which were obtained by numerically evaluating (13) using a constrained optimization. In addition, Equal-power allocation and beamforming are optimal in the high and low SNR regimes, respectively.

Figure 2 demonstrates the convergence of the proposed IWFA for optimal power allocation. In this figure,

the SNR ρ is set to 20 dB, and the algorithm is initialized using $\lambda^0 = 1$. From these results, we see that the proposed IWFA converges after only a few iterations.

Figure 3 compares the ergodic sum capacity upper bound in (25) with the Monte-Carlo simulation results of the ergodic mutual information by averaging over a large number of independent realizations of \mathbf{H} with $\{N_t^{(1)} = 2, N_t^{(2)} = 2, N_r = 2\}$, $\{N_t^{(1)} = 2, N_t^{(2)} = 2, N_r = 4\}$ and $\{N_t^{(1)} = 4, N_t^{(2)} = 4, N_r = 8\}$. It shows that the ergodic sum capacity upper bound is tight. Employing this result of the ergodic sum capacity upper bound, the proposed IWFA for optimal power allocation can be achieved the true channel capacity.

To investigate the beamforming optimality condition, the inequality (36) can be rewritten as

$$\frac{\left(1 - \mathbb{E} \left[\frac{1}{1 + \rho \Gamma_t^{(j)} t_1^{(j)} + b_{t_1}^{(j)} A_t^{(j)} b_{t_1}^{(j)}} \right] \right)^{t_1^{(j)}}}{\text{tr}[\mathbb{E}(A_t^{-1} \mathbf{A}_t)] - \rho \Gamma_t^{(j)} t_1^{(j)} + \mathbb{E} \left[\frac{(b_{t_1}^{(j)} A_t^{(j)} b_{t_1}^{(j)})^2}{1 + \rho \Gamma_t^{(j)} t_1^{(j)} + b_{t_1}^{(j)} A_t^{(j)} b_{t_1}^{(j)}} \right]} > \rho \Gamma_t^{(j)}, \text{ for } j = 2, \dots, N_t^{(j)} \text{ and } \forall t. \quad (38)$$

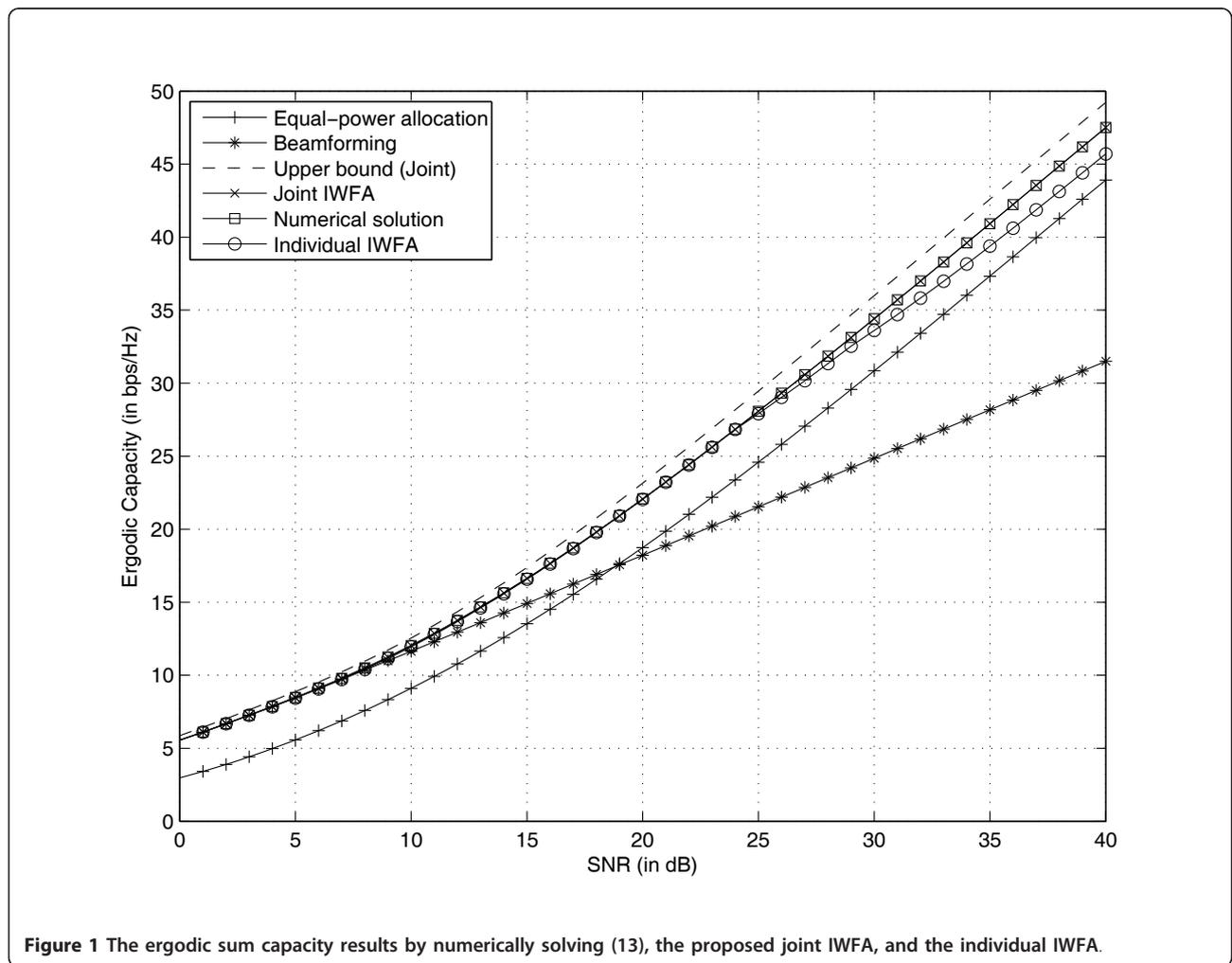


Figure 1 The ergodic sum capacity results by numerically solving (13), the proposed joint IWFA, and the individual IWFA.

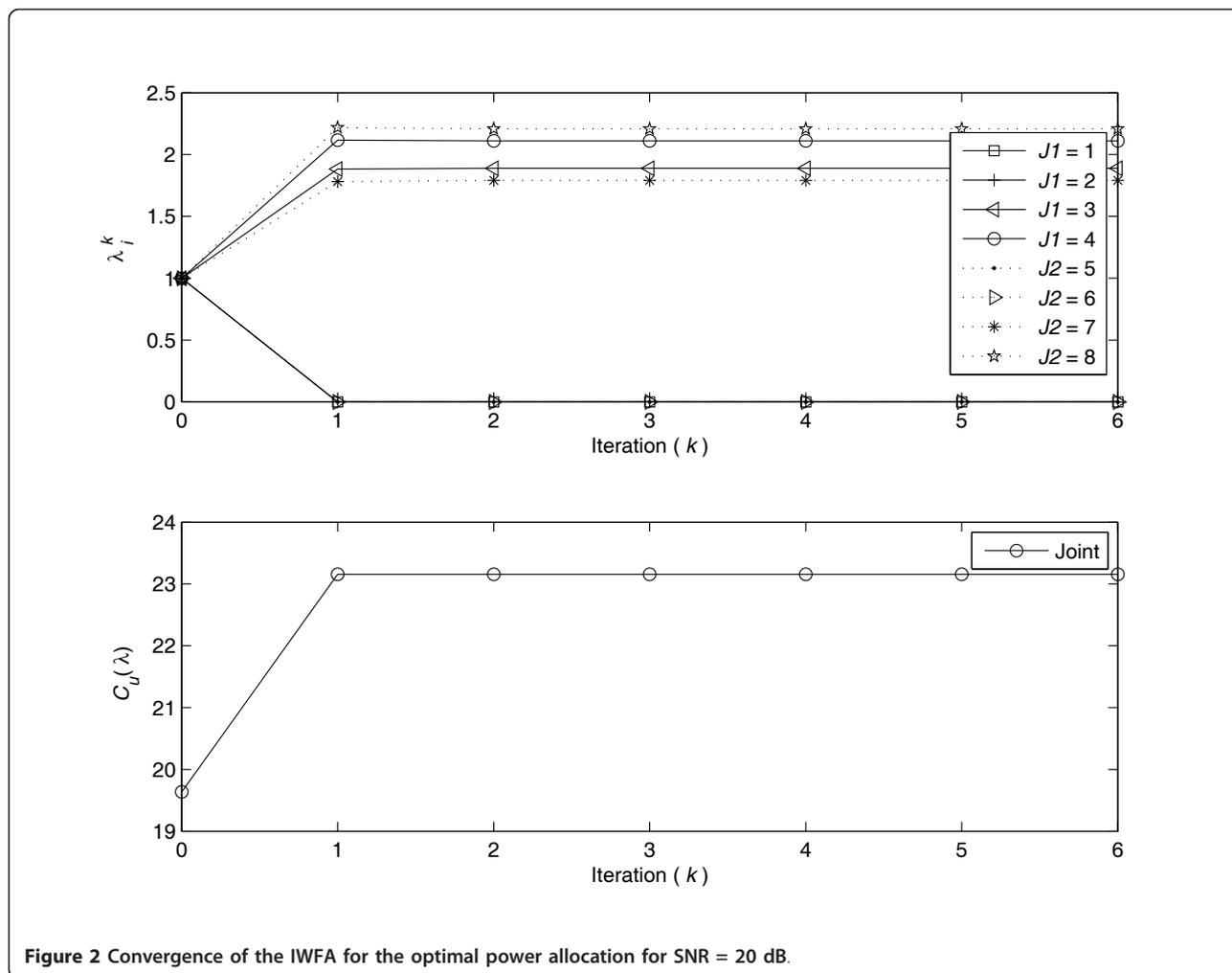


Figure 2 Convergence of the IWFA for the optimal power allocation for SNR = 20 dB.

We set the LOS $\mathbf{D} = \mathbf{0}$ and $\Gamma_1 = \Gamma_2 = 1$, the number of transmit antennas $N_t^{(i)} = 2$ and the number of receive antennas $N_r = 4$, with the same SCM as before. Figure 4 shows the beamforming optimality condition (38) for different numbers of BSs. The condition is plotted as a function of $\rho_1 \tau_1^{(1)}$ and $\rho_1 \tau_2^{(1)}$. We see that the region where beamforming is optimal gets larger with increasing number of BSs. Note that these curves lie below the $\tau_1^{(1)} = \tau_2^{(1)}$ line because $\tau_1^{(1)}$ is the largest transmit eigenmode. It can be seen that as m increases, the curves get closer to the $\tau_1^{(1)} = \tau_2^{(1)}$ line.

Figure 5 illustrates the beamforming optimality condition (38) for different antenna spacing. We set the number of BSs $m = 5$. The condition is plotted as a function of $\rho_1 \tau_1^{(1)}$ and $\rho_1 \tau_2^{(1)}$. We see that the region where beamforming is optimal gets larger as the user antenna spacing decreases and remains almost unchanged as the BS spacing changes.

5 Conclusion

We optimized the transmission for jointly correlated MIMO channels using statistical CSI over cooperative cellular networks with multiple BSs and a single cell-edge receiver user. We proposed an optimal transmit scheme to maximize the ergodic capacity and showed that the transmitted signals of all the BSs should be mutually independent and the optimal transmitted directions are the eigenvectors of the BS's own transmit correlation matrix. We also derived a closed-form tight upper bound for the ergodic capacity, based on which we developed low-complexity power allocation solutions using convex optimization techniques and a simple IWFA. Finally, we derived the beamforming optimality conditions for all the BSs.

Appendix 1: The definitions and properties of matrix permanents

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

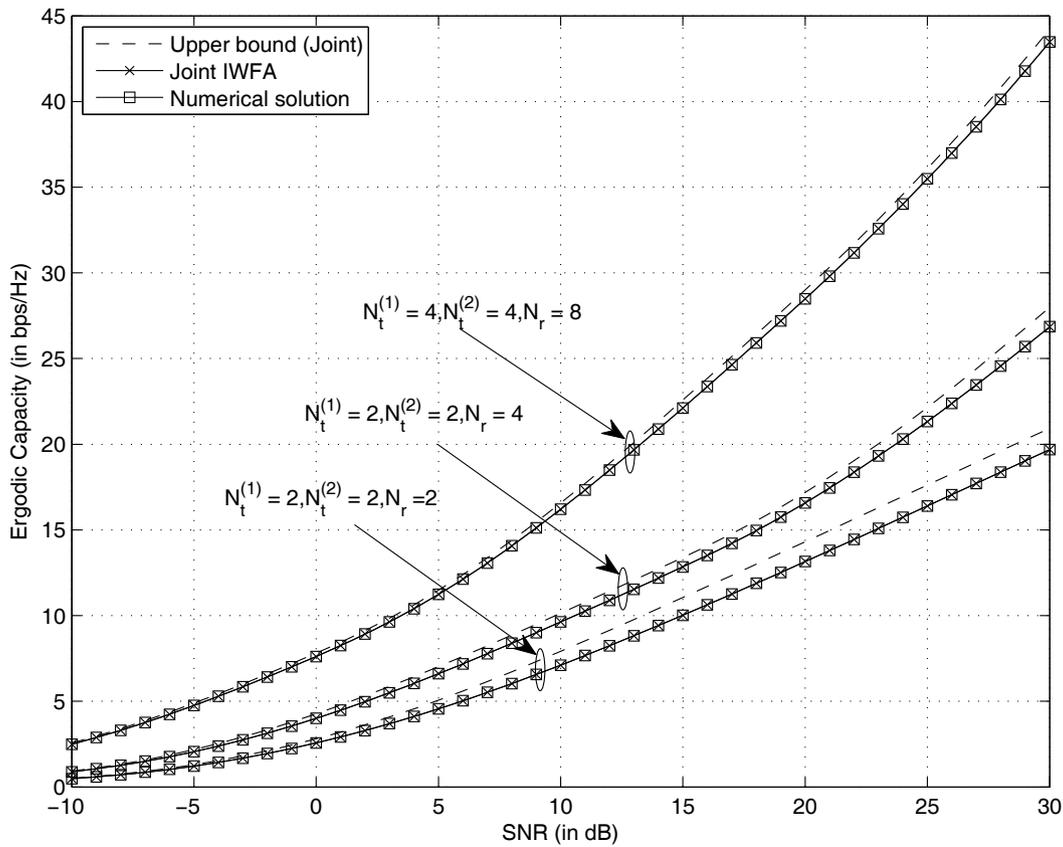


Figure 3 Comparison of the ergodic sum capacity upper bound in (25) with the exact ergodic mutual information ($\{N_t^{(1)} = 2, N_t^{(2)} = 2, N_r = 2\}$, $\{N_t^{(1)} = 2, N_t^{(2)} = 2, N_r = 4\}$, and $\{N_t^{(1)} = 4, N_t^{(2)} = 4, N_r = 8\}$).

$$\text{Per}(\mathbf{A}) = \begin{cases} \sum_{\hat{\alpha}_M \in \mathcal{S}_N^M} \prod_{i=1}^M a_{i, \alpha_i}, & \text{if } M \leq N, \\ \sum_{\hat{\beta}_N \in \mathcal{S}_M^N} \prod_{i=1}^N a_{\beta_i, i}, & \text{if } M > N, \end{cases} \quad (39)$$

where $a_{i,j}$ denotes the (i,j) th element of \mathbf{A} ; \mathcal{S}_N^k denotes the set of all size- k permutations of the numbers $\{1, 2, \dots, N\}$, for $k \leq N$; $\hat{\alpha}_k = (\alpha_1, \alpha_2, \dots, \alpha_k)$, $\alpha_i \in \{1, 2, \dots, N\}$ for $1 \leq i \leq k$, and $\alpha_i \neq \alpha_j$ for $1 \leq i, j \leq k$ and $i \neq j$.

Definition 2 The extended permanent of \mathbf{A} is defined as

$$\underline{\text{Per}}(\mathbf{A}) = \text{Per}(\mathbf{I}_M \mathbf{A}) = \text{Per}(\mathbf{I}_N \mathbf{A}^T). \quad (40)$$

According to the above definition, one can easily establish a number of important properties of the matrix permanent, as given in the following lemma [28].

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, we have

$$\text{Per}(\mathbf{A}) = \text{Per}(\mathbf{A}^T) \quad (41)$$

$$\text{Per}(\mathbf{a}) = \sum_{i=1}^M a_i \quad (42)$$

$$\text{Per}(\text{diag}(\mathbf{a})) = \det(\text{diag}(\mathbf{a})) \quad (43)$$

$$\text{Per}(\mu \mathbf{A}) = \mu^{\min(M,N)} \text{Per}(\mathbf{A}) \quad (44)$$

$$\text{Per}(\text{diag}(\mathbf{a}) \mathbf{A}) = \det(\text{diag}(\mathbf{a})) \text{Per}(\mathbf{A}), \quad \text{for } M \leq N \quad (45)$$

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

$$\text{Per}(\mathbf{A}) = \begin{cases} \sum_{\hat{\sigma}_k \in \mathcal{S}_N^{(k)}} \text{Per}(\mathbf{A}_{\hat{\sigma}_k}^{\hat{\alpha}_k}) \text{Per}(\mathbf{A}_{\hat{\sigma}_k}^{\hat{\alpha}'_k}), & \text{if } M \leq N \\ \sum_{\hat{\sigma}_k \in \mathcal{S}_N^{(k)}} \text{Per}(\mathbf{A}_{\hat{\beta}_k}^{\hat{\sigma}_k}) \text{Per}(\mathbf{A}_{\hat{\beta}_k}^{\hat{\beta}'_k}), & \text{if } M > N, \end{cases} \quad (47)$$

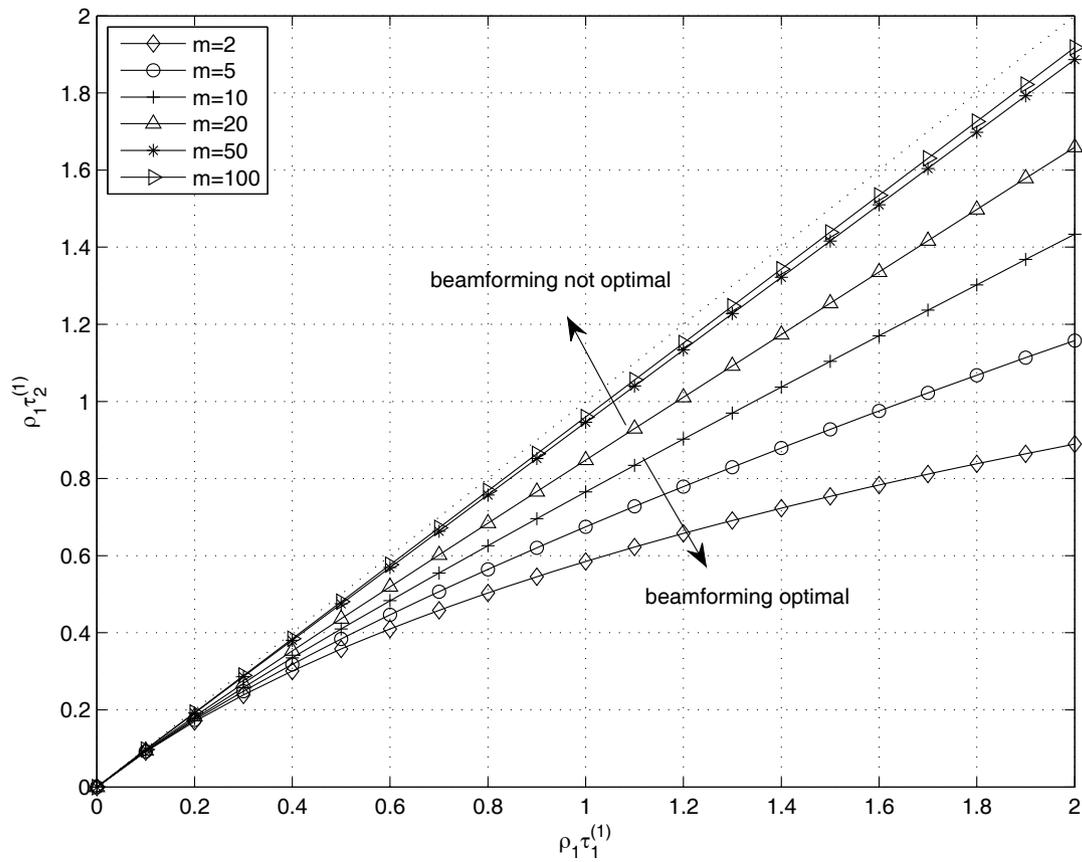


Figure 4 The optimality of beamforming for different numbers of BSs for systems with $N_r = 4, N_t^{(i)} = 2$ for $i = 1, \dots, m$.

where $\mathcal{S}_N^{(k)}$ denotes the set of all ordered length- k subsets of the numbers $\{1, 2, \dots, N\}$; $\hat{\alpha}_k = (\alpha_1, \alpha_2, \dots, \alpha_k)$, $\alpha_i \in \{1, 2, \dots, N\}$ for $1 \leq i \leq k$, and $\alpha_1 < \alpha_2 < \dots < \alpha_k$; $\hat{\beta}_k \in \mathcal{S}_N^{(k)}$ and $\hat{\beta}'_k \in \mathcal{S}_N^{(k)}$ with $1 \leq k \leq \min(M, N)$; $\mathbf{A}_{\hat{\beta}_k}^{\hat{\alpha}_k}$ is the sub-matrix of \mathbf{A} obtained by selecting the rows and columns indexed by $\hat{\alpha}_k$ and $\hat{\beta}_k$ respectively; $\hat{\alpha}'_k$ and $\hat{\beta}'_k$ denote the sequences complementary to $\hat{\alpha}_k$ and $\hat{\beta}_k$ in $\{1, 2, \dots, M\}$ and $\{1, 2, \dots, N\}$, respectively; $\mathbf{A}_{\hat{\beta}'_k}^{\hat{\alpha}'_k}$ is the sub-matrix of \mathbf{A} obtained by deleting the rows and columns indexed by $\hat{\alpha}_k$ and $\hat{\beta}_k$, respectively.

Appendix 2: Proof of Theorem 3

Let $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, \dots, \boldsymbol{\mu}_m^T]^T$ and $\mathbf{v} = [v_1, v_2, \dots, v_m]^T$ be the Lagrange multipliers for the inequality constraint $\boldsymbol{\lambda} \geq 0$ and the equality constraint $1^T \boldsymbol{\lambda}_i = P'_i$ for $i = 1, 2, \dots, m$, respectively, where $\boldsymbol{\mu}_i^T = [\mu_1^{(i)}, \mu_2^{(i)}, \dots, \mu_{N_t^{(i)}}^{(i)}]^T$, for $i = 1,$

$2, \dots, m$. Then, the KKT conditions satisfied by the optimal $\boldsymbol{\lambda}$ can be found from solving

$$\frac{\partial \tilde{C}_u(\boldsymbol{\lambda})}{\partial \lambda_j^{(i)}} + \mu_j^{(i)} + v_i = 0, \quad (48)$$

$$\lambda_i \geq 0, 1^T \boldsymbol{\lambda}_i = P'_i, \mu_i \geq 0, \mu_j^{(i)} \lambda_j^{(i)} = 0, \quad (49)$$

where $\frac{\partial \tilde{C}_u(\boldsymbol{\lambda})}{\partial \lambda_j^{(i)}}$ denotes the partial derivative of $\tilde{C}_u(\boldsymbol{\lambda})$ with respect to $\lambda_j^{(i)}$, for $1 \leq j \leq N_t^{(i)}$ and $1 \leq i \leq m$. By using the derivative chain rule, we then have

$$\frac{\partial \tilde{C}_u(\boldsymbol{\lambda})}{\partial \lambda_j^{(i)}} = \frac{1}{E(\boldsymbol{\lambda})} \frac{\partial E(\boldsymbol{\lambda})}{\partial \lambda_j^{(i)}}. \quad (50)$$

Now, letting $\mathbf{B} = [\Gamma_1 \boldsymbol{\Omega}_1 \boldsymbol{\Lambda}_1, \dots, \Gamma_m \boldsymbol{\Omega}_m \boldsymbol{\Lambda}_m]$, we can write $E(\boldsymbol{\lambda}) = \text{Per}(\boldsymbol{\gamma} \mathbf{B})$. To evaluate the remaining derivatives in (50), we apply the Laplace expansion property of

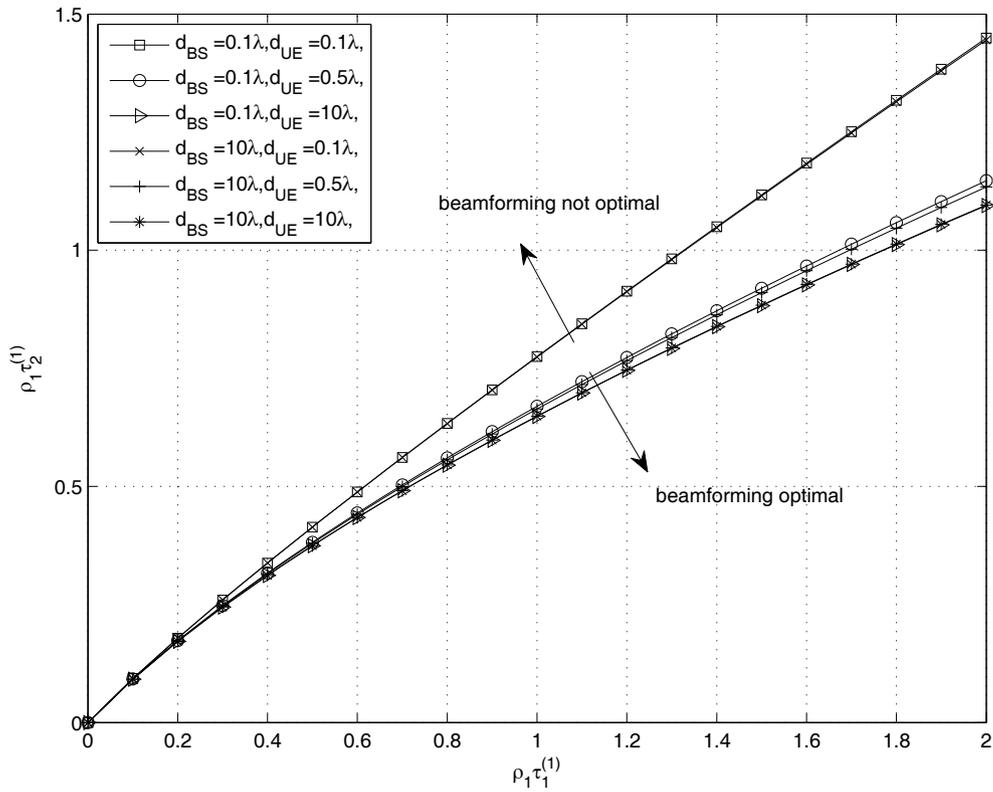


Figure 5 The optimality of beamforming for different antenna spacing for systems with $m = 5$, $N_r = 4$, $N_t^{(i)} = 2$ for $i = 1, \dots, 5$.

permanents, given by *Lemma 2* in Appendix 1 when $k = 1$, to express $E(\lambda)$ as

$$E(\lambda) = p(\lambda_{i(j)}) + \lambda_i q(\lambda_{i(j)}), \quad (51)$$

where $p(\lambda_{i(j)})$ and $q(\lambda_{i(j)})$ are given by (33) and (34), respectively. Thus, (50) becomes

$$\frac{\partial \tilde{C}_u(\lambda)}{\partial \lambda_j^{(i)}} = \frac{q(\lambda_{i(j)})}{p(\lambda_{i(j)}) + \lambda_i q(\lambda_{i(j)})}. \quad (52)$$

Substituting (52) into (48) and eliminating the slack variable μ , the KKT conditions become (31) and (32), where $(a)^+ = \max\{0, a\}$ and $\tilde{v}_i = \frac{1}{v_i}$.

Appendix 3: Proof of Theorem 4

It is obvious that (13) is equivalent to the following convex optimization problem:

$$\min -\mathbb{E} \left\{ \log_2 \left| \mathbf{I}_{N_r} + \gamma \sum_{i=1}^m \Gamma_i \sum_{j=1}^{N_t^{(i)}} \lambda_j^{(i)} \tilde{\mathbf{h}}_{ij} \tilde{\mathbf{h}}_{ij}^H \right| \right\}, \quad (53)$$

$$\text{s.t. } \sum_{j=1}^{N_t^{(i)}} \lambda_j^{(i)} = P'_i, \quad \text{for } i = 1, \dots, m, \quad (54)$$

with $\tilde{\mathbf{h}}_{ij}$ denoting the j -th column of $\tilde{\mathbf{H}}_i$. The Lagrangian of the problem is given by

$$L = -\mathbb{E} \left\{ \log_2 \left| \mathbf{I}_{N_r} + \gamma \sum_{i=1}^m \Gamma_i \sum_{j=1}^{N_t^{(i)}} \lambda_j^{(i)} \tilde{\mathbf{h}}_{ij} \tilde{\mathbf{h}}_{ij}^H \right| \right\} + \sum_{i=1}^m \mu_i \left(\sum_{j=1}^{N_t^{(i)}} \lambda_j^{(i)} - P'_i \right), \quad (55)$$

with μ_i denoting the Lagrange multiplier corresponding to the transmit power constraint of BS $_i$. Since the problem is convex, we can derive necessary and sufficient conditions for the optimal solution using the KKT conditions. The KKT conditions for BS i can be found as

$$\gamma \Gamma_i \mathbb{E} \left\{ \tilde{\mathbf{h}}_{ij}^H \left(\mathbf{I}_{N_r} + \gamma \sum_{i=1}^m \Gamma_i \sum_{j=1}^{N_t^{(i)}} \lambda_j^{(i)} \tilde{\mathbf{h}}_{ij} \tilde{\mathbf{h}}_{ij}^H \right)^{-1} \tilde{\mathbf{h}}_{ij} \right\} \leq \mu_i, \quad \text{for } j = 1, \dots, N_t^{(i)}, \quad (56)$$

where the conditions are met with equality only if, for a given j , the corresponding eigenvalue $\lambda_j^{(i)}$ of the input covariance matrix is non-zero. Thus, from (35),

beamforming is optimal if (56) is satisfied as an equality for $j = 1$, and as a strict inequality for all other j . This means that beamforming of BS_i along the strongest transmit eigenmode is optimal. That is,

$$E_{i1} = \gamma \Gamma_i \mathbb{E} \left\{ \tilde{\mathbf{h}}_{i1}^H \mathbf{A}^{-1} \tilde{\mathbf{h}}_{i1} \right\} = \mu_i, \quad (57)$$

$$E_{ij} = \gamma \Gamma_i \mathbb{E} \left\{ \tilde{\mathbf{h}}_{ij}^H \mathbf{A}^{-1} \tilde{\mathbf{h}}_{ij} \right\} < \mu_i, \quad \text{for } j \neq 1. \quad (58)$$

where

$$\mathbf{A} = \mathbf{I}_{N_r} + \gamma \sum_{i=1}^m \Gamma_i P_i' \tilde{\mathbf{h}}_{i1} \tilde{\mathbf{h}}_{i1}^H \quad (59)$$

Equivalently, the beamforming optimality conditions for BS_i can be written as

$$\frac{E_{i1}}{E_{ij}} > 1, \quad \text{for } j = 2, \dots, N_t^{(i)}. \quad (60)$$

We now use the matrix inversion formula [45] to give

$$\mathbf{A}^{-1} = \mathbf{A}_i^{-1} - \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1} \left(\frac{1}{\rho_i \Gamma_i} + \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1} \right)^{-1} \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \quad (61)$$

and as a result, we get

$$\begin{aligned} \tilde{\mathbf{h}}_{i1}^H \mathbf{A}^{-1} \tilde{\mathbf{h}}_{i1} &= \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1} - \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1} \left(\frac{1}{\rho_i \Gamma_i} + \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1} \right)^{-1} \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1} \\ &= \frac{1}{\rho_i \Gamma_i} \left(1 - \frac{1}{1 + \rho_i \Gamma_i \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1}} \right), \end{aligned} \quad (62)$$

where

$$\mathbf{A}_i = \mathbf{A} - \Gamma_i P_i' \tilde{\mathbf{h}}_{i1} \tilde{\mathbf{h}}_{i1}^H, \quad (63)$$

Therefore, (57) and (58) can be rewritten as

$$E_{i1} = \frac{1}{P_i'} - \frac{1}{P_i'} \mathbb{E} \left\{ \frac{1}{1 + \rho_i \Gamma_i \tilde{\mathbf{h}}_{i1}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{i1}} \right\} \quad (64)$$

and

$$E_{ij} = \gamma \Gamma_i \text{tr} \left[\mathbb{E} \left\{ \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{ij} \tilde{\mathbf{h}}_{ij}^H \right\} \right] - \gamma \Gamma_i \mathbb{E} \left\{ \left(\tilde{\mathbf{h}}_{ij}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{ij} \right)^2 \left(\frac{1}{\rho_i \Gamma_i} + \tilde{\mathbf{h}}_{ij}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{ij} \right)^{-1} \right\}, \quad \text{for } j \neq 1. \quad (65)$$

Note that $\tilde{\mathbf{h}}_{ij}$ for $j \neq 1$ and \mathbf{A}_i are independent, and $\tilde{\mathbf{h}}_{ij} = \mathbf{d}_{ij} + \mathbf{m}_{ij} \odot \mathbf{h}_{\text{id},ij}$ where \mathbf{d}_{ij} , \mathbf{m}_{ij} and $\mathbf{h}_{\text{id},ij}$ are the j th column of \mathbf{D}_i , \mathbf{M}_i and $\mathbf{H}_{\text{id},i}$ respectively. Then,

$$\mathbb{E} \left\{ \tilde{\mathbf{h}}_{ij} \tilde{\mathbf{h}}_{ij}^H \right\} = \mathbf{d}_{ij} \mathbf{d}_{ij}^H + \mathbf{m}_{ij} \mathbf{m}_{ij}^H \odot \mathbf{I}_{N_r}, \quad \text{for } j = 2, \dots, N_t^{(i)} \quad (66)$$

and (65) can be simplified as

$$E_{ij} = \gamma \Gamma_i \text{tr} \left[\mathbb{E} \left\{ \mathbf{A}_i^{-1} \right\} \left(\mathbf{d}_{ij} \mathbf{d}_{ij}^H + \mathbf{m}_{ij} \mathbf{m}_{ij}^H \odot \mathbf{I}_{N_r} \right) \right] - \gamma \rho_i \Gamma_i^2 \mathbb{E} \left\{ \frac{\left(\tilde{\mathbf{h}}_{ij}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{ij} \right)^2}{1 + \rho_i \Gamma_i \tilde{\mathbf{h}}_{ij}^H \mathbf{A}_i^{-1} \tilde{\mathbf{h}}_{ij}} \right\}, \quad \text{for } j \neq 1. \quad (67)$$

Combining (60), (64), and (67), we obtain the desired condition in (36).

Acknowledgements

The authors would like to thank the reviewers for their valuable comments and suggestions. This study was supported by the National Natural Science Foundation of China under Grants 60925004, 60902009 and 61101089, and the Supporting Program for New Century Excellent Talents in University.

Author details

¹National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China ²Department of Electronic and Electrical Engineering, University College London, London WC1E 7JE, UK

Competing interests

The authors declare that they have no competing interests.

Received: 25 October 2011 Accepted: 12 April 2012

Published: 12 April 2012

References

1. IE Telatar, Capacity of multi-antenna Gaussian channels. *Eur Trans Telecommun.* **10**(6), 585–595 (1999). doi:10.1002/ett.4460100604
2. S Shamai (Shitz), BM Zaidel, Enhancing the cellular downlink capacity via co-processing at the transmission end, in *Proc IEEE Veh Tech Conf Spring*, vol. 3. Rhodes Island, Greece, pp. 1745–1749 (2001)
3. S Shamai (Shitz), O Somekh, BM Zaidel, Multi-cell communications: An information theoretic perspective, in *Proc Joint Workshop Commun and Coding*, Donnini, Florence, Italy, pp. 14–17 (2004)
4. O Somekh, BM Zaidel, S Shamai (Shitz), Sum rate characterization of joint multiple cell-site processing. *IEEE Trans Inf Theory.* **53**(12), 4473–4497 (2007)
5. O Somekh, O Simeone, Y Bar-Ness, AM Haimovich, S Shamai (Shitz), Cooperative multicell zero-forcing beamforming in cellular downlink channels. *IEEE Trans Inf Theory.* **2009**(7), 3206–3219 (2009)
6. H Zhang, H Dai, Co-channel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks. *EURASIP J Wirel Commun Netw.* **2**, 222–235 (2004)
7. MK Karakayali, GJ Foschini, RA Valenzuela, RD Yates, On the maximum common rate achievable in a coordinated network, in *Proc IEEE Int Conf Commun*, vol. 9. Istanbul, Turkey, pp. 4333–4338 (2006)
8. S Jing, DNC Tse, J Hou, JB Soriaga, JE Smee, R Padovani, Multi-cell downlink capacity with coordinated processing. *EURASIP J Wirel Commun Netw.* **2008**, 1–19 (2008)
9. S Venkatesan, A Lozano, R Valenzuela, Network MIMO: overcoming intercell interference in indoor wireless systems, in *Proc Asilomar Conf Sig Sys and Comput*, Pacific Grove, CA, USA, pp. 83–87 (2007)
10. H Huang, M Trivellato, A Hottinen, M Shafi, P Smith, R Valenzuela, Increasing downlink cellular throughput with limited network MIMO coordination. *IEEE Trans Wirel Commun.* **8**(6), 2983–2989 (2009)
11. G Caire, S Shamai, On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Trans Inf Theory.* **49**(7), 1691–706 (2003). doi:10.1109/TIT.2003.813523
12. P Viswanath, D Tse, Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality. *IEEE Trans Inf Theory.* **49**(8), 1912–1921 (2003). doi:10.1109/TIT.2003.814483
13. W Yu, J Cioffi, Sum capacity of Gaussian vector broadcast channels. *IEEE Trans Inf Theory.* **50**(9), 1875–1892 (2004). doi:10.1109/TIT.2004.833336
14. S Vishwanath, N Jindal, A Goldsmith, Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels. *IEEE Trans Inf Theory.* **49**(10), 2658–2668 (2003). doi:10.1109/TIT.2003.817421
15. W Yu, T Lan, Transmitter optimization for the multi-antenna downlink with per-antenna power constraints. *IEEE Trans Signal Process.* **55**(6), 2646–2660 (2007)
16. L Zhang, R Zhang, YC Liang, Y Xin, HV Poor, On the Gaussian MIMO BC-MAC duality with multiple transmit covariance constraints. *IEEE Trans Inf Theory.* **58**(4), 2064–2078 (2012)
17. H Huh, H Papadopoulos, G Caire, MIMO broadcast channel optimization under general linear constraints, in *Proc IEEE Int Sym Info Theory*, Seoul, Korea, pp. 2664–2668 (2009)

18. SA Ramprashad, G Caire, Cellular vs. network MIMO: a comparison including channel state information overhead, in *Proc IEEE Int Sym Personal, Indoor Mobile Radio Commun*, Tokyo, Japan, pp. 878–884 (2009)
19. SA Ramprashad, G Caire, HC Papadopoulos, Cellular and network MIMO architectures: MU-MIMO spectral efficiency and costs of channel state information, in *Proc Asilomar Conf on Sig Sys and Comp*, Pacific Grove, CA, USA, pp. 1811–1818 (2009)
20. G Caire, SA Ramprashad, HC Papadopoulos, Rethinking network MIMO: cost of CSIT, performance analysis, and architecture comparisons, in *Proc Info Theory and App Workshop*, San Diego, USA, pp. 1–10 (2010)
21. G Barriac, U Madhow, Space-time communication for OFDM with implicit channel feedback. *IEEE Trans Inf Theory*. **50**(12), 3111–3129 (2004). doi:10.1109/TIT.2004.838379
22. G Barriac, U Madhow, Space-time precoding for mean and covariance feedback: application to wideband OFDM. *IEEE Trans Commun*. **54**, 96–107 (2006)
23. DS Shui, GJ Foschini, MJ Gans, JM Kahn, Fading correlation and its effect on the capacity of multielement antenna systems. *IEEE Trans Commun*. **48**(3), 502–513 (2000). doi:10.1109/26.837052
24. H Shin, JH Lee, Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole. *IEEE Trans Inf Theory*. **49**(10), 2636–2647 (2003). doi:10.1109/TIT.2003.817439
25. AM Tulino, A Lozano, S Verdú, Impact of antenna correlation on the capacity of multiantenna channels. *IEEE Trans Inf Theory*. **51**(7), 2491–2509 (2005). doi:10.1109/TIT.2005.850094
26. W Weichselberger, M Herdin, H Ozelik, E Bonek, A stochastic MIMO channel model with joint correlation of both link ends. *IEEE Trans Wirel Commun*. **5**, 90–100 (2006)
27. Y Zhou, M Herdin, A Sayeed, E Bonek, Experimental study of MIMO channel statistics and capacity via the virtual channel representation. UW Tech Report http://dune.ece.wisc.edu/pdfs/zhou_meas.pdf (2007)
28. XQ Gao, B Jiang, X Li, AB Gershman, MR McKay, Statistical eigenmode transmission over jointly correlated MIMO channels. *IEEE Trans Inf Theory*. **55**(8), 3735–3750 (2009)
29. SA Jafar, A Goldsmith, Transmitter optimization and optimality of beamforming for multiple antenna systems. *IEEE Trans Wirel Commun*. **3**(4), 1165–1175 (2004). doi:10.1109/TWC.2004.830822
30. E Jorswieck, H Boche, Channel capacity and capacity range of beamforming in MIMO wireless systems under correlated fading with covariance feedback. *IEEE Trans Wirel Commun*. **3**(5), 1543–1553 (2004). doi:10.1109/TWC.2004.833523
31. X Li, S Jin, XQ Gao, MR Mckay, KK Wong, Transmitter optimization and beamforming optimality conditions for doublescattering MIMO channels. *IEEE Trans Wirel Commun*. **7**(9), 3647–3654 (2008)
32. H Wan, RR Chen, YB Liang, Optimality of beamforming for MIMO multiple access channels via virtual representation. *IEEE Trans Signal Process*. **58**(10), 5458–5463 (2010)
33. A Soysal, S Ullukus, Optimality of beamforming in fading MIMO multiple access channels. *IEEE Trans Commun*. **57**(4), 1171–1183 (2009)
34. X Li, S Jin, XQ Gao, MR Mckay, Capacity bounds and low complexity transceiver design for double-scattering MIMO multiple access channels. *IEEE Trans Signal Process*. **58**(5), 2809–2822 (2010)
35. CK Wenm, KK Wong, Asymptotic analysis of spatially correlated MIMO multiple-access channels with arbitrary signaling inputs for joint and separate decoding. *IEEE Trans Inf Theory*. **53**, 252–268 (2007)
36. CK Wen, KK Wong, JC Chen, Spatially correlated MIMO multiple-access systems with macro-diversity: asymptotic analysis via statistical physics. *IEEE Trans Commun*. **55**(3), 477–488 (2007)
37. AM Tulino, A Lozano, S Verdú, Capacity-achieving input covariance for single-user multi-antenna channels. *IEEE Trans Wirel Commun*. **5**(3), 662–671 (2006)
38. SK Jayaweera, HV Poor, On the capacity of multi-antenna systems in the presence of Rician fading, in *Proc IEEE Veh Tech Conf*, Vancouver, BC, Canada, pp. 1963–1967 (2002)
39. SK Jayaweera, HV Poor, On the capacity of multiple antenna systems in Rician fading. *IEEE Trans Wirel Commun*. **4**(3), 1102–1111 (2005)
40. E Visotsky, U Madhow, Space-time transmit precoding with imperfect feedback. *IEEE Trans Inf Theory*. **47**(6), 2632–2639 (2001). doi:10.1109/18.945281
41. V Veeravalli, Y Liang, AM Sayeed, Correlated MIMO Rayleigh fading channels: capacity, optimal signaling and asymptotics. *IEEE Trans Inf Theory*. **51**(6), 2058–2072 (2005). doi:10.1109/TIT.2005.847724
42. H Minc, *Permanents* (Addison-Wesley Publishing Company, Mass, 1978)
43. Spatial channel model for multiple input multiple output (MIMO) simulations. 3GPP, TR 25.996, V7.0.0 (2007)
44. J Salo, G Del Galdo, J Salmi, P Kyosti, M Milojevic, D Laselva, C Schneider, MATLAB implementation of the 3GPP spatial channel model (3GPP TR 25.996). <http://www.tkk.fi/Units/Radio/scm/> (2005)
45. RA Horn, CR Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985)

doi:10.1186/1687-6180-2012-81

Cite this article as: Zhang et al.: On base station cooperation using statistical CSI in jointly correlated MIMO downlink channels. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:81.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
