Cooperative Multicell Precoding: Rate Region Characterization and Distributed Strategies with Instantaneous and Statistical CSI

Emil Björnson, *Student Member, IEEE,* Randa Zakhour, *Student Member, IEEE,* David Gesbert, *Senior Member, IEEE,* and Björn Ottersten, *Fellow, IEEE*

Abstract

Base station cooperation is an attractive way of increasing the spectral efficiency in multiantenna communication. By serving each terminal through several base stations in a given area, inter-cell interference can be coordinated and higher performance achieved, especially for terminals at cell edges. Most previous work in the area has assumed that base stations have *common* knowledge of both data dedicated to all terminals and full or partial channel state information (CSI) of all links. Herein, we analyze the case of *distributed* cooperation where each base station has only *local* CSI, either instantaneous or statistical. In the case of instantaneous CSI, the beamforming vectors that can attain the outer boundary of the achievable rate region are characterized for an arbitrary number of multiantenna transmitters and single-antenna receivers. This characterization only requires local CSI and justifies distributed precoding design based on a novel virtual SINR framework, which can handle an arbitrary SNR and achieves the optimal multiplexing gain. The local power allocation between terminals is solved heuristically. Conceptually, analogous results for the achievable rate region characterization and precoding design are derived in the case of local statistical CSI. The

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant Agreement No. 228044. This work is supported in part by the FP6 project Cooperative and Opportunistic Communications in Wireless Networks (COOPCOM), Project Number: FP6-033533. It is also supported in part by the French ANR-funded project ORMAC. Parts of this work was previously presented at the IEEE Global Communications Conference (GLOBECOM) 2009.

E. Björnson and B. Ottersten are with the Signal Processing Laboratory, ACCESS Linnaeus Center, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden (e-mail: emil.bjornson@ee.kth.se; bjorn.ottersten@ee.kth.se). B. Ottersten is also with securityandtrust.lu, University of Luxembourg, L-1359 Luxembourg-Kirchberg, Luxembourg.

R. Zakhour and D. Gesbert are with the Mobile Communications Department, EURECOM, 06560 Sophia Antipolis, France (e-mail: randa.zakhour@eurecom.fr; david.gesbert@eurecom.fr).

benefits of distributed cooperative transmission are illustrated numerically, and it is shown that most of the performance with centralized cooperation can be obtained using only local CSI.

Index Terms

Coordinated multipoint (CoMP), network MIMO, base station cooperation, distributed precoding, rate region, virtual SINR.

I. Introduction

The performance of cellular communication systems can be greatly improved by multiple-input multiple-output (MIMO) techniques. Many algorithms have been proposed for the single-cell downlink scenario, where a base station communicates simultaneously with multiple terminals [1]. These approaches exploit various amounts of channel state information (CSI) and improve the throughput by optimizing the received signal gain and limiting the intra-cell interference. In multicell scenarios, these single-cell techniques are however obliged to treat the interference from adjacent cells as noise, resulting in a fundamental limitation on the performance [2]–[5]—especially for terminals close to cell edges.

In recent years, base station coordination (also known as network MIMO [4]) has been analyzed as a means of handling inter-cell interference. In principle, all base stations might share their CSI and data through backhaul links, which enable coordinated transmission that manages the interference as in a single cell with either total [6] or per-group-of-antennas power constraints [7]. Unfortunately, the demands on backhaul capacity and computational power scale rapidly with the number of cells [8], [9], which makes this approach unsuitable for practical systems. Thus, there is a great interest in distributed forms of cooperation that reduce the backhaul signaling and precoding complexity, while still benefiting from robust interference control [9]–[12]. Two major considerations in the design of such schemes are to which extent the cooperation is managed centrally (requires CSI sharing) and whether each terminal should be served by multiple base stations (requires data sharing).

We consider the scenario of base stations equipped with multiple antennas and terminals with a single antenna each. In this context, the multiple-input single-output interference channel (MISO IC) represents the special case when each base station only serves a single unique terminal, but can share CSI to manage co-terminal interference. Although each base station aims at maximizing the rate achieved by its own terminal, cooperation over the MISO IC can greatly improve the performance [13]. The achievable rate region was characterized in [14] and the authors proposed a game-theoretic precoding design based on full CSI sharing [13]. Distributed precoding that only exploits locally available CSI can be achieved when each base station balances the ratio between signal gain at the

intended terminal and the interference caused at other terminals [15]–[17]. Recently, this approach has been shown to attain optimal rate points [17].

Herein, we address the problem of distributed network MIMO where the cooperating base stations share knowledge of the data symbols but have local CSI only, thereby reducing the feedback load on the uplink and avoiding cell-to-cell CSI exchange. The fundamental difference from the MISO IC is that multiple base stations can cooperate on serving each terminal, which means that the achievable rate region is larger [18]. In addition, the number of terminals is not limited by the number of base stations. In this paper, we derive a characterization of the optimal linear precoding strategy, which formally justifies a distributed approach that treats the system as a superposition of broadcast channels. This leads to novel distributed beamforming and power allocation strategies. The major contributions are:

- We characterize the achievable rate region for network MIMO with an arbitrary number of links and antennas at the transmitters, and either instantaneous or statistical CSI. The optimal beamformers are shown to belong to a certain subspace defined using local CSI. This parametrization provides understanding and a structure for heuristic precoding.
- We propose a distributed virtual SINR framework based on uplink-downlink duality theory [19]. This framework is used for distributed beamforming design and power allocation with local instantaneous CSI, and handles an arbitrary number of links. It achieves the optimal multiplexing gain and numerical examples show good and stable performance at all SNRs, which makes it more practical than distributed maximum ratio transmission (MRT) and zero-forcing (ZF).
- We extend this framework to handle beamforming design with local statistical CSI, for cases when instantaneous fading information is unavailable. A heuristic power allocation scheme is also proposed under these conditions.

Preliminary results with two base stations and two terminals were presented in [18]. The performance and complexity differences between centralized and distributed precoding are discussed in [20]. An alternative approach based on superposition of ICs is analyzed in [21].

Notations: Boldface (lower case) is used for column vectors, $\mathbf{x}$, and (upper case) for matrices, $\mathbf{X}$. Let $\mathbf{X}^T$, $\mathbf{X}^H$, and $\mathbf{X}^*$ denote the transpose, the conjugate transpose, and the conjugate of $\mathbf{X}$, respectively. The orthogonal projection matrix onto the column space of $\mathbf{X}$ is $\mathbf{\Pi}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H$, and that onto its orthogonal complement is $\mathbf{\Pi}_{\mathbf{X}}^{\perp} = \mathbf{I} - \mathbf{\Pi}_{\mathbf{X}}$, where $\mathbf{I}$ is the identity matrix. $\mathcal{CN}(\bar{\mathbf{x}}, \mathbf{Q})$ is used to denote circularly symmetric complex Gaussian random vectors, where $\bar{\mathbf{x}}$ is the mean and $\mathbf{Q}$ is the covariance matrix.

II. System Model

The communication scenario herein consists of K_r single-antenna receivers[1] (e.g., active mobile terminals) and K_t transmitters (e.g., base stations in a cellular system) equipped with N_t antennas each. The jth transmitter and kth receiver are denoted BS$_j$ and MS$_k$, respectively, for $j \in \{1,\ldots,K_t\}$ and $k \in \{1,\ldots,K_r\}$. This setup is illustrated in Fig. 1 for $N_t = 8$. Let $\mathbf{x}_j \in \mathbb{C}^{N_t}$ be the signal transmitted by BS$_j$ and let the corresponding received signal at MS$_k$ be denoted by $y_k \in \mathbb{C}$. The propagation channel to MS$_k$ is assumed to be narrowband, flat and Rayleigh fading with the system model

$$y_k = \sum_{j=1}^{K_t} \mathbf{h}_{jk}^H \mathbf{x}_j + n_k \tag{1}$$

where $\mathbf{h}_{jk} \in \mathcal{CN}(\mathbf{0}, \mathbf{Q}_{jk})$ represents the channel between BS$_j$ and MS$_k$ and $n_k \in \mathcal{CN}(0,\sigma_k^2)$ is white additive noise. The channel correlation matrix $\mathbf{Q}_{jk} = \mathbb{E}\{\mathbf{h}_{jk}\mathbf{h}_{jk}^H\} \in \mathbb{C}^{n_T \times n_T}$ is positive semi-definite. Throughout the paper, each receiver MS$_k$ has full local CSI (i.e., perfect estimates of $\mathbf{h}_{jk}$ for $j = 1,\ldots,K_t$). At the transmitter side, we will distinguish between two different types of local CSI:

- *Local Instantaneous CSI*: BS$_j$ knows the current channel vector $\mathbf{h}_{jk}$ and the noise power σ_k^2, for $k = 1,\ldots,K_r$.
- *Local Statistical CSI*: BS$_j$ knows the statistics of $\mathbf{h}_{jk}$ (e.g., type of distribution and $\mathbf{Q}_{jk}$) and the noise power σ_k^2, for $k = 1,\ldots,K_r$.

Observe that in both cases, the philosophy is that transmitters only have CSI that can be obtained locally (either through feedback or reverse-link estimation [22]). Hence, there is no exchange of CSI between them, thus allowing the scalability of multicell cooperation to large and dense networks. For simplicity, each transmitter has CSI for its links to all receivers, which is non-scalable when the resources for CSI acquisition are limited. However, it is still a good model for large networks as most terminals will be far away from any given transmitter and thus have negligibly weak channel gains.

A. Cooperative Multicell Precoding

Let $s_k \in \mathcal{CN}(0,1)$ be the data symbol intended for MS$_k$. Unlike the MISO IC [13]–[17], we assume that the data symbols intended for all receivers are available at all transmitters. This enables cooperative precoding techniques, where each receiver is served simultaneously by all the transmitters in the area[2]. Herein, we will consider distributed linear precoding where each transmitter selects its

[1]The results herein also apply to simple multi-antenna receivers that fix their receive beamforming (e.g., antenna selection) prior to base station optimization.

[2]This assumption will ease the exposition, but in practice the subset of terminals served by a given base station will be determined by a scheduler. This scheduling problem is however beyond the scope of this paper.

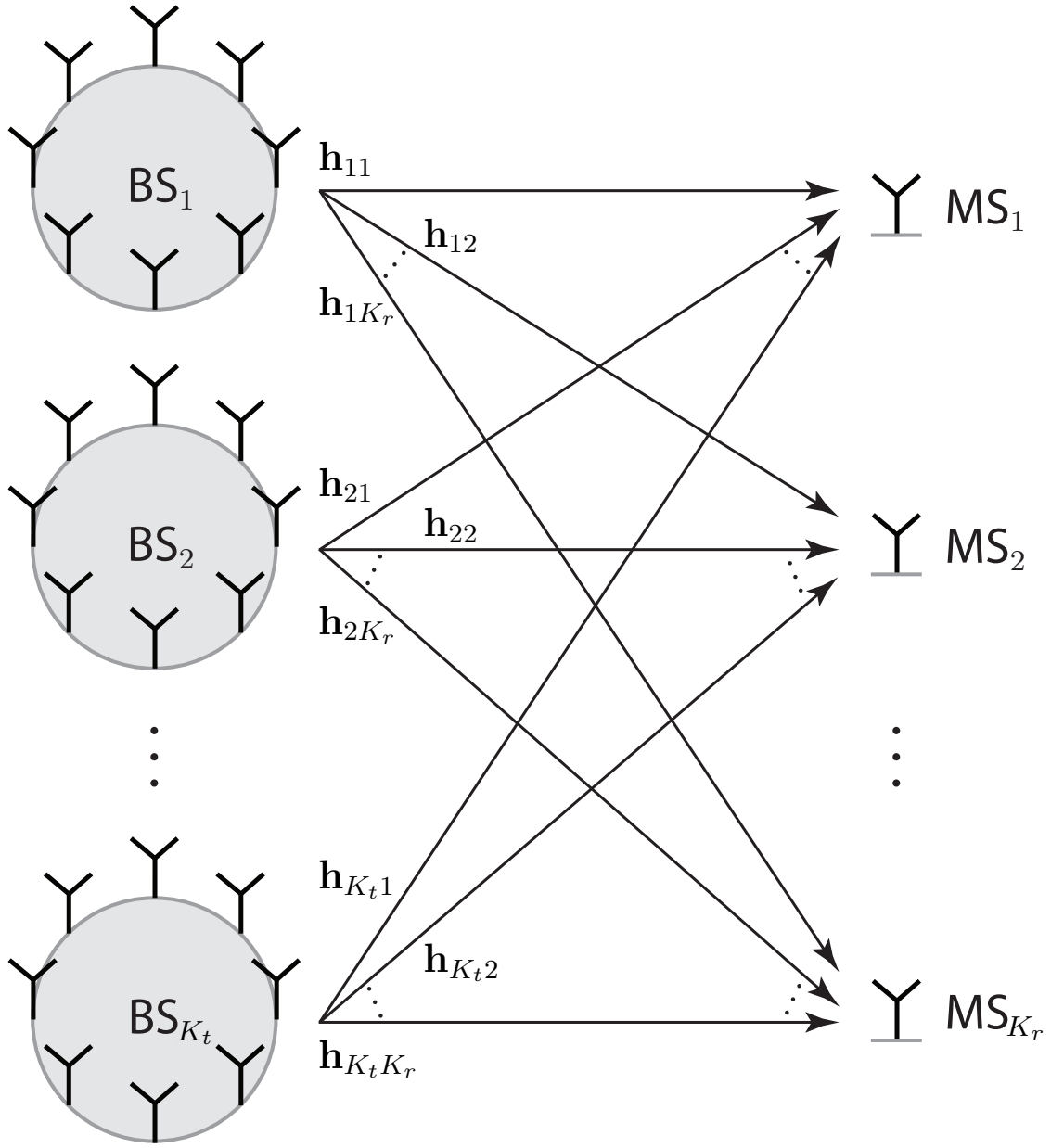


Fig. 1. The basic scenario with K_t base stations and K_r terminals (illustrated for $N_t = 8$ transmit antennas).

beamforming vectors independently using only local CSI, as defined above. Proper transmission synchronization is however required to avoid inter-symbol interference. The signal transmitted by BS_j is

$$\mathbf{x}_j = \sum_{k=1}^{K_r} \sqrt{p_{jk}} \mathbf{w}_{jk} s_k \tag{2}$$

where the beamforming vectors $\mathbf{w}_{jk}$ have unit norms (i.e., $\|\mathbf{w}_{jk}\| = 1$) and p_{jk} represents the power allocated for transmission to MS_k from BS_j. BS_j is subject to an individual average power constraint of P_j; that is, $\mathbb{E}\{\|\mathbf{x}_j\|^2\} = \sum_{k=1}^{K_r} p_{jk} \leq P_j$. Thus, the main differences between the scenario at hand

and the MISO broadcast channel (BC) is that in the latter *all* antennas are controlled by a central unit with CSI of *all* links and a joint power constraint.

When the receivers treat co-terminal interference as noise, the instantaneous SINR at MS_k is

$$\mathrm{SINR}_k = \frac{\left|\sum_{j=1}^{K_t} \sqrt{p_{jk}} \mathbf{h}_{jk}^H \mathbf{w}_{jk}\right|^2}{\sum_{\substack{\bar{k}=1 \\ \bar{k}\neq k}}^{K_r} \left|\sum_{j=1}^{K_t} \sqrt{p_{j\bar{k}}} \mathbf{h}_{jk}^H \mathbf{w}_{j\bar{k}}\right|^2 + \sigma_k^2} \quad \text{for } k = 1, \ldots, K_r. \tag{3}$$

The maximal achievable instantaneous transmission rate is accordingly $R_k = \log_2(1 + \mathrm{SINR}_k)$.

In the case of local statistical CSI at each transmitter, we introduce the notation $a_{\bar{k}k} \triangleq \sum_{j=1}^{K_t} \sqrt{p_{j\bar{k}}} \mathbf{h}_{jk}^H \mathbf{w}_{j\bar{k}}$, $\mathbf{S}_k \triangleq \sum_{j=1}^{K_t} \mathbf{W}_j^H \mathbf{Q}_{jk} \mathbf{W}_j$, and $\mathbf{W}_j \triangleq [\sqrt{p_{j1}}\mathbf{w}_{j1} \ldots \sqrt{p_{jK_r}}\mathbf{w}_{jK_r}]$. The collective beamforming matrix $\mathbf{W}_j$ is statistically independent of the instantaneous channel realizations $\mathbf{h}_{jk}$, since it is only based on statistical CSI. Then, the stochastic behavior of the SINR in (3), seen by the transmitters, is clarified by the alternative expression

$$\mathrm{SINR}_k = \frac{|a_{kk}|^2}{\sum_{\substack{\bar{k}=1 \\ \bar{k}\neq k}}^{K_r} |a_{\bar{k}k}|^2 + \sigma_k^2} \quad \text{with } \mathbf{a}_k = [a_{1k} \ \ldots \ a_{K_r k}]^H \in \mathcal{CN}(\mathbf{0}, \mathbf{S}_k). \tag{4}$$

When the transmitters only have statistical CSI, they can only optimize an average performance measure. Herein, we therefore consider the expected achievable transmission rate, $\mathbb{E}\{R_k\} = \mathbb{E}\{\log_2(1 + \mathrm{SINR}_k)\}$. Using the notation introduced above, it can be calculated using the next theorem. The results will be used for precoding design in Section IV-B.

Theorem 1. Let $\tilde{\mathbf{S}}_k$ be the matrix obtained by removing the kth column and kth row of $\mathbf{S}_k$. Then,

$$\begin{aligned} \mathbb{E}\{\log_2(1 + \mathrm{SINR}_k)\} = &\sum_{m=1}^{\mathrm{rank}(\mathbf{S}_k)} \frac{e^{\frac{\sigma_k^2}{\mu_m}} E_1\left(\frac{\sigma_k^2}{\mu_m}\right)}{\log(2) \prod_{l\neq m} (1 - \frac{\mu_l}{\mu_m})} \\ &- \sum_{m=1}^{\mathrm{rank}(\tilde{\mathbf{S}}_k)} \frac{e^{\frac{\sigma_k^2}{\lambda_m}} E_1\left(\frac{\sigma_k^2}{\lambda_m}\right)}{\log(2) \prod_{l\neq m} (1 - \frac{\lambda_l}{\lambda_m})} \end{aligned} \tag{5}$$

where $\mu_1, \ldots, \mu_{\mathrm{rank}(\mathbf{S}_k)}$ and $\lambda_1, \ldots, \lambda_{\mathrm{rank}(\tilde{\mathbf{S}}_k)}$ are the non-zero distinct eigenvalues of $\mathbf{S}_k$ and $\tilde{\mathbf{S}}_k$, respectively. Here, $E_1(x) = \int_1^\infty e^{-xu}/u \, du$ is the exponential integral.

Proof: The proof is given in Appendix A. ■

As stated in Theorem 1, a requirement for the expressions in (5) is that all non-zero eigenvalues of $\mathbf{S}_k$ and $\tilde{\mathbf{S}}_k$ are distinct. In the unlikely event of non-distinct eigenvalues, general expressions can be derived using the theory of [23] and [24].

III. Characterization of the Pareto Boundary

In this section, we analyze the achievable rate region for the scenario at hand, which will provide a precoding structure that is used for practical precoding design in Section IV. Since the receivers are assumed to treat co-channel interference as noise (i.e., not attempting to decode and subtract the interference), the achievable rate region will in general be smaller than the information theoretic capacity region. This limiting assumption is however relevant in the case of simple receiver structures. In the case of instantaneous CSI, we define the achievable rate region as

$$\mathcal{R}_{\text{instant}} = \bigcup_{\substack{\{\mathbf{w}_{jk}\in\mathbb{C}^{N_t}\ \forall j,k;\ \|\mathbf{w}_{jk}\|=1\}\\ \{p_{jk}\ \forall j,k;\ p_{jk}\geq 0,\sum_{k=1}^{K_r} p_{jk}\leq P_j\}}} (R_1,\ldots,R_{K_r}) \tag{6}$$

while in the case of statistical CSI we define the achievable expected rate region as

$$\mathcal{R}_{\text{statistic}} = \bigcup_{\substack{\{\mathbf{w}_{jk}\in\mathbb{C}^{N_t}\ \forall j,k;\ \|\mathbf{w}_{jk}\|=1\}\\ \{p_{jk}\ \forall j,k;\ p_{jk}\geq 0,\sum_{k=1}^{K_r} p_{jk}\leq P_j\}}} (\mathbb{E}\{R_1\},\ldots,\mathbb{E}\{R_{K_r}\}). \tag{7}$$

Observe that all rates are functions of all $\mathbf{w}_{jk}$ and p_{jk}, although not written explicitly. The above rate regions characterize, respectively, all rate tuples $(R_1,\ldots,R_{K_r})$ and expected rate tuples $(\mathbb{E}\{R_1\},\ldots,\mathbb{E}\{R_{K_r}\})$ that are achievable with feasible precoding strategies, regardless of how these strategies are obtained. Our assumption of local CSI at the transmitters determines which rate tuples can be reached by practical algorithmic selection of $\mathbf{w}_{jk}$ and p_{jk}, since it restricts the latter to be functions of the local knowledge alone, as opposed to being functions of the whole channel knowledge as traditionally assumed (see Section IV).

The outer boundary of $\mathcal{R}$ is known as the Pareto boundary. The rate tuples on this boundary are Pareto optimal, which means that the rate achieved by MS_k cannot be increased without decreasing the rate of any of the other receivers. Each Pareto optimal rate tuple maximizes a certain weighted sum rate. We have the following definition of the Pareto boundary in the case of instantaneous CSI:

Definition 1. Consider all achievable rate tuples $(R_1,\ldots,R_{K_r})$. The Pareto boundary consists of all such tuples for which there exist no *non-identical* achievable rate tuple $(S_1,\ldots,S_{K_r})$ with $S_k \geq R_k$ for all $k = 1,\ldots,K_r$.

The corresponding definition with statistical CSI is achieved by replacing all rates by their expectations. Next, we will parameterize the Pareto boundary of $\mathcal{R}_{\text{instant}}$ by showing that beamformers, $\mathbf{w}_{jk}$, that can be used to attain the boundary lie in a certain subspace defined by only local CSI and that full transmit power ($\sum_{k=1}^{K_r} p_{jk} = P_j$) should be used by all base stations. In Section III-B, we derive a similar characterization of the Pareto boundary of $\mathcal{R}_{\text{statistic}}$ for systems with statistical CSI.

A. Characterization with Instantaneous CSI

Two classic beamforming strategies are maximum ratio combining (MRT) and zero-forcing (ZF), which maximizes the received signal power and minimizes the co-terminal interference, respectively. In the special case of $K_r = 2$, these beamforming vectors are aligned with $\mathbf{h}_{jk}$ and $\mathbf{\Pi}^{\perp}_{\mathbf{h}_{j\bar{k}}}\mathbf{h}_{jk}$, respectively, for $\bar{k} \neq k$. It was shown in [14] that the Pareto boundary of the MISO IC and BC with $K_t = K_r = 2$ can only be attained by beamformers that are linear combinations of MRT and ZF. This optimal strategy is interesting from a game theoretical perspective, since it can be interpreted as a combination of the selfish MRT and the altruistic ZF approach.

The system defined in Section II represents cooperative multicell precoding with data sharing. This scenario is fundamentally different from the MISO IC as the data sharing enables terminals to be served by multiple transmitters, and thus the achievable rate region can be considerably larger. The following theorem derives the optimal precoding characterization for this scenario, which turns out to be a conceptually similar combination of MRT and ZF. It also constitutes a novel extension to an arbitrary number of transmitters/receivers.

Theorem 2. For each rate tuple $(R_1, \ldots, R_{K_r})$ on the Pareto boundary it holds that

i) It can be achieved by beamformers $\mathbf{w}_{jk}$ that fulfill

$$\mathbf{w}_{jk} \in \operatorname{span}\left(\{\mathbf{h}_{jk}\} \bigcup_{\bar{k}\neq k} \{\mathbf{\Pi}^{\perp}_{\mathbf{h}_{j\bar{k}}}\mathbf{h}_{jk}\} \right) \quad \text{for all } j,k; \tag{8}$$

ii) If $\mathbf{h}_{jk} \notin \operatorname{span}(\bigcup_{\bar{k}\neq k}\{\mathbf{h}_{j\bar{k}}\})$ for some k, then a necessary condition for Pareto optimality is that BS_j uses full power (i.e., $\sum_k p_{jk} = P_j$) and selects $\mathbf{w}_{jk}$ that satisfy (8) for all k.

Proof: The proof is given in Appendix A. ■

The theorem implies that to attain rate tuples on the Pareto boundary, all transmitters are required to use full transmit power (except in a special case with zero probability) and use beamforming vectors that can be expressed as

$$\mathbf{w}_{jk} = \gamma_{jk}^{(k)}\mathbf{h}_{jk} + \sum_{\substack{\bar{k}=1 \\ \bar{k}\neq k}}^{K_r} \gamma_{jk}^{(\bar{k})}\mathbf{\Pi}^{\perp}_{\mathbf{h}_{j\bar{k}}}\mathbf{h}_{jk}, \quad \text{for all } j,k, \tag{9}$$

for some coefficients $\gamma_{jk}^{(1)}, \ldots, \gamma_{jk}^{(K_r)} \in \mathbb{C}$. This is a linear combination of $K_r - 1$ zero-forcing vectors $\mathbf{\Pi}^{\perp}_{\mathbf{h}_{j\bar{k}}}\mathbf{h}_{jk}$ (each inflicting zero interference at $\mathrm{MS}_{\bar{k}}$) and the following MRT beamformer:

Definition 2 (*Maximum Ratio Transmission*).

$$\mathbf{w}_{jk}^{(\mathrm{MRT})} = \frac{\mathbf{h}_{jk}}{\|\mathbf{h}_{jk}\|} \quad \text{for all } j,k.$$

Complete ZF beamforming that inflicts zero interference on all co-terminals exists if $N_t \geq K_r$ and can be defined in the following way, but observe that it can also be expressed as the linear combination in (9).

Definition 3 (*Zero-Forcing Beamforming*). If $N_t \geq K_r$,

$$\mathbf{w}_{jk}^{(\mathrm{ZF})} = \frac{\left(\mathbf{I} - \sum_{l=1}^{m_{jk}} \mathbf{\Pi}_{\mathbf{e}_{jk}^{(l)}}\right)\mathbf{h}_{jk}}{\left\|\left(\mathbf{I} - \sum_{l=1}^{m_{jk}} \mathbf{\Pi}_{\mathbf{e}_{jk}^{(l)}}\right)\mathbf{h}_{jk}\right\|}$$

where $\mathbf{e}_{jk}^{(1)}, \ldots, \mathbf{e}_{jk}^{(m_{jk})}$ is an orthogonal basis of $\mathrm{span}(\bigcup_{\bar{k}\neq k}\{\mathbf{h}_{j\bar{k}}\})$, for all j, k.

Several important conclusions can be drawn from the theorem. Firstly, the precoding characterization reduces the precoding complexity if $N_t > K_r$ (since the beamforming vectors we are looking for each lie in a K_r-dimensional subspace[3]), especially if N_t is large. Secondly, the optimal beamforming approach can be interpreted as a linear combination of the selfish MRT approach and altruistic interference control towards each co-terminal. This behavior has been pointed out in [14] for the MISO IC, but not for multicell precoding systems. Thirdly, the characterization is defined using only local CSI, while global information is required to find the optimal coefficients $\gamma_{jk}^{(\bar{k})}$. In Section IV, we discuss heuristic approaches for distributed computation of the coefficients and evaluate their performance in Section V. Apart from selecting beamforming vectors, it is necessary to perform optimal power allocation to attain the Pareto boundary. Some power allocation strategies that exploit local CSI are also provided in Section IV.

B. Characterization with Statistical CSI

Next, we characterize the Pareto boundary in a similar manner as in Theorem 2, but for the case of statistical CSI. It was shown in [25], for the MISO IC, that an exact parametrization can be derived when the correlation matrices are rank deficient. This is however rarely the case in practice and therefore we concentrate on general spatially correlated channels and characterize their correlation matrices, $\mathbf{Q}_{jk}$. Depending on the antenna distance and the amount of scattering, the channels from transmit antennas to the receiver have varying spatial correlation; large antenna spacing and rich scattering correspond to low spatial correlation, and vice versa. High correlation translates into large eigenvalue spread in $\mathbf{Q}_{jk}$ and low correlation to almost identical eigenvalues. The existence of strongly structured spatial correlation has been verified experimentally, in both outdoor [26] and indoor [27] scenarios, and we will show herein how to exploit these results in the context of multicell precoding. In particular, these results suggest the existence of a dominating subspace.

Similar to [28], we partition the eigenvalue decomposition $\mathbf{Q}_{jk} = \mathbf{U}_{jk}\mathbf{\Lambda}_{jk}\mathbf{U}_{jk}^H$ of the correlation matrix $\mathbf{Q}_{jk}$ in signal and interference subspaces based on the size of the eigenvalues. Assume that the

[3]In practice, this dimension can be further reduced by ignoring inactive receivers and those with negligible link gains to the transmitter.

eigenvalues in the diagonal matrix $\mathbf{\Lambda}_{jk}$ are ordered decreasingly with the corresponding eigenvectors as columns of the unitary matrix $\mathbf{U}_{jk}$. Then, we partition $\mathbf{U}_{jk}$ as

$$\mathbf{U}_{jk} = [\mathbf{U}_{jk}^{(D)} \ \mathbf{U}_{jk}^{(0)}] \tag{10}$$

where $\mathbf{U}_{jk}^{(D)} \in \mathbb{C}^{N_t \times N_d}$ spans the subspace associated with the $N_d \leq N_t$ *dominating* eigenvalues. The transmit power allocated to these eigendirections will have large impact on the SINR. Consequently, data transmission should take place in the range of eigendirections included in $\mathbf{U}_{jk}^{(D)}$, while one should avoid receiving interference in these directions. Assuming that the *non-dominating* eigenvalues associated with the remaining eigenvectors in $\mathbf{U}_{jk}^{(0)} \in \mathbb{C}^{N_t \times N_t - N_d}$ are much smaller than the dominating ones, the interference in this subspace will be limited. The design parameter N_d depends strongly on the amount of spatial correlation, and can be a small fraction of N_t in an outdoor cellular scenario. In completely uncorrelated environments, the partitioning can be ignored since $N_d = N_t$. Feedback of instantaneous channel norms and receive beamforming (in the case of multi-antenna receivers) can increase the effective spatial correlation and thereby decrease N_d [23]. In practice, careful measurements are necessary to determine the value that maximizes the average throughput.

Now, we will characterize the Pareto boundary of the achievable expected rate region for cooperative multicell precoding. It will be done in an approximate manner, using the eigenvector partitioning in (10). The following theorem is more general than its counterpart for the MISO IC in [25] as it considers an arbitrary number of transmitters/receivers and correlation matrices of full rank.

Theorem 3. Let the sum of non-dominating eigenvalues in $\mathbf{Q}_{1k}, \ldots, \mathbf{Q}_{K_t k}$ be denoted $\epsilon_k \triangleq \sum_{j=1}^{K_t} \mathrm{tr}((\mathbf{U}_{jk}^{(0)})^H \mathbf{Q}_{jk} \mathbf{U}_{jk}^{(0)})$, for all k. For each expected rate tuple $(\mathbb{E}\{R_1\}, \ldots, \mathbb{E}\{R_{K_r}\})$ on the Pareto boundary, there exists with probability one another achievable tuple $(\mathbb{E}\{\widetilde{R}_1\}, \ldots, \mathbb{E}\{\widetilde{R}_{K_r}\})$ that fulfills $\mathbb{E}\{R_k\} = \mathbb{E}\{\widetilde{R}_k\} + o(\epsilon_k)$ for all k, where the small o function means that $\mathbb{E}\{R_k\} - \mathbb{E}\{\widetilde{R}_k\} \to 0$ as $\epsilon_k \to 0$. For the rate tuple $(\mathbb{E}\{\widetilde{R}_1\}, \ldots, \mathbb{E}\{\widetilde{R}_{K_r}\})$ it holds that

i) It can be reached using beamforming vectors

$$\mathbf{w}_{jk} \in \mathrm{span}\left(\{\mathbf{U}_{jk}^{(D)}\} \bigcup_{\bar{k} \neq k} \{\mathbf{\Pi}_{\mathbf{U}_{j\bar{k}}^{(D)}}^{\perp} \mathbf{U}_{jk}^{(D)}\} \right) \quad \text{for all } j, k; \tag{11}$$

ii) If $\mathrm{span}(\bigcup_{k=1}^{K_r} \{\mathbf{U}_{jk}^{(D)}\}) \neq \mathbb{C}^{N_t}$ for some j, it can be reached when BS_j uses full power.

Proof: The proof is given in Appendix A. ■

Observe that in the special case of zero-valued eigenvalues within each non-dominating eigenspace $\mathbf{U}_{jk}^{(0)}$ (i.e., $\epsilon_k = 0$), the theorem gives an exact characterization since $\mathbb{E}\{R_k\} = \mathbb{E}\{\widetilde{R}_k\}$.

There are clear similarities between the precoding characterization in (8) for instantaneous CSI, and its counterpart in (11) for statistical CSI. In both cases, all interesting beamforming vectors are linear combinations of eigenvectors that (selfishly) provide strong signal gain and that (altruistically)

limit the interference at co-terminals. These eigenvectors are defined using local CSI, which enables distributed precoding in a structured manner (see Section IV). The results with statistical CSI are however weaker, which is natural since each channel vector belongs (approximately) to a subspace of rank N_d while the channels with instantaneous CSI are known vectors (i.e., rank one). In the special case of $N_d = 1$, the characterization in Theorem 3 becomes essentially the same as in Theorem 2.

From these observations, it is natural to consider the two extremes that satisfy the beamforming characterization, namely MRT and ZF. Analogously to the MRT and ZF approaches with instantaneous CSI in Definitions 2 and 3, we propose extensions to the case of statistical CSI. The straightforward generalization of MRT is to use the dominating eigenvector of $\mathbf{Q}_{jk}$ as beamformer in $\mathbf{w}_{jk}$. We denote the normalized eigenvector associated with the largest eigenvalue of $\mathbf{Q}_{jk}$ by $\mathbf{u}_{jk}^{(D)}$. The generalization of ZF is to maximize the average received signal power under the condition that the beamformer lies in the non-dominating eigen-subspace $\mathbf{U}_{j\bar{k}}^{(0)}$ of all co-terminals. Formally, we have the following precoding approaches.

Definition 4 (*Generalized MRT*).

$$\mathbf{w}_{jk}^{\text{(G-MRT)}} = \mathbf{u}_{jk}^{(D)} \quad \text{for all } j, k.$$

Definition 5 (*Generalized ZF*).

$$\mathbf{w}_{jk}^{\text{(G-ZF)}} = \mathbf{v}_{jk}$$

where $\mathbf{v}_{jk}$ is the normalized dominating eigenvector of $\mathbf{\Pi}_{\mathcal{S}_{jk}}^{\perp} \mathbf{Q}_{jk} \mathbf{\Pi}_{\mathcal{S}_{jk}}^{\perp}$ with $\mathcal{S}_{jk} = \text{span}(\bigcup_{\bar{k} \neq k} \{\mathbf{U}_{j\bar{k}}^{(D)}\})$, for all j, k.

Observe that generalized ZF only exists for certain combinations of N_d and K_r as it is necessary that $\text{rank}(\mathcal{S}_{jk}) < N_t$. The generalizations in Definition 4 and 5 are made for multicell precoding. For broadcast channels, and in general, other generalizations are possible.

IV. Distributed Precoding with Local CSI

In the previous section, we characterized the beamforming vectors that can be used to attain the Pareto boundary of the achievable rate region. These are all linear combinations of MRT and ZF, the two extremes in beamforming. Intuitively, MRT is the asymptotically optimal strategy at low SNR, while ZF works well at high SNR or as the number of antennas increases. In general, the optimal strategy lies in between these extremes and cannot be determined without global CSI. Next, we use these insights to solve distributed precoding at an arbitrary SNR using only local CSI. The proposed beamforming approach is inspired by uplink-downlink duality for broadcast channels [19] and the transmit power is allocated heuristically by solving local optimization problems. The approach is asymptotically optimal at high SNR and the numerical evaluation in Section V shows a limited performance loss at all SNRs.

A. Transmission design with Local Instantaneous CSI

In general, we would like the precoding to solve

$$\underset{\mathbf{w}_{jk}\in\mathbb{C}^{N_t},p_{jk}\geq 0\ \forall j,k}{\text{maximize}} \sum_{k=1}^{K_r} \log_2(1+\text{SINR}_k) \tag{12}$$

subject to $\|\mathbf{w}_{jk}\| = 1$ and $\sum_{k=1}^{K_r} p_{jk} \leq P_j$ for all j and k. Unfortunately, none of the transmitters or receivers have sufficient CSI to calculate the sum rate, which makes the optimization problem in (12) intractable in a truly distributed scenario. Thus, we will look for distributed design criteria that allow approximated beamforming vectors, $\mathbf{w}_{jk}$, and power allocation coefficients, p_{jk}, of BS_j to be determined locally at the transmitter. The goal will still be to achieve performance close to the maximum sum rate. An important feature of the precoding characterization in Theorem 2 is that the optimal $\mathbf{w}_{jk}$ fulfills

$$\mathbf{w}_{jk} \in \text{span}\left(\{\mathbf{h}_{jk}\} \bigcup_{\bar{k}\neq k} \{\mathbf{\Pi}^{\perp}_{\mathbf{h}_{j\bar{k}}}\mathbf{h}_{jk}\} \right) \tag{13}$$

where all the spanning vectors are known locally at BS_j. In other words, the beamforming design consists of determining the coefficients of the linear combination in (9). To find heuristic coefficients, we exploit the following result based on the uplink-downlink duality theory of [19]:

Theorem 4. Assume that BS_j is the only active base station. Then, each Pareto optimal rate tuple of the corresponding achievable rate region is achieved by beamforming vectors

$$\mathbf{w}_{jk} = \underset{\|\mathbf{w}\|^2=1}{\arg\max} \frac{\beta_{jk}|\mathbf{h}_{jk}^H\mathbf{w}|^2}{\frac{\sigma_k^2}{P_j} + \sum_{\bar{k}\neq k}\beta_{j\bar{k}}|\mathbf{h}_{j\bar{k}}^H\mathbf{w}|^2} \tag{14}$$

for some positive coefficients β_{jk} with $\sum_{k=1}^{K_r}\beta_{jk} = 1$.

Proof: The proof is given in Appendix A. ■

Thus, in the special case of a MISO broadcast channel, the optimal beamforming vectors are achieved by maximizing the SINR-like expression in (14) where the signal power that BS_j generates at MS_k is balanced against the noise and interference power generated at all other receivers. We call it a *virtual SINR* as it originates from the dual virtual uplink [19] and does not directly represent the SINR of any of the links in the downlink. However, it is easy to show[4] that solutions to (14) are of the type described in (9) and (13). In fact, by varying the coefficients β_{jk}, different solutions within the span of Theorem 2 can be achieved. In general, the coefficients that provides the largest sum rate can only be found using global CSI.

Network MIMO can be seen as a superposition of K_t broadcast channels. We propose to exploit this fact for distributed precoding by letting each base station optimize its performance based on

[4]Observe that $\mathbf{w}$ should lie in the span of $\mathbf{h}_{jk}$, for all k, as no other directions will affect (14). We achieve (13) by rewriting this span following the approach in the proof of Theorem 2.

Theorem 4. In the superposition case, the noise term of (14) should be modified to compensate for the interference from other base stations, or equivalently the coefficients β_{jk} should be increased beyond what is allowed for pure broadcast channels. To account for stronger interference we therefore select $\beta_{jk} = 1$ (i.e., equal to its upper bound) and arrive at a novel distributed virtual SINR (DVSINR) beamforming approach:

Strategy 1. *BS_j should select its beamformers as*

$$\mathbf{w}_{jk}^{(DVSINR)} = \arg\max_{\|\mathbf{w}\|^2=1} \frac{|\mathbf{h}_{jk}^H \mathbf{w}|^2}{\frac{\sigma_k^2}{P_j} + \sum_{\bar{k}\neq k} |\mathbf{h}_{j\bar{k}}^H \mathbf{w}|^2} \quad \textit{for all } k. \tag{15}$$

Observe that the virtual SINR in (15) is a Rayleigh quotient and thus the maximization can be solved by straightforward eigenvalue techniques. For example,

$$\mathbf{w}_{jk}^{(\text{DVSINR})} = \frac{\mathbf{C}_{jk}^{-1}\mathbf{h}_{jk}}{\|\mathbf{C}_{jk}^{-1}\mathbf{h}_{jk}\|} \text{ where } \mathbf{C}_{jk} \triangleq \frac{\sigma_k^2}{P_j}\mathbf{I} + \sum_{\bar{k}\neq k} \mathbf{h}_{j\bar{k}}\mathbf{h}_{j\bar{k}}^H.$$

The solution to (15) is non-unique, since the virtual SINR is unaffected by phase shifts in $\mathbf{w}$. However, the expression above was selected to make $\mathbf{h}_{jk}^H\mathbf{w}$ positive and real-valued, which means that the signals arriving at a given terminal from different base stations will do so constructively. By its very definition, maximization of a virtual SINR effectively balances between the useful signal power at a target terminal and the interference generated at others; along with judicious power allocation coefficients p_{jk} for all j, k, this can be shown to provide good performance at all SNRs (see Section V). Observe that (15) gives solutions similar to MRT and ZF in the SNR regimes where these methods perform well (i.e., low SNR and high SNR, respectively). Asymptotic optimality conditions are provided by the following theorem.

Theorem 5. Assume an arbitrary power allocation which guarantees that each terminal is allocated non-zero total transmit power (i.e., $\sum_{j=1}^{K_t} p_{jk} > 0$ for all k). If $N_t \geq K_r$[5] and each $p_{jk} > 0$ increases with P_j, then with probability one DVSINR beamforming achieves the full multiplexing gain of K_r asymptotically in $P = \min_j P_j$.

Proof: The proof is given in Appendix A. ■

This means that the sum rate behaves as $K_r \log_2(P) + \text{constant}$ at high P. Thus, the absolute performance loss compared with optimal centralized precoding is bounded at high SNR and the relative loss goes to zero. The absolute loss is primarily due to the fact that distributed beamforming limits the magnitude of interference from *each* transmitter to *every* terminal, while the global solution can coordinate and cancel out the sum of interference from different transmitters. However, such centralized interference coordination is practically questionable even under optimal conditions [29].

[5]The constraint $N_t \geq K_r$ can be removed if each BS_j have non-negligible channel gain to at most N_t terminals.

The power allocation has a clear impact on the practical performance, although Theorem 5 holds for any allocation. Next, we propose a heuristic power allocation scheme for BS_j. This is based on the observation that with proper beamforming, the interference is negligible at both low and high SNR. Assuming constructive signal contributions from all base stations, the sum rate then becomes

$$\begin{aligned}&\sum_{k=1}^{K_r} \log_2(1+\mathrm{SINR}_k)\\ &\quad\approx \sum_{k=1}^{K_r} \log_2\left(1+\left|\sqrt{p_{jk}}\underbrace{\frac{\mathbf{h}_{jk}^H\mathbf{w}_{jk}}{\sigma_k}}_{=c_{jk}}+\underbrace{\sum_{\bar{j}\neq j}\sqrt{p_{\bar{j}k}}\frac{\mathbf{h}_{\bar{j}k}^H\mathbf{w}_{\bar{j}k}}{\sigma_k}}_{=d_{jk}}\right|^2\right)\end{aligned} \tag{16}$$

where $|c_{jk}|^2$ denotes the channel gain between BS_j and MS_k and $|d_{jk}|^2$ is the signal gain from the other transmitters (including power allocation). All c_{jk} and d_{jk} can be taken as positive real-valued, due to the assumption of transmission synchronization. For fixed values on all c_{jk} and d_{jk}, the power allocation at BS_j is solved by the following lemma.

Lemma 1. For a given j and some positive constants c_{jk}, d_{jk}, the optimization problem

$$\begin{aligned}\underset{p_{j1}\geq 0,\ldots,p_{jK_r}\geq 0}{\text{maximize}} \ &\sum_{k=1}^{K_r}\log_2\left(1+(\sqrt{p_{jk}}c_{jk}+d_{jk})^2\right)\\ \text{subject to} \ &\sum_{k=1}^{K_r} p_{jk}\leq P_j\end{aligned} \tag{17}$$

is solved by

$$\begin{aligned}\sqrt{p_{jk}} = \max\Bigg(&\sqrt[3]{a_{jk}+\sqrt{a_{jk}^2+b_{jk}^3}}\\ &-\frac{b_{jk}}{\sqrt[3]{a_{jk}+\sqrt{a_{jk}^2+b_{jk}^3}}}-\frac{2d_{jk}}{3c_{jk}},0\Bigg),\\ a_{jk}=\frac{18d_{jk}+2d_{jk}^3+9\alpha c_{jk}^2 d_{jk}}{54c_{jk}^3},&\quad b_{jk}=\frac{3-d_{jk}^2}{9c_{jk}^2}-\frac{\alpha}{3}.\end{aligned} \tag{18}$$

If $d_{jk}=0$, this reduces to $\sqrt{p_{jk}}=\max(\alpha-1/c_{jk}^2,0)$. The Lagrange multiplier $\alpha\geq 0$ is selected to fulfill the power constraint with equality.

Proof: The maximization of a concave function can be solved by standard Lagrangian methods, using the Karush-Kuhn-Tucker (KKT) conditions [30, Chapter 5.5]. In this case, the optimal power allocation follows from straightforward differentiation, solving of a third-order polynomial equation with respect to $\sqrt{p_{jk}}$, and identifying the two false roots. ■

The local channel gains c_{jk} are known at each BS_j, while the contributions d_{jk} from other transmitters are unknown when having local CSI. Thus, BS_j needs to estimate these parameters.

To avoid that all transmitters believe that someone else serves a given terminal, the estimate should be pessimistic. In a symmetric environment, the selection

$$d_{jk}^{\text{symmetric}} = \sqrt{\frac{P_j}{N_t} \sum_{\bar{k}=1}^{K_r} \frac{|\mathbf{h}_{j\bar{k}}^H \mathbf{w}_{j\bar{k}}|^2}{K_r \sigma_{\bar{k}}^2}} \tag{19}$$

represents that one other transmitter uses $1/N_t$ of its power to serve the terminal (the channel gain is estimated as the average gain from BS_j to all terminals). In other cases, the worst-case selection

$$d_{jk}^{\text{worst-case}} = 0 \tag{20}$$

can give better performance or robustness. In practice, d_{jk} should be considered design parameters and tuned based on measured properties of the actual propagation environment. For given d_{jk} and beamforming vectors $\mathbf{w}_{jk}$, we use Lemma 1 to propose the following power allocation scheme.

Strategy 2. *Using local CSI, an efficient power allocation at BS_j for $N_t \geq K_r$ is given by Lemma 1 using $c_{jk} = |\mathbf{h}_{jk}^H \mathbf{w}_{jk}|/\sigma_k$ and some d_{jk} that reflects the propagation environment.*

In the case $N_t < K_r$, the interference can in general not be considered negligible as was assumed in the heuristic power allocation. Thus, alternative power allocation schemes should be considered—for example, the simple scheme $p_{jk} = P_j \|\mathbf{h}_{jk}\|^2 / \sum_j \|\mathbf{h}_{jk}\|^2$ evaluated in [20] and [21].

Observe that the power allocation in Strategy 2 has the waterfilling behavior, which means that zero power is allocated to weak terminals. Thus, terminals far from the base station are disregarded automatically, which limits the computational complexity as K_t and K_r increases.

B. Transmission design with Local Statistical CSI

Next, we extend the precoding design in the previous subsection to the case of local statistical CSI. As in the previous case, maximizing a virtual SINR will balance the generated signal and interference powers. We propose the following novel extension where the Rayleigh quotient represents maximization of an SINR where expectation has been applied to the numerator and denominator (using that $\mathbb{E}\{|\mathbf{h}_{jk}^H \mathbf{w}|^2\} = \mathbf{w}^H \mathbf{Q}_{jk} \mathbf{w}$).

Strategy 3. *For given power allocation coefficients, BS_j should select its beamformers as*

$$\mathbf{w}_{jk}^{\text{(G-DVSINR)}} = \arg\max_{\|\mathbf{w}\|^2=1} \frac{\mathbf{w}^H \mathbf{Q}_{jk} \mathbf{w}}{\frac{\sigma_k^2}{P_j} + \sum_{\bar{k} \neq k} \mathbf{w}^H \mathbf{Q}_{j\bar{k}} \mathbf{w}} \quad \textit{for all } k. \tag{21}$$

Unlike the case of instantaneous CSI, beamforming design with statistical CSI cannot guarantee coherent arrival of useful signals at a given receiver, but an increase in signal power will improve the average rate. The distributed SINR beamforming vectors of (21) satisfy (approximately) the beamforming characterization in Theorem 3.

Finally, we derive a distributed power allocation scheme. Since the expected rate expression in (5) is complicated, we simplify it by neglecting the interference. For MS_k, the expected rate in Theorem 1 becomes

$$\begin{aligned}\mathbb{E}\{\log_2(1+\text{SINR}_k)\} &\approx \frac{e^{\sigma_k^2/\mu_1}E_1\left(\sigma_k^2/\mu_1\right)}{\log(2)} \\ &\approx \log_2\left(1+p_{jk}\underbrace{\frac{\mathbf{w}_{jk}^H\mathbf{Q}_{jk}\mathbf{w}_{jk}}{\sigma_k^2}}_{=f_{jk}}+\underbrace{\sum_{\bar{j}\neq j}\frac{p_{\bar{j}k}\mathbf{w}_{\bar{j}k}^H\mathbf{Q}_{\bar{j}k}\mathbf{w}_{\bar{j}k}}{\sigma_k^2}}_{=g_{jk}}\right)\end{aligned} \tag{22}$$

using the upper part of the bound $\frac{1}{2}\log(1+2\mu_1/\sigma_k^2) < e^{\sigma_k^2/\mu_1}E_1\left(\sigma_k^2/\mu_1\right) < \log(1+\mu_1/\sigma_k^2)$. Here, f_{jk} denotes the average channel gain between BS_j and MS_k and g_{jk} is an estimation of the average signal gain from the other transmitters (including power allocation). For fixed values on all g_{jk}, the power allocation at BS_j is solved by the following lemma.

Lemma 2. For a given j and some positive f_{jk}, g_{jk}, the optimization problem

$$\begin{aligned}&\text{maximize}\sum_{k=1}^{K_r}\log_2(1+p_{jk}f_{jk}+g_{jk}) \\ &\text{subject to}\sum_{k=1}^{K_r}p_{jk}\leq P_j,\, p_{j1}\geq 0,\ldots,p_{jK_r}\geq 0\end{aligned} \tag{23}$$

is solved by $p_{jk}=\max\left(\alpha-(1+g_{jk})/f_{jk},0\right)$, where the Lagrange multiplier $\alpha\geq 0$ is selected to fulfill the power constraint with equality.

Proof: The solution to this convex optimization problem follows from straightforward Lagrangian methods, see the proof of Lemma 1 for details. ■

Using only local statistical CSI, the average local channel gains f_{jk} are known at BS_j, while the contributions g_{jk} from other transmitters are unknown. Thus, BS_j needs to estimate these parameters, which can be done similarly to (19) and (20):

$$\begin{aligned}g_{jk}^{\text{symmetric}} &= \frac{P_j}{N_t}\sum_{\bar{k}=1}^{K_r}\frac{\mathbf{w}_{j\bar{k}}^H\mathbf{Q}_{j\bar{k}}\mathbf{w}_{j\bar{k}}}{K_r\sigma_{\bar{k}}^2}, \\ g_{jk}^{\text{worst-case}} &= 0.\end{aligned} \tag{24}$$

For given g_{jk} and beamforming vectors $\mathbf{w}_{jk}$, we use Lemma 2 to propose the following power allocation scheme.

Strategy 4. *Using local CSI, an efficient power allocation at BS_j is given by Lemma 2 using $f_{jk}=\mathbf{w}_{jk}^H\mathbf{Q}_{jk}\mathbf{w}_{jk}/\sigma_k^2$ and some g_{jk} that reflects the propagation environment.*

V. Numerical Examples

In this section, the performance of the distributed beamforming and power allocation strategies in Section IV will be illustrated numerically. The DVSINR approach in Strategy 1 will be compared

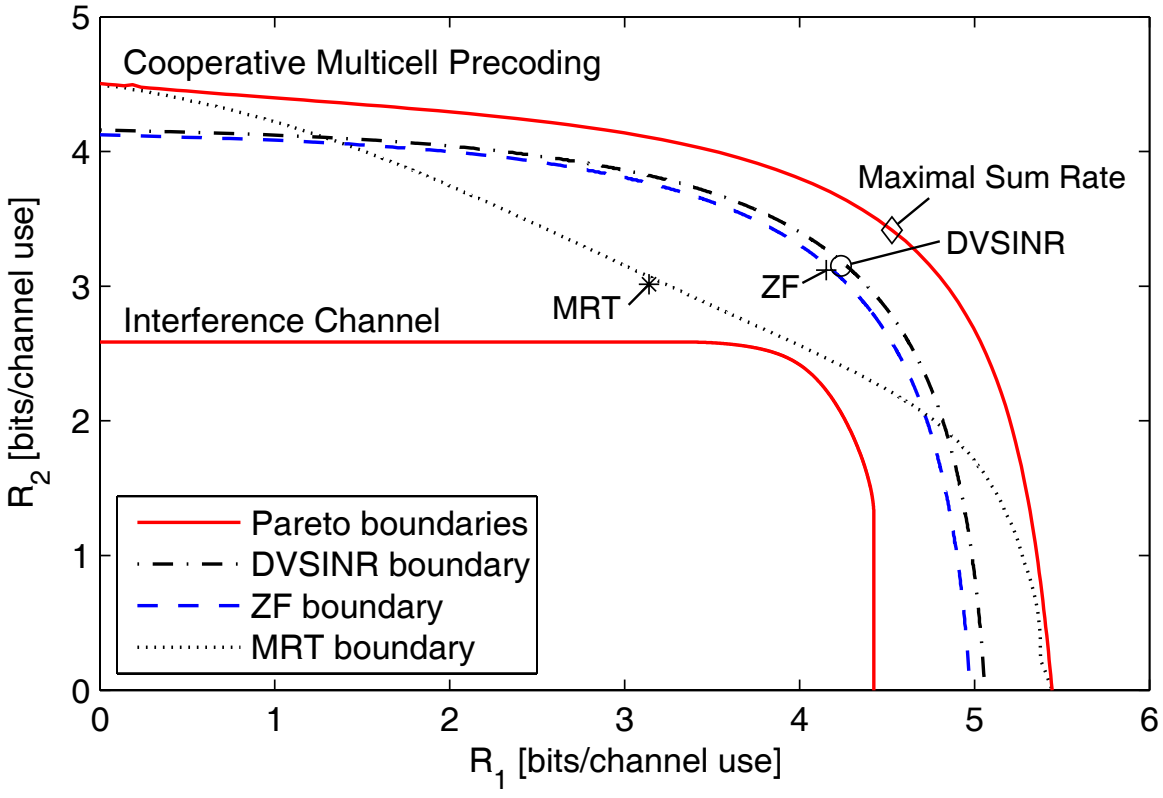


Fig. 2. Achievable rate regions with different beamforming and power allocation for $K_t = K_r = 2$, $N_t = 3$, and a random realization with $\mathbf{Q}_{11} = \mathbf{Q}_{22} = \mathbf{I}$, $\mathbf{Q}_{12} = \mathbf{Q}_{21} = 0.5\mathbf{I}$, and SNR 5 dB. The sum rate point and the points achieved by MRT, ZF, and DVSINR with the power allocation in Strategy 2 are shown for comparison.

with what we call *distributed MRT* and *distributed ZF*. These two approaches use the beamforming vectors in Definition 2 and 3, respectively. Observe that there are major differences from regular MRT and ZF for broadcast and interference channels, namely that the same message is sent from multiple transmitters with individual power constraints. When used, distributed MRT and ZF need to be combined with some power allocation, for example the one proposed in Strategy 2.

A. Transmitter-Receiver Pairs with Varying Cross-Links

First, consider the case of two transmitter-receiver links where the strengths of the cross-links are varied. The environment is spatially uncorrelated with $N_t = 3$, $\mathbf{Q}_{11} = \mathbf{Q}_{22} = \mathbf{I}$, and $\mathbf{Q}_{12} = \mathbf{Q}_{21} = \beta\mathbf{I}$, where β is the average cross link power. This represents a two-cell scenario where β determines how close the terminals are to the common cell edge. The SNR is defined as SNR $= P_j \mathrm{tr}(\mathbf{Q}_{jj})/N_t$ (with normalization $\sigma_k^2 = 1$) and represents the average SNR for beamforming to the own terminal.

In Figure 2, the Pareto boundary with $\beta = 0.5$ and an average SNR of 5 dB is given for a random realization of $\mathbf{h}_{jk}$ drawn from $\mathcal{CN}(\mathbf{0}, \mathbf{Q}_{jk})$, for all j, k. As a comparison, we give the Pareto boundary of the MISO IC [14] and show the outer boundaries of the achievable rate regions with the DVSINR

approach in Strategy 1, distributed MRT, and distributed ZF. The rate tuples achieved with the power allocation in Strategy 2 (with $d_{jk} = 0$) and the sum rate maximizing point are given as references. For the selected realization there is a clear performance gain of allowing cooperative multicell precoding as compared with forcing each transmitter to only communicate with its own receiver. As expected, MRT is useful to maximize the rate of only one of the terminals, while ZF and DVSINR are quite close to the optimal sum rate point. The proposed power allocation scheme provides performance close to the boundary of each achievable rate region.

In Figure 3, the average sum rates (over channel realizations) are given with optimal linear precoding (i.e., sum rate maximization through exhaustive search) and with the DVSINR approach in Strategy 1, distributed MRT, and distributed ZF (all three using the power allocation in Strategy 2 with $d_{jk} = 0$ to ensure robustness), and the VSINR approach in [17] for the MISO IC. The performance is shown for varying cross link power β and at an SNR of 0 or 10 dB. We observe that MRT is good at low SNR and/or weak cross link power, while ZF is better at high SNR and/or strong cross link power. However, the DVSINR approach is the most versatile strategy as it provides higher performance at low SNR and combines the benefits of MRT and ZF at high SNR. The three cooperative approaches clearly yield better performance than the non-cooperative VSINR approach.

In practice, two common terminal locations are close to a base station (i.e., high SNR with weak cross link power) and close to the cell edge (i.e., low SNR with strong cross link power). From Figure 3 it is clear that DVSINR is the only of the distributed schemes that provides good performance in both cases, which is an important property as both types of terminals appear simultaneously in practice. Thus, although distributed MRT and ZF achieve performance comparable to DVSINR in special cases, it is fair to say that the DVSINR scheme is the most versatile. Due to the distributed nature of the schemes, there is some performance loss compared with sum rate maximizing precoding. However, we argue that the backhaul and computational demands required to achieve the optimal solution may not be motivated in light of the small performance loss.

B. Quadratic Multicell Area

Next, we evaluate a scenario with terminals located in both cell centers and at cell edges. The scenario consists of four uniformly distributed terminals in a square with base stations in each of the corners. The power decay is proportional to $1/r^4$, where r is the distance from a transmitter, the SNR is defined as SNR $= P_j \mathrm{tr}(\mathbf{Q}_{jk})/N_t$ (with normalization $\sigma_k^2 = 1$), and its value in the center of the square represents the cell edge SNR. This represents a scenario where terminals are moving around in the area covered by four base stations. We will illustrate the performance with both instantaneous and statistical CSI.

In Figure 4, the average sum rate (over terminal locations and channel realizations) with instanta-

neous CSI and no spatial correlation is shown as a function of the SNR. In the case of $N_t = 4$, the DVSINR approach is superior to MRT and ZF at most SNRs, although ZF approaches DVSINR at very high cell edge SNR. The performance loss compared with optimal precoding is at most 15-20 percent, depending on the SNR, and will asymptotically approach zero since DVSINR achieves the optimal multiplexing gain (see Theorem 5). This raises the question of whether the high backhaul demands for achieving the optimal solution are justifiable in practice. In the case of $N_t = 2$ (with the power allocation in [20, Eq. (10)]), the performance of both DVSINR and MRT saturates at high SNR since $K_r > N_t$, but DVSINR still constitutes a major performance improvement compared with MRT. Distributed ZF does not exist for this number of antennas.

In Figure 5, the expected sum rate (over terminal locations) with $N_t = 6$, statistical CSI, and an angular spread of 10 degrees (as seen from a transmitter) is shown as a function of the SNR. The G-DVSINR approach in Strategy 3, G-MRT, and G-ZF (all using the power allocation in Strategy 4 with $g_{jk}^{\text{symmetric}}$) are compared with equal time sharing between the terminals and an upper bound consisting of the broadcast GZF approach in [28] that requires both statistical CSI and perfect instantaneous norm feedback. In this scenario, the G-DVSINR approach is clearly the better choice among the distributed methods; it even beats the upper bound at high SINR, since the performance GZF approach saturates at around 15 dB SNR. All the cooperative approaches outperform time sharing.

VI. Conclusion

We have considered cooperative multicell precoding in a system with an arbitrary number of multi-antenna transmitters and single-antenna receivers. The outer boundary of the achievable rate region was characterized for transmitters with either instantaneous or statistical CSI. At each transmitter, the spans of beamforming vectors that can attain this boundary only depend on local CSI, and can be interpreted as linear combinations of MRT and ZF vectors. This enables distributed precoding in a structured manner that only requires local CSI and processing. By viewing the multicell system as a superposition of broadcast channels, we propose a novel framework of distributed virtual SINR (DVSINR) beamforming that satisfies the optimal beamforming characterization and achieves the optimal multiplexing gain. It was applied for distributed beamforming with instantaneous and statistical CSI, along with two heuristic power allocation schemes. The performance of this approach was illustrated and shown to combine the benefits of conventional MRT and ZF, and outperform them at most SNRs. Finally, the loss in performance of having only local CSI is rather small, compared with the backhaul and computational demands of sharing and processing global CSI.

Acknowledgement

The authors would like to thank the reviewers for their insightful comments and suggestions, which have strengthened the analysis and improved the precoding performance.

APPENDIX A
COLLECTION OF LEMMAS AND PROOFS

Lemma 3. Let $b_k \in \mathcal{CN}(0,\mu_k)$ be independent random variables with distinct variances $\mu_k > 0$ for $k = 1,\ldots,K$, and let $\sigma_k^2 \geq 0$. Then,

$$\begin{aligned}&\mathbb{E}\left\{\log_2\left(\sigma_k^2+\sum_{k=1}^{K}|b_k|^2\right)\right\}\\&\quad=\log_2(\sigma_k^2)+\sum_{k=1}^{K}\frac{e^{\frac{\sigma_k^2}{\mu_k}}E_1\left(\frac{\sigma_k^2}{\mu_k}\right)}{\log(2)\prod_{l\neq k}(1-\frac{\mu_l}{\mu_k})}\end{aligned} \tag{25}$$

where $E_1(x) = \int_1^\infty e^{-xu}/u\,du$ is the exponential integral.

Proof: Let $z = \sum_{k=1}^K |b_k|^2/\sigma_k^2$ and observe that

$$\begin{aligned}&\mathbb{E}\left\{\log_2\left(\sigma_k^2+\sum_{k=1}^{K}|b_k|^2\right)\right\}\\&\quad=\log_2(\sigma_k^2)+\int_0^\infty\frac{\log(1+z)}{\log(2)}\sum_{k=1}^{K}\frac{\sigma_k^2e^{-\frac{\sigma_k^2}{\mu_k}z}}{\mu_k\prod_{l\neq k}(1-\frac{\mu_l}{\mu_k})}dz\end{aligned} \tag{26}$$

using the PDF expression for z in [23, Eq. 5]. The integrand that contains z is

$$\begin{aligned}\int_0^\infty\log(1+z)e^{-\frac{\sigma_k^2}{\mu_k}z}dz&=\int_1^\infty\log(\tilde{z})e^{-\frac{\sigma_k^2}{\mu_k}(\tilde{z}-1)}d\tilde{z}\\&=e^{\frac{\sigma_k^2}{\mu_k}}\Bigg(\underbrace{\Big[-\frac{\mu_k}{\sigma_k^2}\log(\tilde{z})e^{-\frac{\sigma_k^2}{\mu_k}\tilde{z}}\Big]_1^\infty}_{=0}+\frac{\mu_k}{\sigma_k^2}E_1\Big(\frac{\sigma_k^2}{\mu_k}\Big)\Bigg)\end{aligned} \tag{27}$$

where the first equality follows from the variable substitution $\tilde{z} = 1+z$ and the second from integration by parts. Substitution into (26) gives the final expression. ■

Proof of Theorem 1

Using the notation for the SINR in (4), the expected rate can be divided as

$$\begin{aligned}\mathbb{E}\{\log_2(1+\mathrm{SINR}_k)\}=&\mathbb{E}\Big\{\log_2\Big(\sum_{\bar{k}=1}^{K_r}|a_{\bar{k}k}|^2+\sigma_k^2\Big)\Big\}\\&-\mathbb{E}\Big\{\log_2\Big(\sum_{\substack{\bar{k}=1\\\bar{k}\neq k}}^{K_r}|a_{\bar{k}k}|^2+\sigma_k^2\Big)\Big\}.\end{aligned} \tag{28}$$

Observe that $\|\mathbf{a}_k\|^2 = \sum_{\bar{k}=1}^{K_r}|a_{\bar{k}k}|^2$ and since the Euclidean norm is invariant under unitary transformations, $\sum_{\bar{k}=1}^{\mathrm{rank}(\mathbf{S}_k)}|b_{\bar{k}}|^2$ has identical distribution for independent variables $b_{\bar{k}} \in \mathcal{CN}(0,\mu_{\bar{k}})$. Using these variables, we can apply Lemma 3 and achieve the first term in (5). The second term is achieved by a similar transformation based on eigenvalues of $\tilde{\mathbf{S}}_k$.

Proof of Theorem 2

Consider a rate tuple $(R_1, \ldots, R_{K_r})$ on the Pareto boundary that is achieved by beamforming vectors $\mathbf{w}_{jk}$ and power allocation p_{jk} for all j, k. The following approach can be taken (for each j, k) to replace $\mathbf{w}_{jk}$ with a beamformer that fulfills (8) and reduces the power usage, while achieving the same rate tuple. Let $\mathcal{A}_{jk} = \{\mathbf{h}_{jk}\} \bigcup_{\bar{k} \neq k} \{\mathbf{\Pi}^{\perp}_{\mathbf{h}_{j\bar{k}}} \mathbf{h}_{jk}\}$ and observe that the vector $\mathbf{w}_{jk}$ can be expressed as the linear combination

$$\mathbf{w}_{jk} = \gamma_k \mathbf{h}_{jk} + \sum_{\substack{\bar{k}=1 \\ \bar{k} \neq k}}^{K_r} \gamma_{\bar{k}} \mathbf{\Pi}^{\perp}_{\mathbf{h}_{j\bar{k}}} \mathbf{h}_{jk} + \sum_{l=\mathrm{rank}(\mathcal{A}_{jk})+1}^{N_t} \gamma_l \mathbf{v}_l \tag{29}$$

for some complex-valued coefficients γ_k and some orthogonal basis $\{\mathbf{v}_l\}_{l=\mathrm{rank}(\mathcal{A}_{jk})+1}^{N_t}$ for the orthogonal complement to $\mathcal{A}_{jk}$. Now, observe that

$$\mathbf{h}_{j\bar{k}} = \frac{\|\mathbf{h}_{j\bar{k}}\|}{\|\mathbf{\Pi}_{\mathbf{h}_{j\bar{k}}} \mathbf{h}_{jk}\|} \left(\mathbf{h}_{jk} - \mathbf{\Pi}^{\perp}_{\mathbf{h}_{j\bar{k}}} \mathbf{h}_{jk} \right)$$

for all $\bar{k}$ that are non-orthogonal to $\mathbf{h}_{jk}$ (while orthogonal channels can be removed from $\mathcal{A}_{jk}$, since these directions only create interference). Thus, $\mathbf{h}_{j\bar{k}}^H \mathbf{v}_l = 0$ for $\bar{k} = 1, \ldots, K_r$ and $l = \mathrm{rank}(\mathcal{A}_{jk}) + 1, \ldots, N_t$. Since $\mathbf{w}_{jk}$ only appears in the SINR expression in (3) as inner products with $\mathbf{h}_{jk}$ and $\mathbf{h}_{j\bar{k}}$, the identical rate tuple is achieved by the beamforming vector

$$\begin{aligned} \widetilde{\mathbf{w}}_{jk} = & \frac{\gamma_k}{\sqrt{1 - \sum_{l=\mathrm{rank}(\mathcal{A}_{jk})+1}^{N_t} |\gamma_l|^2}} \mathbf{h}_{jk} \\ & + \sum_{\substack{\bar{k}=1 \\ \bar{k} \neq k}}^{K_r} \frac{\gamma_{\bar{k}}}{\sqrt{1 - \sum_{l=\mathrm{rank}(\mathcal{A}_{jk})+1}^{N_t} |\gamma_l|^2}} \mathbf{\Pi}^{\perp}_{\mathbf{h}_{j\bar{k}}} \mathbf{h}_{jk} \end{aligned} \tag{30}$$

and transmit power $\tilde{p}_{jk} = p_{jk}(1 - \sum_{l=\mathrm{rank}(\mathcal{A}_{jk})+1}^{N_t} |\gamma_l|^2) \leq p_{jk}$. Thus, we have proved that all rate tuples on the Pareto boundary can be achieved by beamforming vectors $\widetilde{\mathbf{w}}_{jk} \in \mathrm{span}(\mathcal{A}_{jk})$.

Next, we will show that if $\mathbf{h}_{jk} \notin \mathrm{span}(\bigcup_{\bar{k} \neq k} \{\mathbf{h}_{j\bar{k}}\})$ for some j, k, then BS_j needs to use full power to reach the Pareto boundary. The given property corresponds to that $\sum_{l=1}^{m_{jk}} \mathbf{\Pi}_{\mathbf{e}_{jk}^{(l)}} \mathbf{h}_{jk} \neq \mathbf{h}_{jk}$, where $\mathbf{e}_{jk}^{(1)}, \ldots, \mathbf{e}_{jk}^{(m_{jk})}$ is an orthogonal basis of $\mathrm{span}(\bigcup_{\bar{k} \neq k} \{\mathbf{h}_{j\bar{k}}\})$. Consequently, there should exist a zero-forcing vector $\mathbf{u} = (\mathbf{I} - \sum_{l=1}^{m_{jk}} \mathbf{\Pi}_{\mathbf{e}_{jk}^{(l)}}) \mathbf{h}_{jk} \neq \mathbf{0}$ that satisfies $\mathbf{h}_{j\bar{k}}^H \mathbf{u} = 0$ for all $\bar{k} \neq k$.

Now, assume for the purpose of contradiction that the Pareto boundary is attained for a set of beamforming vectors $\{\mathbf{w}_{jk}\}$ and power allocations $\{p_{jk}\}$ that fulfills $\sum_{k=1}^{K_r} p_{jk} < P_j$. Then, we can replace $\mathbf{w}_{jk}$ and p_{jk} by

$$\begin{aligned} p_{jk}^{\text{new}} &= P_j - \sum_{\substack{\bar{k}=1 \\ \bar{k} \neq k}}^{K_r} p_{j\bar{k}} \quad \text{and} \\ \mathbf{w}_{jk}^{\text{new}} &= \sqrt{\frac{p_{jk}}{p_{jk}^{\text{new}}}} \mathbf{w}_{jk} + \alpha \mathbf{u} e^{i \arg(\mathbf{h}_{jk}^H \mathbf{w}_{jk})} \end{aligned} \tag{31}$$

for some positive parameter α that makes $\|\mathbf{w}_{jk}^{\text{new}}\| = 1$. This corresponds to increasing the power in the zero-forcing direction and making sure that the signal powers add up constructively at the intended

receiver. Thus, BS_j can increase the signal power at MS_k as $p_{jk}^{\text{new}}|\mathbf{h}_{jk}^H\mathbf{w}_{jk}^{\text{new}}|^2 = p_{jk}(|\mathbf{h}_{jk}^H\mathbf{w}_{jk}| + \alpha\mathbf{h}_{jk}^H\mathbf{u})^2 > p_{jk}|\mathbf{h}_{jk}^H\mathbf{w}_{jk}|^2$, without affecting the co-terminal interference. In other words, R_k has been increased without affecting $R_{\bar{k}}$ for $\bar{k} \neq k$, which is a contradiction to the assumption that the initial rate tuple belonged to the Pareto boundary. Thus, full transmit power is required to attain the Pareto boundary. The condition in (8) also becomes a necessary condition, because otherwise we can decrease the power by the approach in the first part of the proof and then increased it again using (31).

Proof of Theorem 3

Consider an expected rate tuple $(\mathbb{E}\{R_1\}, \ldots, \mathbb{E}\{R_{K_r}\})$ on the Pareto boundary that is achieved by beamforming vectors $\mathbf{w}_{jk}$ and power allocation p_{jk} for all j, k. The beamformers $\mathbf{w}_{jk}$ can in general be expressed as the linear combination

$$\mathbf{w}_{jk} = \sum_{l=1}^{m} \gamma_l \mathbf{v}_l + \sum_{l=m+1}^{N_t} \gamma_l \mathbf{u}_l \tag{32}$$

where $\{\mathbf{v}_l\}_{l=1}^m$ is an orthogonal basis of the m-dimensional given by $\text{span}(\bigcup_{k=1}^{K_r}\{\mathbf{U}_{jk}^{(D)}\})$ and $\{\mathbf{u}_l\}_{l=m+1}^{N_t}$ is an orthogonal basis of the orthogonal complement. The coefficients γ_l are complex-valued and fulfill $\sum_{l=1}^{N_t}|\gamma_l|^2 = 1$, since $\mathbf{w}_{jk}$ is expanded in terms of an orthonormal basis. To avoid allocating power to the weak eigenvalues in the orthogonal complement, we can replace $\mathbf{w}_{jk}$ by

$$\widetilde{\mathbf{w}}_{jk} = \frac{1}{\sqrt{\sum_{l=1}^{m}|\gamma_l|^2}} \sum_{l=1}^{m} \gamma_l \mathbf{v}_l \tag{33}$$

and reduce the transmit power to $\tilde{p}_{jk} = p_{jk}\sum_{l=1}^{m}|\gamma_l|^2 \leq p_{jk}$. This new precoding satisfy (11) and will achieve a new rate tuple $(\mathbb{E}\{\widetilde{R}_1\}, \ldots, \mathbb{E}\{\widetilde{R}_{K_r}\})$. Next, we show that the difference in performance is bounded by $o(\epsilon_k)$. With the new precoding, the change in the covariance matrix $\mathbf{S}_k$ in (4) is limited since $\mathbf{S}_k = \sum_{j=1}^{K_t}\widetilde{\mathbf{W}}_j^H\mathbf{Q}_{jk}\widetilde{\mathbf{W}}_j + \mathbf{E}$, where $\tilde{\mathbf{W}}_j = [\sqrt{\tilde{p}_{j1}}\widetilde{\mathbf{w}}_{j1} \ldots \sqrt{\tilde{p}_{jK_r}}\widetilde{\mathbf{w}}_{jK_r}]$ and the elements of the symmetric perturbation matrix $\mathbf{E}$ are bounded as $o(\epsilon_k)$. By applying the eigenvalue perturbation result in [31, Theorem 7.2.2] when deriving $\mathbb{E}\{\widetilde{R}_k\}$ in (5), the eigenvalues μ_m and λ_m can be replaced by $\tilde{\mu}_m = \mu_m + o(\epsilon_k)$ and $\tilde{\lambda}_m = \lambda_m + o(\epsilon_k)$, respectively. Observe that each term in (5) has the structure

$$\frac{e^{\frac{\sigma_k^2}{\mu_m+o(\epsilon_k)}}E_1\left(\frac{\sigma_k^2}{\mu_m+o(\epsilon_k)}\right)}{\log(2)\prod_{l\neq m}(1-\frac{\mu_l+o(\epsilon_k)}{\mu_m+o(\epsilon_k)})} = \frac{e^{\frac{\sigma_k^2}{\mu_m}}E_1\left(\frac{\sigma_k^2}{\mu_m}\right)}{\log(2)\prod_{l\neq m}(1-\frac{\mu_l}{\mu_m})} + o(\epsilon_k) \tag{34}$$

where the equality follows from straightforward appliance of l'Hospital's rule. Thus, by applying this result to each term in (5), we achieve $\mathbb{E}\{R_k\} = \mathbb{E}\{\widetilde{R}_k\} + o(\epsilon_k)$. To finalize the proof of the first part, observe that for arbitrary covariance matrices it holds with probability one that $\text{span}(\mathbf{\Pi}_{\mathbf{U}_{j\bar{k}}^{(D)}}\mathbf{U}_{jk}^{(D)}) = \text{span}(\mathbf{U}_{j\bar{k}}^{(D)})$ for all $\bar{k}$. Since,

$$\mathbf{\Pi}_{\mathbf{U}_{j\bar{k}}^{(D)}}\mathbf{U}_{jk}^{(D)} = (\mathbf{I} - \mathbf{\Pi}_{\mathbf{U}_{j\bar{k}}^{(D)}}^{\perp})\mathbf{U}_{jk}^{(D)}$$

it follows that

$$\mathrm{span}(\bigcup_{k=1}^{K_r}\{\mathbf{U}_{jk}^{(D)}\}) = \mathrm{span}(\{\mathbf{U}_{jk}^{(D)}\}\bigcup_{\bar{k}\neq k}\{\mathbf{\Pi}_{\mathbf{U}_{j\bar{k}}^{(D)}}^{\perp}\mathbf{U}_{jk}^{(D)}\}).$$

Finally, consider the case when $\mathrm{span}(\bigcup_{k=1}^{K_r}\{\mathbf{U}_{jk}^{(D)}\}) \neq \mathbb{C}^{N_t}$ for some j. If $\sum_k p_{jk} < P_j$, we propose the following way of increasing the power usage while guaranteeing the same type performance. We assume that the beamforming vector $\mathbf{w}_{jk}$ fulfills (11), otherwise we can follow the approach in first part of the proof to decrease the power usage, retain the performance, and fulfill (11). Select a unit vector $\mathbf{d} \notin \mathrm{span}(\bigcup_{k=1}^{K_r}\{\mathbf{U}_{jk}^{(D)}\})$. If we replace the beamformer and the power allocation with

$$\begin{aligned} p_{jk}^{\text{new}} &= P_j - \sum_{\bar{k}=1,\bar{k}\neq k}^{K_r} p_{j\bar{k}} \quad \text{and} \\ \mathbf{w}_{jk}^{\text{new}} &= \sqrt{\frac{p_{jk}}{p_{jk}^{\text{new}}}}\mathbf{w}_{jk} + \sqrt{1-\frac{p_{jk}}{p_{jk}^{\text{new}}}}\mathbf{d}, \end{aligned} \tag{35}$$

the difference in signal and interference variance yields a perturbation in $\mathbf{S}_k$ on the order of $o(\epsilon_k)$ and we can use the approach above to show that resulting rate tuple fulfills $\mathbb{E}\{R_k\} = \mathbb{E}\{\widetilde{R}_k\} + o(\epsilon_k)$. Hence, full transmit power can be used to achieve $(\mathbb{E}\{\widetilde{R}_1\},\ldots,\mathbb{E}\{\widetilde{R}_{K_r}\})$.

Proof of Theorem 4

Let $R_k^{\text{opt}} = \log_2(1+\mathrm{SINR}_k^{\text{opt}})$ represent an arbitrary Pareto optimal rate tuple. Observe that this rate tuple is achieved by solving

$$\begin{aligned} \underset{\mathbf{w}_{jk}\in\mathbb{C}^{N_t},p_{jk}\geq 0\ \forall j,k}{\text{maximize}} \quad & \min_{1\leq k\leq K_r} \frac{R_k}{R_k^{\text{opt}}} \\ \text{subject to} \quad & \sum_{k=1}^{K_r} p_{jk} \leq P_j, \ \ \|\mathbf{w}_{jk}\|^2 = 1 \text{ for all } \ k \end{aligned} \tag{36}$$

since [19, Lemma 1] shows that all solutions to (36) must satisfy $R_1/R_1^{\text{opt}} = \ldots = R_{K_r}/R_{K_r}^{\text{opt}}$. Thus, with $R_1^{\text{opt}},\ldots,R_{K_r}^{\text{opt}}$ as rate constraints for the different terminals, the uplink-downlink duality result of [19, Theorem 1] can be applied. This means that the optimal beamforming vectors for the downlink problem in (36) should also maximize the virtual uplink SINRs in (14) for each user, and the parameters β_{jk} represents the optimal power allocation in the virtual dual uplink.

Proof of Theorem 5

Let the arbitrary power allocation be denoted $p_{jk} = P_j\tilde{p}_{jk}$ with (normalized) coefficients $\tilde{p}_{jk}$. As $P = \min_j P_j$, we let $P_j = P\alpha_j$ for some parameters $0 < \alpha_j < 1$ for all j. Using the SINR expression

in (3) and the given power allocation, the sum rate becomes

$$\sum_{k=1}^{K_r} \log_2 \left(1 + \frac{\left| \sum_{j=1}^{K_t} \sqrt{P\alpha_j \tilde{p}_{jk}} \mathbf{h}_{jk}^H \mathbf{w}_{jk}^{(\text{DVSINR})} \right|^2}{\sum_{\substack{\bar{k}=1 \\ \bar{k}\neq k}}^{K_r} \left| \sum_{j=1}^{K_t} \sqrt{P\alpha_j \tilde{p}_{j\bar{k}}} \mathbf{h}_{jk}^H \mathbf{w}_{j\bar{k}}^{(\text{DVSINR})} \right|^2 + \sigma_k^2} \right)$$

$$= K_r \log_2(P) \tag{37}$$

$$+ \sum_{k=1}^{K_r} \log_2 \left(\frac{1}{P} + \frac{\left| \sum_{j=1}^{K_t} \sqrt{\alpha_j \tilde{p}_{jk}} \mathbf{h}_{jk}^H \mathbf{w}_{jk}^{(\text{DVSINR})} \right|^2}{\sum_{\substack{\bar{k}=1 \\ \bar{k}\neq k}}^{K_r} \left| \sum_{j=1}^{K_t} \sqrt{P\alpha_j \tilde{p}_{j\bar{k}}} \mathbf{h}_{jk}^H \mathbf{w}_{j\bar{k}}^{(\text{DVSINR})} \right|^2 + \sigma_k^2} \right).$$

If $N_t \geq K_r$, then with probability $\mathbf{h}_{jk} \notin \text{span}(\bigcup_{\bar{k}\neq k} \{\mathbf{h}_{j\bar{k}}\})$ for all j, k. By analyzing the expression $\mathbf{w}_{j\bar{k}}^{(\text{DVSINR})} = \mathbf{C}_{j\bar{k}}^{-1} \mathbf{h}_{j\bar{k}} / \|\mathbf{C}_{j\bar{k}}^{-1} \mathbf{h}_{j\bar{k}}\|$ it is straightforward to show that $\sqrt{P} \mathbf{h}_{jk}^H \mathbf{w}_{j\bar{k}}^{(\text{DVSINR})} \to 0$ as $P \to \infty$. Thus, the last term of (37) is bounded as $P \to \infty$ and the multiplexing gain is K_r.

REFERENCES

[1] D. Gesbert, M. Kountouris, R. Heath, C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, 2007.

[2] S. Shamai and B. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE VTC'01-Spring*, vol. 3, 2001, pp. 1745–1749.

[3] H. Zhang and H. Dai, "Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks," *EURASIP J. Wirel. Commun. Netw.*, vol. 2, pp. 222–235, 2004.

[4] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun. Mag.*, vol. 13, no. 4, pp. 56–61, 2006.

[5] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network MIMO: Overcoming intercell interference in indoor wireless systems," in *Proc. IEEE ACSSC'07*, 2007, pp. 83–87.

[6] P. Viswanath and D. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.

[7] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 361–374, 2006.

[8] P. Marsch and G. Fettweis, "On multicell cooperative transmission in backhaul-constrained cellular systems," *Ann. Telecommun.*, vol. 63, pp. 253–269, 2008.

[9] S. Jing, D. Tse, J. Soriaga, J. Hou, J. Smee, and R. Padovani, "Multicell downlink capacity with coordinated processing," *EURASIP J. Wirel. Commun. Netw.*, 2008.

[10] B. L. Ng, J. Evans, S. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, 2008.

[11] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, "Downlink multicell processing with limited-backhaul capacity," *EURASIP J. on Adv. in Signal Process.*, 2009.

[12] M. Kobayashi, M. Debbah, and J. Belfiore, "Outage efficient strategies in network MIMO with partial CSIT," in *Proc. IEEE ISIT'09*, 2009, pp. 249–253.

[13] E. Larsson and E. Jorswieck, "Competition versus cooperation on the MISO interference channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, 2008.

[14] E. Jorswieck, E. Larsson, and D. Danev, "Complete characterization of the Pareto boundary for the MISO interference channel," *IEEE Trans. Signal Process.*, vol. 56, no. 10, 2008.

[15] N. Hassanpour, J. Smee, J. Hou, and J. Soriaga, "Distributed beamforming based on signal-to-caused-interference ratio," in *Proc. IEEE ISSSTA'08*, 2008, pp. 405–410.

[16] B. O. Lee, H. W. Je, I. Sohn, O.-S. Shin, and K. B. Lee, "Interference-aware decentralized precoding for multicell MIMO TDD systems," in *Proc. IEEE GLOBECOM'08*, 2008.

[17] R. Zakhour and D. Gesbert, "Coordination on the MISO interference channel using the virtual SINR framework," in *Proc. International ITG Workshop on Smart Antennas*, 2009.

[18] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Distributed multicell and multiantenna precoding: Characterization and performance evaluation," in *Proc. IEEE GLOBECOM'09*, 2009.

[19] M. Schubert and H. Boche, "Solution of the multiuser downlink beamforming problem with individual SINR constraints," *IEEE Trans. Veh. Technol.*, vol. 53, no. 1, pp. 18–28, 2004.

[20] E. Björnson and B. Ottersten, "On the principles of multicell precoding with centralized and distributed cooperation," in *Proc. WCSP'09*, 2009.

[21] R. Zakhour and D. Gesbert, "Distributed multicell-MIMO precoding using the layered virtual SINR framework," in *IEEE Trans. Wireless Commun.*, submitted.

[22] E. Björnson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, 2010.

[23] E. Björnson, D. Hammarwall, and B. Ottersten, "Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 4027–4041, 2009.

[24] E. Björnson and B. Ottersten, "Post-user-selection quantization and estimation of correlated Frobenius and spectral channel norms," in *Proc. IEEE PIMRC'08*, 2008.

[25] J. Lindblom, E. Karipidis, and E. Larsson, "Selfishness and altruism on the MISO interference channel: The case of partial transmitter CSI," *IEEE Commun. Lett.*, vol. 13, no. 9, pp. 667–669, 2009.

[26] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, "Multiple-input-multiple-output measurements and modeling in Manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 321–331, 2003.

[27] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Modeling of wide-band MIMO radio channels based on NLoS indoor measurements," *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 655–665, 2004.

[28] D. Hammarwall, M. Bengtsson, and B. Ottersten, "Utilizing the spatial information provided by channel norm feedback in SDMA systems," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3278–3293, 2008.

[29] H. Zhang, N. Mehta, A. Molisch, J. Zhang, and H. Dai, "Asynchronous interference mitigation in cooperative base station systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 155–165, 2008.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[31] G. Golub and C. V. Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.

PLACE PHOTO HERE

Emil Björnson (S07) was born in Malmö, Sweden, in 1983. He received the M.S. degree in engineering mathematics from Lund University, Lund, Sweden, in 2007. He is currently working towards the Ph.D. degree in telecommunications at the Signal Processing Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden.

His research interests include wireless communications, resource allocation, estimation theory, stochastic signal processing, and mathematical optimization.

For his work on multicell MIMO communications, he received a Best Paper Award at the 2009 International Conference on Wireless Communications and Signal Processing (WCSP 2009).

PLACE PHOTO HERE

Randa Zakhour (S'00) received her B.E. in Computer and Communications Engineering degree from the American University of Beirut, Lebanon in 2002. In June 2006, she received her M.Sc. in Communications engineering from the RWTH Aachen, Germany. She obtained her Ph.D. degree in Electronics and Communications from Telecom ParisTech inApril 2010: she carried out her Ph.D. research work at the Mobile Communications Department at EURECOM, Sophia Antipolis.

Her current research interests include signal processing, optimization, and feedback and resource allocation issues in multicell multi-user MIMO systems.

PLACE PHOTO HERE

David Gesbert (S'96-M'99-SM'05) is Professor in the Mobile Communications Dept., EURECOM, France, where he heads the Communications Theory Group. He obtained the Ph.D degree from Ecole Nationale Superieure des Telecommunications, France, in 1997. From 1997 to 1999 he has been with the Information Systems Laboratory, Stanford University. In 1999, he was a founding engineer of Iospan Wireless Inc, San Jose, Ca., a startup company pioneering MIMO-OFDM (now Intel). Between 2001 and 2003 he has been with the Department of Informatics, University of Oslo as an adjunct professor. D. Gesbert has published about 170 papers and several patents all in the area of signal processing, communications, and wireless networks.

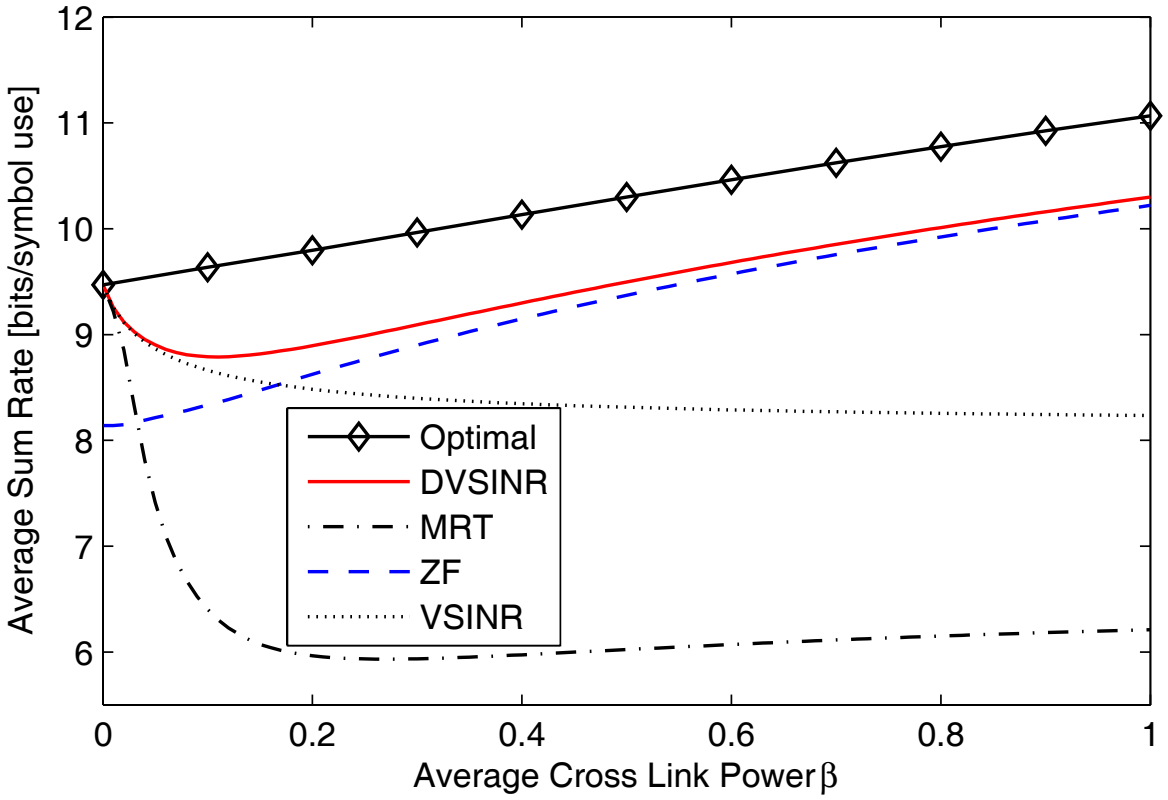
D. Gesbert was a co-editor of several special issues on wireless networks and communications theory, for JSAC (2003, 2007, 2010), EURASIP Journal on Applied Signal Processing (2004, 2007), Wireless Communications Magazine (2006). He served on the IEEE Signal Processing for Communications Technical Committee, 2003-2008. He's an associate editor for IEEE Transactions on Wireless Communications and the EURASIP Journal on Wireless Communications and Networking. He authored or co-authored papers winning the 2004 IEEE Best Tutorial Paper Award (Communications Society) for a 2003 JSAC paper on MIMO systems, 2005 Best Paper (Young Author) Award for Signal Proc. Society journals, and the Best Paper Award for the 2004 ACM MSWiM workshop. He co-authored the book Space time wireless communications: From parameter estimation to MIMO systems, Cambridge Press, 2006.

PLACE
PHOTO
HERE

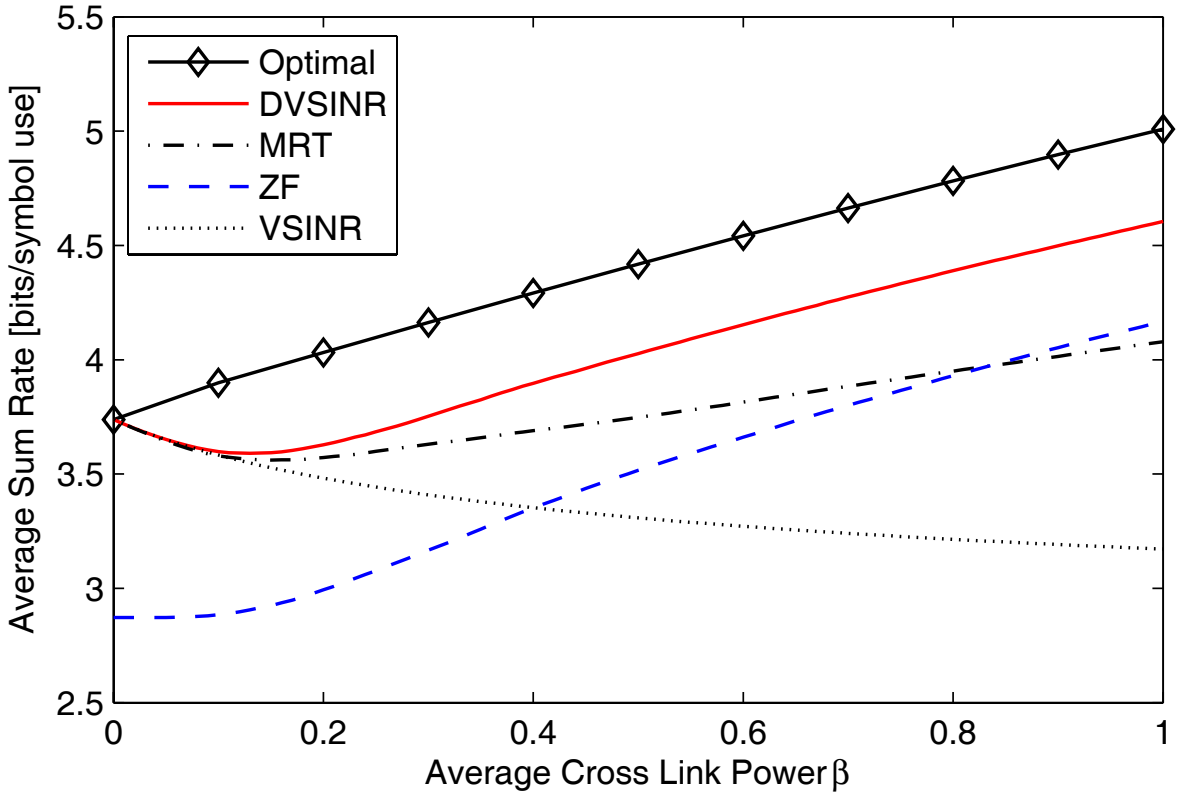
Björn Ottersten (S'87-M'89-SM'99-F'04) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linköping University, Linköping, Sweden, in 1986 and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1989.

He has held research positions at the Department of Electrical Engineering, Linköping University; the Information Systems Laboratory, Stanford University; and the Katholieke Universiteit Leuven, Leuven, Belgium. During 19961997, he was Director of Research at ArrayComm Inc., San Jose, CA, a start-up company based on Otterstens patented technology. In 1991, he was appointed Professor of Signal Processing at the Royal Institute of Technology (KTH), Stockholm, Sweden. From 2004 to 2008, he was Dean of the School of Electrical Engineering at KTH, and from 1992 to 2004 he was head of the Department for Signals, Sensors, and Systems at KTH. He is also Director of security and trust at the University of Luxembourg. His research interests include wireless communications, stochastic signal processing, sensor array processing, and time-series analysis.

Dr. Ottersten has coauthored papers that received an IEEE Signal Processing Society Best Paper Award in 1993, 2001, and 2006. He has served as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and on the Editorial Board of the IEEE Signal Processing Magazine. He is currently Editor-in-Chief of the EURASIP Signal Processing Journal and a member of the Editorial Board of the EURASIP Journal of Advances in Signal Processing. He is a Fellow of EURASIP. He is one of the first recipients of the European Research Council advanced research grant.

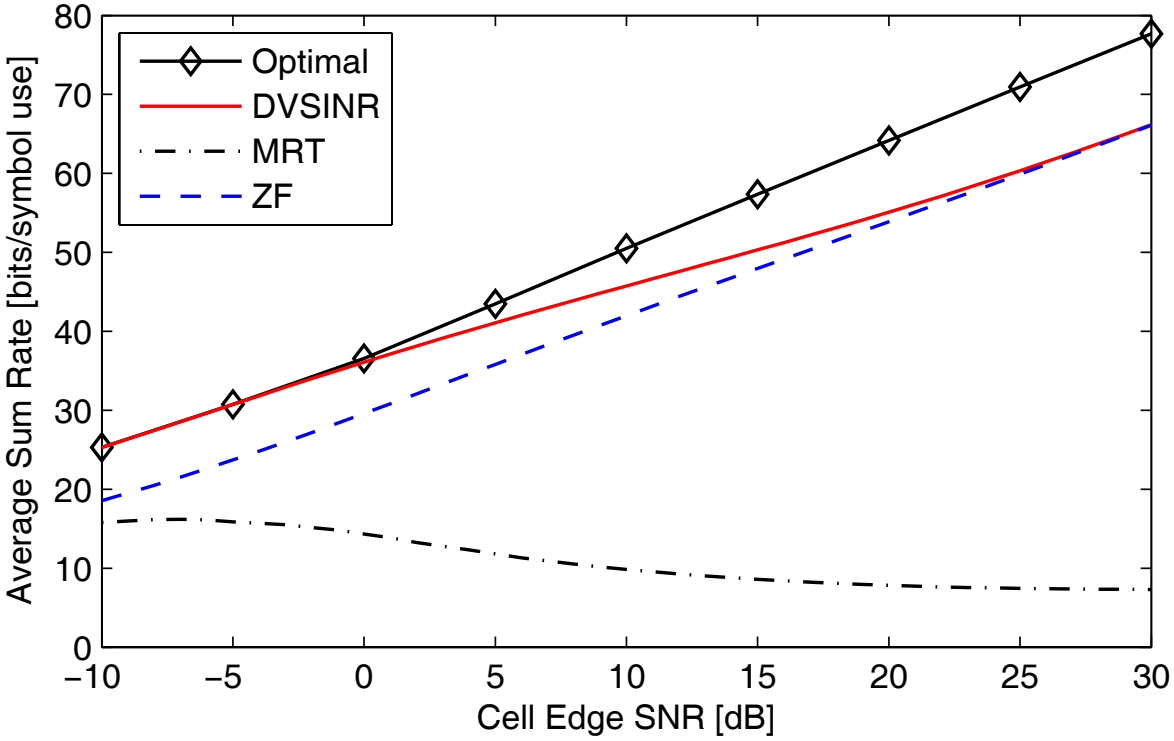


(a) SNR 10 dB.

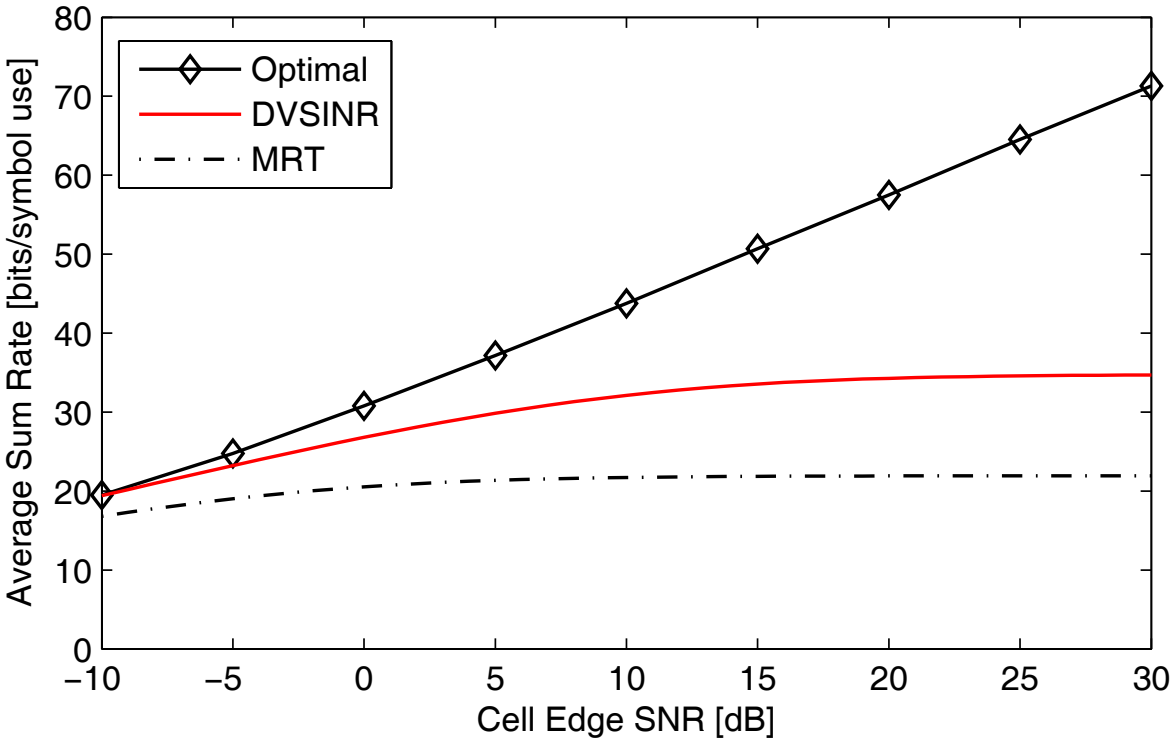


(b) SNR 0 dB.

Fig. 3. Average sum rate (over channel realizations) in a system with local instantaneous CSI, $K_t = K_r = 2$, $N_t = 3$, and a varying average cross link power: $\mathbf{Q}_{11} = \mathbf{Q}_{22} = \mathbf{I}$, $\mathbf{Q}_{12} = \mathbf{Q}_{21} = \beta\mathbf{I}$.



(a) $N_t = 4$.



(b) $N_t = 2$.

Fig. 4. Average sum rate (over terminal locations and channel realizations) in a system with local instantaneous CSI and $K_t = K_r = 4$. The scenario considers uniformly distributed terminals within a square with base stations in each corner.

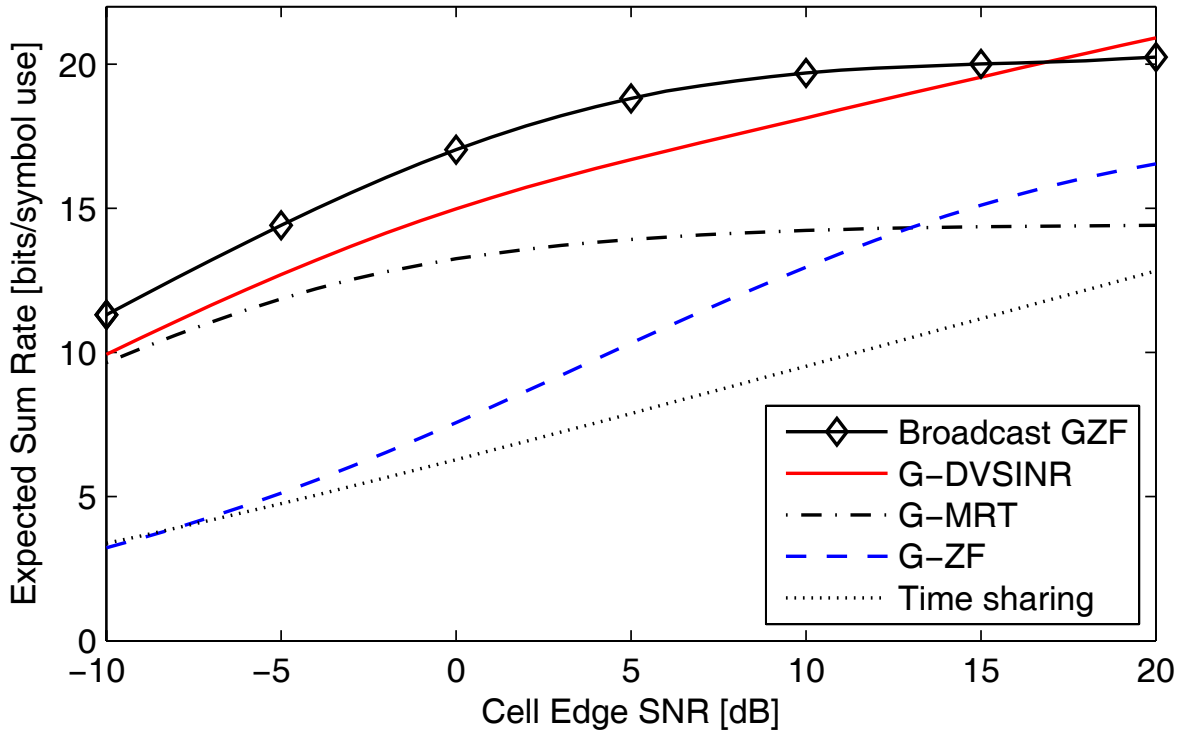


Fig. 5. Expected sum rate (over terminal locations and channel statistics) in a system with local statistical CSI, $K_t = K_r = 4$, $N_t = 6$, and an angular spread of 10 degrees (as seen from each base station). The scenario considers uniformly distributed terminals within a square with base stations in each corner.