Large System Analysis of Sum Capacity in the Gaussian MIMO Broadcast Channel

Christian Guthy, Wolfgang Utschick, *Senior Member, IEEE*, and Michael L. Honig, *Fellow, IEEE*

***Abstract*—We analyze the achievable sum rate of the Gaussian MIMO broadcast channel. We first consider Multiple-Input Single-Output (MISO) channels and derive the large system limit of the sum capacity as the number of users and transmit antennas go to infinity with a fixed ratio. We then consider Multiple-Input Multiple-Output (MIMO) broadcast channels and fix the number of users and let the number of transmit and receive antennas tend to infinity with fixed ratio. As in this case an asymptotic expression for sum capacity is hard to obtain, we evaluate the large system sum rate corresponding to successive zero-forcing beamforming with Dirty-Paper Coding. The analysis gives a lower bound on the large system sum capacity, which is numerically observed to be quite close. In addition, large system analysis is applied to estimate the relatively small performance losses with respect to sum capacity of successive zero-forcing beamforming with and without Dirty-Paper Coding in finite MISO systems.**

***Index Terms*—Broadcast Channels, Large system analysis, Multiple-Input Multiple-Output (MIMO) systems.**

I. Introduction

To achieve the sum capacity of the Multiple-Input Multiple Output (MIMO) broadcast channel, numerically complex iterative methods are needed to determine the optimum transmit covariance matrices [1], [2], [3]. This has motivated a wide variety of algorithms that aim to approximate the sum capacity closely with reduced computational complexity. For example, one approach is to decompose the MIMO broadcast channel into a system of scalar interference free subchannels, which are allocated across users [4]. Interference can be suppressed through a combination of spatial zero-forcing and/or Dirty Paper Coding (DPC) [5].

In addition to simplifying the computation of optimal covariance matrices, approximate algorithms can also facilitate performance analysis. Here we analyze the performance of the Successive Encoding Successive Allocation Method (SESAM) proposed in [6]. This method assigns beams sequentially according to a greedy criterion, and eliminates interference through both zero-forcing and DPC. A key feature is that it is *non-iterative*, meaning that the spatial beams are successively computed one at a time. A similar decomposition for Multiple-Input Single-Output (MISO) systems into scalar subchannels has been proposed in [4], and the greedy allocation stems from [7]. These methods have been observed to be very efficient in terms of reducing numerical complexity with only marginal performance degradation.

Our goal in this paper is to derive an analytical expression that accurately estimates the average sum capacity of the MIMO Gaussian broadcast fading channel with *i.i.d.* Rayleigh fading. This is quite challenging for finite-size systems (i.e., finite number of users and antennas), hence we instead consider large system performance limits in which the number of users and/or antennas tend to infinity. Evaluating a large system limit of this sum capacity, which accurately estimates the sum capacity of a finite system, still appears to be quite challenging. Thus, we first consider the special case of non-cooperating receive antennas i.e., a MISO system, and present an exact analytical expression for the sum capacity in the limit as the number of transmit antennas and users each tend to infinity with fixed ratio. This expression only serves as a weak lower capacity bound for MIMO channels (in which case some receivers cooperate). We therefore proceed to derive the sum capacity of SESAM (with MIMO channels) in the large system limit in which the number of users is fixed, and the number of transmit and receive antennas each tend to infinity with fixed ratio. This serves as a tight lower bound for the asymptotic sum capacity of the MIMO broadcast channel.

We also apply the preceding results to the MISO broadcast channel, and quantify the loss in sum capacity incurred when beams are assigned successively, as proposed in [7] and [8], relative to the optimal beams. Our results show that this performance loss is relatively minor. Finally, numerical results show that the asymptotic results for both the MIMO and MISO successive assignment schemes give accurate estimates of the performance of finite-size systems.

Other suboptimal methods in which zero-forcing beamforming is used alone to avoid the high complexity associated with DPC are proposed in [8], [9].[1] The performance of those algorithms has been demonstrated primarily by simulation results. Asymptotic results as the number of users becomes infinite are also presented in [7], [8] and [9]. However, those results cannot be used to estimate the performance of a finite-size system. This is especially true for the algorithms without DPC in [8] and [9], which are optimum in the limit of infinitely

Manuscript received February 1, 2012, accepted August 29, 2012. This work has been supported by Deutsche Forschungsgemeinschaft (DFG) with the funded project Ut36/8–1 within the DFG focus program 1163 and by the US National Science Foundation under grant CCF-1018578/001.

Christian Guthy and Wolfgang Utschick are with the Fachgebiet Methoden der Signalverarbeitung, Technische Universität München, Theresienstr. 90, D-80290 München, Germany, e-mail: {guthy, utschick}@tum.de. Michael L. Honig is with Northwestern University, Evanston, IL, USA, e-mail: mh@eecs.northwestern.edu

[1] Those algorithms have been developed for MISO systems and can be extended to MIMO systems by considering each eigenmode of the channel matrices as a "virtual" MISO user. A more advanced approach including the receive filters in the successive optimization can be found in [10].

many users, but show substantial performance degradation compared to the optimum, when the number of users is finite.

Related work in which large system analysis is used to estimate the performance of different system models is given in [11] and [12]. There the ergodic performance of MIMO systems is evaluated in the limit of infinitely many transmit and receive antennas given statistical channel knowledge at the transmitter. In [13] and in [14] MISO systems with infinitely many transmit antennas and users are analyzed, and asymptotically optimum parameters are applied to finite systems, giving near-optimum performance results. Large system analysis is applied to multi-cell MISO systems in [15], [16], [17], and [18], where the number of antennas at the base stations and the number of users tend towards infinity.

In the next section we explain the system model and sum capacity objective. The large system limit of sum capacity for the MISO broadcast channel is derived in Section III, and the asymptotic sum capacity of SESAM for the MIMO broadcast channel is presented in Section IV. The analysis of successive beam assignment for MISO channels is carried out in Section V. Numerical results are shown in SectionVI, and the paper concludes with Section VII.

Notation: Bold lower and uppercase letters denote vectors and matrices, respectively. $(\bullet)^{\mathrm{T}}$ and $(\bullet)^{\mathrm{H}}$ describe the transpose and the Hermitian of a matrix, respectively. $\rho_i(\boldsymbol{A})$, $\mathrm{tr}(\boldsymbol{A})$, and $|\boldsymbol{A}|$ are the ith eigenvalue, the trace, and the determinant of the matrix $\boldsymbol{A}$, respectively, and $\boldsymbol{A} \succeq 0$ implies that $\boldsymbol{A}$ is positive semi-definite. $\mathrm{null}\{\boldsymbol{A}\}$ and $\mathrm{span}\{\boldsymbol{A}\}$ denote the null-space and the range of the matrix $\boldsymbol{A}$, respectively. $\mathrm{diag}(a_1, \ldots, a_i)$ denotes a diagonal matrix with the elements $a_1, \ldots, a_i$ on its diagonal and $\mathrm{blockdiag}(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_i)$ is a block-diagonal matrix. $\boldsymbol{I}_i$ and $\boldsymbol{0}_i$ are $i \times i$ identity matrix and $i \times i$ zero matrix,respectively. $\boldsymbol{e}_j$ denotes the j-th canonical unit vector and $\delta(x)$ is the Dirac function.

II. System Model and Problem Formulation

We consider a multiuser MIMO system with a base station having N_{T} antennas and K users, each equipped with N_{R} antennas. Gaussian codebooks are used at the transmitter, the channel matrix between the base station and user k is denoted as $\hat{\boldsymbol{H}}_k \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{T}}}$ and Rayleigh fading is considered. This implies that all entries in the channel matrices $\hat{\boldsymbol{H}}_k$ follow a circularly symmetric Gaussian distribution with zero mean. The variance is set to one for all users and it is furthermore assumed that all entries in the channel matrices are independent. The transmit power at the base station must not exceed the limit P_{Tx}, perfect Channel State Information (CSI) is assumed at the transmitter, and the additive white Gaussian noise has zero mean and variance one.
In this paper we consider the large system limit of number of transmit and receive antennas simultaneously going to infinity at a finite fixed ratio α, i.e., $N_{\mathrm{T}} \to \infty$, $N_{\mathrm{R}} \to \infty$, $\alpha = \frac{N_{\mathrm{T}}}{N_{\mathrm{R}}}$. All other parameters such as transmit power P_{Tx} and number of users K remain finite. We first try to analyze the sum capacity in this limit and explain, why this task is not solved in this paper, but the special case of non-cooperating receive antennas is considered instead and additionally a lower bound for the asymptotic sum capacity with receive antenna cooperation is presented. Sum capacity can be computed via the duality between broadcast channel and the dual multiple access channel [19]. Denoting the transmit covariance matrix of user k in the dual multiple access channel as $\boldsymbol{W}_k \in \mathbb{C}^{N_{\mathrm{R}} \times N_{\mathrm{R}}}$, the sum capacity $R_{\text{sum cap}}$ results from the following optimization problem

$$\begin{aligned} R_{\text{sum cap}} &= \max_{\boldsymbol{W}} \log_2 \left| \boldsymbol{I}_{N_{\mathrm{T}}} + \hat{\boldsymbol{H}}^{\mathrm{H}} \boldsymbol{W} \hat{\boldsymbol{H}} \right|, \\ \text{s.t. } & \mathrm{tr}\left(\boldsymbol{W}\right) \leq P_{\mathrm{Tx}}, \quad \boldsymbol{W} \succeq 0, \quad \boldsymbol{W} \text{ blockdiag}, \end{aligned} \tag{1}$$

where $\hat{\boldsymbol{H}}^{\mathrm{H}} = \left[\hat{\boldsymbol{H}}_1^{\mathrm{H}}, \ldots, \hat{\boldsymbol{H}}_K^{\mathrm{H}}\right] \in \mathbb{C}^{N_{\mathrm{T}} \times K N_{\mathrm{R}}}$ and $\boldsymbol{W} = \mathrm{blockdiag}\left(\boldsymbol{W}_1, \ldots, \boldsymbol{W}_K\right)$. When the number of transmit and receive antennas both go to infinity, it has to be proven that $R_{\text{sum cap}}$ converges to an asymptotic limit independent of the channel realizations $\boldsymbol{H}$. This can be done by showing that the empirical eigenvalue distribution $F_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x)$, where $\hat{\boldsymbol{W}}$ denotes the optimum with respect to (1), becomes independent of the actual channel realizations $\hat{\boldsymbol{H}}_k$ in the large system limit. The empirical eigenvalue distribution states the fraction of eigenvalues of the matrix $\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}$ smaller or equal to x, i.e.,

$$\begin{aligned} F_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x) = \frac{1}{N_{\mathrm{T}}} \Bigg| \Bigg\{ \rho_i \left(\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}} \right), i = 1, \ldots, N_{\mathrm{T}} \mid \\ \rho_i \left(\hat{\boldsymbol{H}}_k^{\mathrm{H}} \hat{\boldsymbol{W}}_k \hat{\boldsymbol{H}}_k \right) \leq x \Bigg\} \Bigg|. \end{aligned} \tag{2}$$

In case it can be shown that $F_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x)$ converges to an asymptotic limit $F^{(\infty)}_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x, \alpha)$ independent of the channel realizations and only dependent on α, the sum capacity also converges to an asymptotic limit. It can be obtained via the asymptotic eigenvalue distribution (a.e.d.) $f^{(\infty)}_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x, \alpha)$ of the matrix $\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}$, which is the derivative of $F^{(\infty)}_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x, \alpha)$, i.e., $f^{(\infty)}_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x, \alpha) = \frac{\partial F^{(\infty)}_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x, \alpha)}{\partial x}$ of $F^{(\infty)}_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x, \alpha)$, according to

$$\begin{aligned} \frac{R_{\text{sum cap}}}{N_{\mathrm{T}}} \underset{\substack{N_{\mathrm{T}}, N_{\mathrm{R}} \to \infty \\ N_{\mathrm{T}} = \alpha N_{\mathrm{R}}}}{\longrightarrow} \int_{-\infty}^{\infty} \log_2\left(1 + x\right) f^{(\infty)}_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(x, \alpha) \,\mathrm{d}\, x = \\ \mathcal{V}_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(1). \end{aligned} \tag{3}$$

$\mathcal{V}_{\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}}(\gamma)$ is the Shannon transform of the matrix $\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}} \hat{\boldsymbol{H}}$ as defined in [20, Definition 2.12]. However, closed form solutions for (3) exist in case $\hat{\boldsymbol{W}}$ is independent of $\hat{\boldsymbol{H}}$ [20, Theorem 2.39] or $\hat{\boldsymbol{H}}^{\mathrm{H}} \hat{\boldsymbol{W}}^{\frac{1}{2}}$ contains i.i.d. entries [20, Theorem 2.39], but not for the problem at hand, where $\hat{\boldsymbol{W}}$ is a function of the channel matrices $\boldsymbol{H}_k$ via the optimization in (1), That is why an asymptotic expression for the sum capacity in MIMO systems is difficult to obtain. Thus, we will consider the special case of non-cooperative receive antennas in the next section and provide a lower bound for the large system sum capacity in the succeeding section.

III. Large System Sum Capacity with Non-cooperative Receive Antennas

When the antennas at the receivers cannot cooperate, i.e., when the signals from different receive antennas at the same user cannot be combined constructively to increase receive SNR, the transmit covariance matrices $\boldsymbol{W}_k$ from Problem (1) in the dual uplink channel are constrained to be diagonal. The following theorem is therefore especially useful in MISO systems with an infinite number of $K' = KN_\mathrm{R}$ users.

Theorem 1: The large system sum capacity $R_{\text{non coop}}$ in the MIMO broadcast channel with non-cooperative receivers is given by

$$\frac{R_{\text{non coop}}}{N_\mathrm{T}} \xrightarrow[\substack{N_\mathrm{T}, N_\mathrm{R}\to\infty,\\ N_\mathrm{T}=N_\mathrm{R}\alpha}]{} \left[\frac{K}{\alpha}\log_2\left(1+\frac{\alpha}{K}P_\mathrm{Tx}m_1\right)- \right. \\ \left. -\log_2(m_1)+(m_1-1)\log_2(\mathrm{e})\right], \quad (4)$$

where e is Euler's number and

$$m_1 = \frac{1}{2}\left(1-\frac{K}{\alpha}\right)-\frac{K}{2P_\mathrm{Tx}\alpha}+ \\ +\sqrt{\left(\frac{1}{2}\left(1-\frac{K}{\alpha}\right)-\frac{K}{2P_\mathrm{Tx}\alpha}\right)^2+\frac{K}{P_\mathrm{Tx}\alpha}}.$$

Proof: Considering the constraint of non-cooperative receivers, Problem (1) reads as

$$R_{\text{non coop}} = \max_{\boldsymbol{W}}\log_2\left|\boldsymbol{I}_{N_\mathrm{T}}+\hat{\boldsymbol{H}}^\mathrm{H}\boldsymbol{W}\hat{\boldsymbol{H}}\right|, \\ \text{s.t. } \operatorname{tr}(\boldsymbol{W})\le P_\mathrm{Tx}, \quad \boldsymbol{W}\succeq 0, \quad \boldsymbol{W} \text{ diagonal.} \quad (5)$$

In order to derive the large system sum capacity with non-cooperative receivers, the following lemma will be needed.

Lemma 1: For an infinite number of transmit and receive antennas, the sum capacity in the MIMO broadcast channel without receive antenna cooperation is achieved by equal power distribution, i.e., $\hat{\boldsymbol{W}}-\boldsymbol{I}_{N_\mathrm{R}K}\frac{P_\mathrm{Tx}}{N_\mathrm{R}K} \xrightarrow[\substack{N_\mathrm{T}, N_\mathrm{R}\to\infty,\\ N_\mathrm{T}=N_\mathrm{R}\alpha}]{} \boldsymbol{0}_{N_\mathrm{R}K}$, with $\hat{\boldsymbol{W}}$ denoting the solution achieving the optimum in (5).

Proof: see Appendix A ■

Lemma 1 suggests that the power allocated to each receive antenna (or user in MISO systems) tends to zero for an infinite number of receive antennas. This effect is however compensated by the fact that the elements in the channel matrices are assumed to have variance one, so that $\mathrm{E}\left[\boldsymbol{e}_j^\mathrm{T}\hat{\boldsymbol{H}}\hat{\boldsymbol{H}}^\mathrm{H}\boldsymbol{e}_j\right] = N_\mathrm{T}$, i.e., the channel gain at each receive antenna goes to infinity in this limit. When the channel matrices $\hat{\boldsymbol{H}}$ are normalized so that they consist of Gaussian i.i.d. entries with zero mean and variance $\frac{1}{N_\mathrm{T}}$ and the normalized channel matrices $\boldsymbol{H} := \frac{1}{\sqrt{N_\mathrm{T}}}\hat{\boldsymbol{H}}$ are used, the asymptotic sum rate reads as $R_{\text{non coop}} = \log_2\left|\boldsymbol{I}_{N_\mathrm{T}}+\frac{P_\mathrm{Tx}}{N_\mathrm{R}K}\hat{\boldsymbol{H}}^\mathrm{H}\hat{\boldsymbol{H}}\right| = \log_2\left|\boldsymbol{I}_{N_\mathrm{T}}+P_\mathrm{Tx}\frac{\alpha}{K}\boldsymbol{H}^\mathrm{H}\boldsymbol{H}\right|$ and power allocation and channel attenuation compensate to a finite value. Similarly to (3), the sum rate $R_{\text{non coop}}$ can be computed in the large system limit via the Shannon transform $\mathcal{V}_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma)$ of the matrix $\boldsymbol{H}^\mathrm{H}\boldsymbol{H}$ according to $\frac{R_{\text{sum, equal}}}{N_\mathrm{T}} \xrightarrow[\substack{N_\mathrm{T}, N_\mathrm{R}\to\infty\\ N_\mathrm{T}=N_\mathrm{R}\alpha}]{} \mathcal{V}_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}\left(P_\mathrm{Tx}\frac{\alpha}{K}\right)$. As the matrix $\boldsymbol{H}$ contains independently and identically Gaussian distributed entries with variance $\frac{1}{N_\mathrm{T}}$, the Shannon transform $\mathcal{V}_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma)$ is given by [20, Theorem 2.39]

$$\mathcal{V}_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma) = \frac{K}{\alpha}\log_2\left(1+\gamma\eta_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma)\right)-\log_2\left(\eta_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma)\right)+ \\ +\left(\eta_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma)-1\right)\log_2(\mathrm{e}), \quad (6)$$

where e is Euler's number. $\eta_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma)$ is the η-transform of the matrix $\boldsymbol{H}^\mathrm{H}\boldsymbol{H}$ as defined in [20, Definition 2.11] and stems from the implicit equation [20, Theorem 2.39]

$$\frac{K}{\alpha}\left(1-\frac{1}{1+\gamma\eta_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma)}\right) = 1-\eta_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma), \quad (7)$$

As by definition, the η- transform takes only values between 0 and 1, $\eta_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma)$ results from the positive solution of (7) and is given by

$$\eta_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}(\gamma) = \frac{1}{2}\left(1-\frac{K}{\alpha}-\frac{1}{\gamma}\right)+\sqrt{\left[\frac{1}{2}\left(1-\frac{K}{\alpha}-\frac{1}{\gamma}\right)\right]^2+\frac{1}{\gamma}}. \quad (8)$$

Introducing the variable $m_1 = \eta_{\boldsymbol{H}^\mathrm{H}\boldsymbol{H}}\left(P_\mathrm{Tx}\frac{K}{\alpha}\right)$, and inserting (8) with $\gamma = P_\mathrm{Tx}\frac{\alpha}{K}$ into (6), leads to the desired result in (4). ■

IV. Lower Bound for the Large System Sum Capacity in the MIMO Broadcast Channel

Theorem 1 can also be used as a lower bound in case the antennas at the receivers can cooperate. However, this bound is not very tight especially in case of a large number of cooperating receive antennas. For this reason we will introduce a signaling scheme in Section IV-A, which can be numerically analyzed in the large system limit and leads to a lower bound for the sum capacity that turns out to be tight by numerical simulations. The large system analysis is then carried out in Section IV-B.

A. Spatial Zero-Forcing with Dirty Paper Coding

In order to avoid the numerical difficulties associated with the solution of (1), one can decompose the MIMO broadcast channel into a system of scalar, interference-free subchannels, where interference suppression is achieved by a combination of DPC and zero-forcing. Once an encoding order for DPC has been determined, interference from previously encoded data streams is perfectly cancelled by DPC and interference suppression from data streams encoded afterwards is assured by beamforming. These concepts have been applied to MISO systems in [7] and to MIMO systems in [6] and simulation results in these references show close to optimum performance. The sum rate $R_{\text{ZF DPC}}$ achievable with successive zero-forcing and DPC computes according to

$$R_{\text{ZF DPC}} = \sum_{i=1}^{\min(N_\mathrm{T},KN_\mathrm{R})}\log_2\left(1+p_{i,\text{ZF DPC}}\lambda_{i,\text{ZF DPC}}\right), \quad (9)$$

where $\lambda_{i,\text{ZF DPC}}$ is the effective channel gain of the data stream encoded in the ith step, which considers the effect of channel attenuation, transmit and receive beamforming. Proceeding according to the algorithm in [6], the subchannel gains $\lambda_{i,\text{ZF DPC}}$ compute according to

$$\lambda_{i,\text{ZF DPC}} = \rho_1\left(\hat{\boldsymbol{H}}_{\pi(i)}\boldsymbol{P}_i\hat{\boldsymbol{H}}_{\pi(i)}^\mathrm{H}\right) \quad (10)$$

where $\pi(i)$ is the user which the ith subchannel is allocated to.

$$\boldsymbol{P}_i = \boldsymbol{I}_{N_\text{T}} - \sum_{n=1}^{i-1} \boldsymbol{t}_n \boldsymbol{t}_n^\text{H} \tag{11}$$

projects into $\text{null}\{\boldsymbol{t}_1^\text{H}, \ldots, \boldsymbol{t}_{i-1}^\text{H}\}$ and $\boldsymbol{t}_n$ denotes the transmit beamforming vector for the nth data stream, which is given by the unit-norm eigenvector corresponding to the principal eigenvalue of the matrix $\boldsymbol{P}_n \boldsymbol{H}_{\pi(n)}^\text{H} \boldsymbol{H}_{\pi(n)} \boldsymbol{P}_n$ [6]. The power $p_{i,\text{ZF DPC}}$ allocated to the ith data stream results from distributing the power budget P_Tx to the scalar interference-free subchannels via water-filling, so that

$$p_{i,\text{ZF DPC}} = \max\left(0, \eta - \frac{1}{\lambda_{i,\text{ZF DPC}}}\right), \tag{12}$$

where the water-level η is given by the implicit equation $\sum_{i=1}^{\min(N_\text{T}, KN_\text{R})} p_{i,\text{ZF DPC}} = P_\text{Tx}$. The algorithm therefore works sequentially. First the transmit beamformers $\boldsymbol{t}_n$ and the user allocation $\pi(i)$ are determined, from which the channel gains $\lambda_{i,\text{ZF DPC}}$ result[2] and then the power allocation is done via water-filling.

To find the optimum user allocation it would be required to perform an exhaustive search over all possible user allocations, which is complex. In [6] it is therefore proposed to allocate the data streams in a successive manner, i.e., to allocate in each step a data stream to that user which leads to the strongest increase in sum rate. That is equivalent to selecting $\pi(i) = \text{argmax}_k\, \rho_1\left(\hat{\boldsymbol{H}}_k \boldsymbol{P}_i \hat{\boldsymbol{H}}_k^\text{H}\right)$. However likewise to the problem of finding the asymptotic sum capacity, this allocation leads to a dependency of the matrices $\hat{\boldsymbol{H}}_{\pi(i)}$ on the matrices $\boldsymbol{P}_i$, which makes an asymptotic analysis of the channel gains in (10) involved. Instead, we therefore consider a simple user allocation, where δm with $m = \min(N_\text{T}, KN_\text{R})$ successively encoded data streams are allocated to the same user with $\frac{1}{m} < \delta \leq 1$ being an a priori chosen constant, where $\frac{1}{\delta} \in \mathbb{N}$. Furthermore these groups of successively encoded data streams are allocated to the users in a round robin fashion. This implies that the data streams $i = 1, \ldots, \delta m$ are allocated to user 1, $i = \delta m + 1 \ldots, 2\delta m$ to user 2 and so and so fourth until data streams $i = K\delta m + 1, \ldots, (K+1)\delta m$ are given to user 1 again. Thus, $\pi(i) = \text{mod}\left(\left\lfloor \frac{i-1}{\delta m} \right\rfloor, K\right) + 1$. This way, the user allocation and consequently the matrices $\boldsymbol{P}_i$ are independent of $\boldsymbol{H}_{\pi(i)}$. The reason for allocation groups of successively encoded subchannels to the same user, i.e., choosing $\delta \neq \frac{1}{m}$, will be explained in the next section.

B. Analytical Expression for a Lower Bound of the Sum Capacity

To obtain a lower bound for sum capacity, a large system expression for the sum rate achievable with the signaling scheme presented in the previous section will be derived in this section. As pointed out there, this algorithm works sequentially, i.e., first transmit beamformers are determined to obtain the scalar subchannel gains and then water-filling is performed. One can therefore first consider the asymptotic limit of the subchannel gains, and in case it turns out that these gains become independent of the current channel realization in the large system limit, water-filling can be performed over these asymptotic gains. In the following, it will therefore be first shown that the empirical distribution of the subchannel gains $\lambda_{i,\text{ZF, DPC}}$ from (10) converges to an asymptotic limit. Analogously to (2), the empirical distribution of subchannel gains denotes the fraction of subchannel gains $\lambda_{i,\text{ZF, DPC}}$ that is smaller or equal to x and an asymptotic limit exists, in case this distribution becomes independent of the channel matrices $\hat{\boldsymbol{H}}_k$ for an infinite number of transmit and receive antennas. For notational convenience we will use the normalized subchannel gains $\frac{\lambda_{i,\text{ZF DPC}}}{N_\text{R}} = \frac{1}{N_\text{R}} \rho_1\left(\hat{\boldsymbol{H}}_{\pi(i)} \boldsymbol{P}_i \hat{\boldsymbol{H}}_{\pi(i)}^\text{H}\right) = \rho_1\left(\boldsymbol{H}_{\pi(i)} \boldsymbol{P}_i \boldsymbol{H}_{\pi(i)}^\text{H}\right)$, where the channel matrices $\boldsymbol{H}_k := \frac{1}{\sqrt{N_\text{R}}} \hat{\boldsymbol{H}}_k$ have been normalized so that its entries are Gaussian i.i.d. with zero mean and variance $\frac{1}{N_\text{R}}$. In case the asymptotic empirical distribution of normalized subchannel gains and its derivative, which will be denoted as $f_\text{ZF DPC}^{(\infty)}(x)$ in the following, becomes independent of the current channel realizations in the large system limit, water-filling can be applied and the sum rate $R_\text{ZF, DPC}$ can be computed according to

$$\begin{aligned} \frac{R_\text{ZF, DPC}}{m} &= \frac{1}{m} \sum_{i=1}^{m} \log_2\left(\max[1, \hat{\eta}\lambda_{i,\text{ZF, DPC}}]\right) \\ &\underset{m\to\infty}{\longrightarrow} \int_{\lambda_\text{min}}^{\infty} \log_2(\hat{\eta}x)\, f_\text{ZF DPC}^{(\infty)}(x)\, \mathrm{d}\,x, \end{aligned} \tag{13}$$

where the scaled water-level $\hat{\eta} = N_\text{R}\eta$ computes as

$$\begin{aligned} \hat{\eta} &= \frac{N_\text{R} P_\text{Tx} + \sum\limits_{i=1}^{m} \frac{1}{\lambda_{i,\text{ZF DPC}}}}{m} \\ &\underset{m\to\infty}{\longrightarrow} \frac{\frac{P_\text{Tx}}{\min(\alpha,K)} + \int\limits_{\lambda_\text{min}}^{\infty} \frac{1}{x} f_\text{ZF DPC}^{(\infty)}(x)\, \mathrm{d}\,x}{\int\limits_{\lambda_\text{min}}^{\infty} f_\text{ZF DPC}^{(\infty)}(x)\, \mathrm{d}\,x}. \end{aligned} \tag{14}$$

Water-filling is done by replacing infinite sums over functions of $\lambda_{i,\text{ZF DPC}}$ by the corresponding integrals and λ_min is either given by the inverse water-level, i.e., $\hat{\eta} = 1/\lambda_\text{min}$, in case $f_\text{ZF DPC}^{(\infty)}(x) \neq 0$ for $x \leq \lambda_\text{min}$, or by the minimum value, for which $f_\text{ZF DPC}^{(\infty)}(x)$ is not zero. Thus, it remains to show that the distribution of subchannel gains $f_\text{ZF DPC}^{(\infty)}(x)$ indeed converges to an asymptotic limit and how it can be actually computed.

According to the allocation scheme described in the previous section, δm successively encoded subchannels are allocated to the same user. With the way the projection matrices $\boldsymbol{P}_i$ are computed according to (11) and due to the fact that the vectors $\boldsymbol{t}_n$ are the eigenvectors corresponding to the principal eigenvalues of the matrices $\hat{\boldsymbol{H}}_{\pi(n)} \boldsymbol{P}_n \hat{\boldsymbol{H}}_{\pi(n)}^\text{H} = \frac{1}{N_\text{R}} \boldsymbol{H}_{\pi(n)} \boldsymbol{P}_n \boldsymbol{H}_{\pi(n)}^\text{H}$, the normalized subchannel gains from step $(j-1)\delta m + 1$ to step $j\delta m$ are given by the δm strongest eigenvalues of the matrix $\boldsymbol{H}_{\hat{\pi}(j)} \hat{\boldsymbol{P}}_j \boldsymbol{H}_{\hat{\pi}(j)}^\text{H}$, where we define $\hat{\boldsymbol{P}}_j := \boldsymbol{P}_{((j-1)\delta m+1)}$ and $\hat{\pi}(j) = \pi((j-1)\delta m + 1)$. In order to determine the asymptotic distribution $f_\text{ZF DPC}^{(\infty)}(x)$ it is therefore necessary to show the existence of and derive the a.e.d.s of the matrices $\boldsymbol{H}_{\hat{\pi}(j)} \hat{\boldsymbol{P}}_j \boldsymbol{H}_{\hat{\pi}(j)}^\text{H}$ for $j = 1, 2, \ldots, \frac{1}{\delta}$. The δm strongest

[2]The corresponding optimum receive filters are given by matched filters [6].

eigenvalues of the matrices $\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{P}}_j\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}$ are also the δm strongest eigenvalues[3] of the matrix

$$\boldsymbol{C}_j := \hat{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{V}}_j \in \mathbb{C}^{[N_\mathrm{T}-(j-1)\delta m]\times[N_\mathrm{T}-(j-1)\delta m]}, \tag{15}$$

where $\hat{\boldsymbol{V}}_j \in \mathbb{C}^{N_\mathrm{T}\times[N_\mathrm{T}-(j-1)\delta m]}$ is an orthonormal basis of $\operatorname{span}\left\{\hat{\boldsymbol{P}}_j\right\}$ so that $\hat{\boldsymbol{P}}_j = \hat{\boldsymbol{V}}_j\hat{\boldsymbol{V}}_j^{\mathrm{H}}$. Assume the a.e.d. of this matrix exists and is given by $f_{\boldsymbol{C}_j}^{(\infty)}(x)$. Then, the δm strongest out of $[N_\mathrm{T}-(j-1)\delta m]$ eigenvalues of this matrix are contained in the interval $[\hat{\lambda}_j;\infty[$, where $\hat{\lambda}_j$ stems from the implicit equation

$$\int_{\hat{\lambda}_j}^{\infty} f_{\boldsymbol{C}_j}^{(\infty)}(x)\,\mathrm{d}\,x = \frac{\delta m}{N_\mathrm{T}-(j-1)\delta m} = \frac{\delta\xi}{1-(j-1)\delta\xi}, \tag{16}$$

where $\xi = \frac{m}{N_\mathrm{T}} = \min\left(1, \frac{K}{\alpha}\right)$. In case an a.e.d. can be found for all matrices $\boldsymbol{C}_j$, $j = 1,\ldots,\frac{1}{\delta}$, the asymptotic distribution $f_{\text{ZF DPC}}^{(\infty)}(x)$ can be obtained by taking the sum of the tails of the a.e.d.s of the $\boldsymbol{C}_j$ in the corresponding intervals $[\hat{\lambda}_j;\infty[$, i.e.,

$$f_{\text{ZF DPC}}^{(\infty)}(x) = \sum_{j=1}^{\frac{1}{\delta}} f_j^{(\infty)}(x) \tag{17}$$

with

$$f_j(x) = \begin{cases} \frac{1-(j-1)\delta\xi}{\xi} f_{\boldsymbol{C}_j}^{(\infty)}(x), & x \geq \hat{\lambda}_j \\ 0, & \text{else} \end{cases}.$$

The pre-factor $\frac{1-(j-1)\delta\xi}{\xi}$ is necessary, as for $f_{\text{ZF DPC}}^{(\infty)}(x)$ each matrix $\boldsymbol{C}_j$ contributes δm out of m subchannel gains, i.e., the integrals $\int_0^\infty f_j^{(\infty)}(x)\,\mathrm{d}\,x$ must be all equal to δ.

In the remainder it will be shown that the a.e.d.s $f_{\boldsymbol{C}_j}^{(\infty)}(x)$ indeed exist and how they can be computed numerically. For $j = 1$, $\boldsymbol{C}_1 = \boldsymbol{H}_{\hat{\pi}(1)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(1)}$ and as $\boldsymbol{H}_{\hat{\pi}}(1)$ contains Gaussian i.i.d. entries, $f_{\boldsymbol{C}_1}^{(\infty)}(x)$ exists and is given by the tail of the Marčenko-Pastur distribution, i.e., $f_{\boldsymbol{C}_1}^{(\infty)}(x) = f_{\text{MP}}(x,\alpha)$ with

$$f_{\text{MP}}(x,\alpha) = \left[1-\frac{1}{\alpha}\right]^+\delta(x) + \frac{\sqrt{[x-a]^+[b-x]^+}}{2\pi\alpha x},$$

where $a = (1-\sqrt{\alpha})^2$ and $b = (1+\sqrt{\alpha})^2$.

For $j = 2,\ldots,K$, the matrices $\hat{\boldsymbol{V}}_j$ are independent of the channel matrices $\boldsymbol{H}_{\hat{\pi}(j)}$, as the projection matrices $\hat{\boldsymbol{P}}_j$ are determined independently of the latter matrices. The matrices $\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{V}}_j$ therefore result from multiplying a matrix with Gaussian i.i.d. entries such as $\boldsymbol{H}_{\hat{\pi}(j)}$ with an independent orthonormal matrix like $\hat{\boldsymbol{V}}_j$, which leads again to matrices with Gaussian i.i.d. entries having the same mean and variance as in the original Gaussian matrices but different dimensions. This way, $\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{V}}_j$ are $N_\mathrm{R}\times[N_\mathrm{T}-(j-1)\delta m]$ matrices with Gaussian i.i.d. entries with zero mean and variance $\frac{1}{N_\mathrm{R}}$. Consequently, the eigenvalues of their Grammian products $\boldsymbol{C}_j = \hat{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{V}}_j$ also converge to an asymptotic limit and the corresponding a.e.d.s are given by

$$f_{\boldsymbol{C}_j}^{(\infty)}(x) = f_{\text{MP}}(x, \alpha(1-(j-1)\delta\xi)), \quad j = 1,\ldots,K. \tag{18}$$

For $j > K$, $f_{\boldsymbol{C}_j}^{(\infty)}(x)$ can however not be stated explicitly anymore, instead it can be derived from Proposition 1.

Proposition 1: The Stieltjes-transform $m_{\boldsymbol{C}_j}(z)$ of the matrix $\boldsymbol{C}_j$ is given by the implicit equation

$$\int_0^{\hat{\lambda}_{j-K}} \frac{f_{\boldsymbol{C}_{j-K}}(x)}{1-\beta_j+(x-z)m_{\boldsymbol{C}_j}(z)}\,\mathrm{d}\,x = \frac{1-(j-K)\delta\xi}{1-(j-K-1)\delta\xi}, \tag{19}$$

where $\beta_j = [1-(j-1)\delta\xi]\,/\,[1-(j-K)\delta\xi]$.

Proof: see Appendix B ■

Unfortunately, there is no explicit solution neither for $m_{\boldsymbol{C}_j}(z)$ nor for $f_{\boldsymbol{C}_j}^{(\infty)}(x)$ from (19). For this reason $f_{\boldsymbol{C}_j}^{(\infty)}(x)$ has to be sampled as described in the following. First Equation (19) is solved for $m_{\boldsymbol{C}_j}(z)$ with $z = \hat{\lambda}_{j-K}$. The imaginary part of $m_{\boldsymbol{C}_j}(\hat{\lambda}_{j-K})$ divided by π is then equal to $f_{\boldsymbol{C}_j}^{(\infty)}(\hat{\lambda}_{j-K})$ [e.g. [20, Eq. (2.45)]]. Due to the projections from step $(j-K)\delta m+1$ to step $(j-1)\delta m$, the principal eigenvalue of the matrix $\boldsymbol{C}_j$ will certainly not be larger than $\hat{\lambda}_{j-K}$, which is the channel gain in step $(j-K)\delta m$ [c.f. (16)], the last step the same user has received a subchannel. Thus, $f_{\boldsymbol{C}_j}^{(\infty)}(x) = 0$ for $x > \hat{\lambda}_{j-K}$ and $\hat{\lambda}_{j-K}$ can be used as a starting point for the sampling process. After $f_{\boldsymbol{C}_j}^{(\infty)}(\hat{\lambda}_{j-K})$ has been computed, z is reduced by a constant sampling distance Δ and Equation (19) is solved for $m_{\boldsymbol{C}_j}(z)$ with $z = \hat{\lambda}_{j-K}-\Delta$. This sampling is continued until $z = 0$. The integrals with $f_{\boldsymbol{C}_j}^{(\infty)}(x)$ required in (16) and (19) can then be evaluated numerically for example with the trapezoidal method (e.g. [21]), where $f_{\boldsymbol{C}_j}^{(\infty)}(x)$ is interpolated linearly between two neighboring samples.

Note that the choice of δ influences the tightness of the lower bound and the numerical accuracy. With large values for δ, the number of summands required to compute $f_{\text{ZF DPC}}^{(\infty)}$ in (17) becomes low. Thus, the implicit Equation (19) has to be solved less often so that the numerical complexity is reduced. On the other hand, it can be shown by numerically evaluating the asymptotic sum rates and comparing the results with the average sum capacity, that for large values of δ, the bound for sum capacity becomes less tight. In case one is interested in the best lower bound, one would therefore let δ go to zero, which would lead to the asymptotic sum rate of the allocation scheme described in the previous section with $\delta = \frac{1}{m}$. However, this scenario becomes numerically intractable due to the sum over infinite a.e.d.s in (17).

V. Asymptotic Performance Losses of Suboptimal Algorithms in MISO Systems

As numerically complex iterative algorithms need to be implemented in order to achieve the sum capacity in the broadcast channel, more efficient methods that are only able to approximate the optimum solution are of high practical interest. In this section we will quantify those losses analytically for two algorithms in the large system limit with non-cooperative receive antennas, i.e., in MISO systems. Numerical simulations can be avoided this way, as the results obtained also serve as a good estimation of the average sum capacity in systems with finite parameters. The large system sum capacity is given by Theorem 1 and in this section we will derive

[3] This is because $\rho_j\left(\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{P}}_j\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\right) = \rho_j\left(\hat{\boldsymbol{P}}_j\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{P}}_j\right) = \rho_j\left(\hat{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{V}}_j\right)$ for all $j = 1,\ldots,\min(N_\mathrm{R},[N_\mathrm{T}-(j-1)\delta m])$

analytical lower bounds for the two signaling schemes from [7] and [8]. Both are based on spatial zero-forcing and successive resource allocation, where the method from [7] requires DPC and the scheme in [8] solely relies on interference suppression through linear beamforming. In [8], the sum rates achievable with the proposed method is analyzed in limit of infinitely many users. The results do however not allow any conclusions to the performance in finite systems.

A. Spatial Zero-Forcing with Dirty Paper Coding

The concept of spatial zero-forcing with DPC has already been introduced in Section IV-A with cooperating receive antennas. In case the receive antennas cannot cooperate, the channel gains in (10) can no longer be achieved but instead compute according to

$$\hat{\lambda}_{i,\text{ZF DPC}} = \boldsymbol{e}_{r(i)}^{\mathsf{T}} \hat{\boldsymbol{H}}_{\pi(i)}^{\mathrm{H}} \boldsymbol{P}_i \hat{\boldsymbol{H}}_{\pi(i)} \boldsymbol{e}_{r(i)}, \tag{20}$$

where $\boldsymbol{P}_i$ given by (11) and $r(i)$ denotes the receive antenna selected in step i, so that the data stream encoded at ith step at the transmitter is intended for user $\pi(i)$ and received at its $r(i)$th antenna. The transmit vectors are chosen in a successive manner according to

$$\boldsymbol{t}_i = \frac{1}{\sqrt{\boldsymbol{e}_{r(i)}^{\mathsf{T}} \hat{\boldsymbol{H}}_{\pi(i)}^{\mathrm{H}} \boldsymbol{P}_i \hat{\boldsymbol{H}}_{\pi(i)} \boldsymbol{e}_{r(i)}}} \boldsymbol{P}_i \hat{\boldsymbol{H}}_{\pi(i)} \boldsymbol{e}_{r(i)}, \tag{21}$$

in order to fulfill the zero-forcing constraints that cannot be assured by DPC. The sum rate is then computed with the channel gains $\hat{\lambda}_{i,\text{ZF DPC}}$ as in (9). The optimum user selection $\pi(i)$ and antenna selection $r(i)$ would require an exhaustive search over all possible allocations, which is infeasible in practice. For this reason in [7] it is proposed to choose the user and antenna in each step so that the sum rate becomes maximum. Here, we consider a further sub-optimum selection, where in each step user and antenna are selected randomly. This leads to a lower bound for the performance of the signaling scheme in the large system limit, which is given by Proposition 2.

Proposition 2:

$$\frac{R_{\text{ZF DPC}}^{\text{lb}}}{N_{\text{T}}} \underset{\substack{N_{\text{T}}, N_{\text{R}} \to \infty \\ N_{\text{T}} = N_{\text{R}}\alpha}}{\longrightarrow} \theta \log_2 \left(\frac{1}{\theta} \left[P_{\text{Tx}} - \ln(1-\theta) \right] \right) - \\ - (1-\theta) \log_2 (1-\theta) - \frac{\theta}{\ln 2}. \tag{22}$$

θ stems from water-filling and is determined as follows. First, the implicit equation $P_{\text{Tx}} - \ln(1-\theta) = \frac{\theta}{1-\theta}$ has to be solved for θ. If the result is smaller than $\min(1, \frac{K}{\alpha})$, the solution for θ has been found, otherwise θ is given by $\theta = \min(1, \frac{K}{\alpha})$.

Proof: see Appendix D. ∎

Note that the lower bound (22) is valuable for analysis purposes anyway, as it can be used to state an upper bound for the loss compared to the optimum.

B. Spatial Zero-Forcing without Dirty Paper Coding

So far, it is still necessary to apply DPC at the transmitter, which requires significant computational resources in practical implementations. For that reason the concept of successive zero-forcing has been extended to scenarios, where DPC is not applied. In this case, interference between the data streams is completely suppressed by linear beamforming. The sum rate is then given by

$$R_{\text{ZF}} = \sum_{i=1}^{N_{\text{stop}}} \log_2 (1 + p_{i,\text{ZF}} \lambda_{i,\text{ZF}}), \tag{23}$$

where $p_{i,\text{ZF}}$ is the power allocated to the ith subchannel and results from water-filling as $p_{i,\text{DPC}}$ in (12). $N_{\text{stop}} \leq \min(N_{\text{T}}, KN_{\text{R}})$ denotes the number of active subchannels. The gains $\lambda_{i,\text{ZF}}$ of these subchannels can be computed according to (e.g [8])

$$\lambda_{i,\text{ZF}}^{-1} = \boldsymbol{e}_i^{\mathsf{T}} \left(\boldsymbol{H}_{\text{comp}} \boldsymbol{H}_{\text{comp}}^{\mathrm{H}} \right)^{-1} \boldsymbol{e}_i,$$

where $\boldsymbol{H}_{\text{comp}} \in \mathbb{C}^{N_{\text{stop}} \times N_{\text{T}}} = \boldsymbol{\Pi} \left[\hat{\boldsymbol{H}}_1^{\mathrm{H}}, \ldots, \hat{\boldsymbol{H}}_k^{\mathrm{H}} \right]^{\mathrm{H}}$ is the composite channel matrix and $\boldsymbol{\Pi} \in \mathbb{C}^{N_{\text{stop}} \times KN_{\text{R}}}$ is a selection matrix that selects those rows of $\left[\hat{\boldsymbol{H}}_1^{\mathrm{H}}, \ldots, \hat{\boldsymbol{H}}_k^{\mathrm{H}} \right]^{\mathrm{H}}$ that correspond to users and receive antennas that are selected for transmission. As in the DPC case, the optimum choice for $\boldsymbol{\Pi}$ would require an exhaustive search over all possible selection matrices, which is infeasible in practice. In [8] it is therefore proposed to allocate the users and receive antennas in a successive manner, i.e., to select in each allocation step this user that leads to the strongest increase in sum rate provided that the previously selected users and receive antennas are also served. The allocation is stopped, in case no increase in sum rate can be observed. Such an allocation scheme is however difficult to analyze in the large system limit, as shown in Appendix E, which is why we consider a random selection of users and receive antennas in the following. This leads to a lower bound for the sum rate in the large system limit, which is given by Proposition 3.

Proposition 3: The sum rate asymptotically achievable with spatial zero-forcing can be lower bounded as

$$\frac{R_{\text{ZF}}^{\text{lb}}}{N_{\text{T}}} \underset{\substack{K, N_{\text{T}} \to \infty \\ N_{\text{T}} = N_{\text{R}}\alpha}}{\longrightarrow} \max_{\gamma} \gamma \log_2 \left(1 + P_{\text{Tx}} \left(\frac{1}{\gamma} - 1 \right) \right), \\ \text{s.t. } \gamma \leq \min \left(1, \frac{K}{\alpha} \right), \tag{24}$$

which is a concave optimization problem and can therefore be solved for example by bisection [22, Chapter 8.2].

Proof: see Appendix E ∎

Despite a sub-optimal user allocation, the expression in (24) is still optimized over the fraction of active data streams $\gamma = \frac{N_{\text{stop}}}{N_{\text{T}}}$. As a consequence, in case the optimum N_{stop} is equal to KN_{R}, i.e., $\boldsymbol{\Pi} = \boldsymbol{I}_{KN_{\text{R}}}$, the bound becomes exact, because the optimum user allocation matrix is then equal to our a priori chosen matrix.

VI. Numerical Results

Figure 1 exhibits the sum rates normalized to the number of transmit antennas and averaged over 1000 circularly symmetric Gaussian channel matrices. The ratio α of transmit antennas to receive antennas is set to $\alpha = \frac{1}{2}$, $K = 1$ and the SNR is equal to 10dB. The large system sum rates of the corresponding algorithms are compared to the average sum

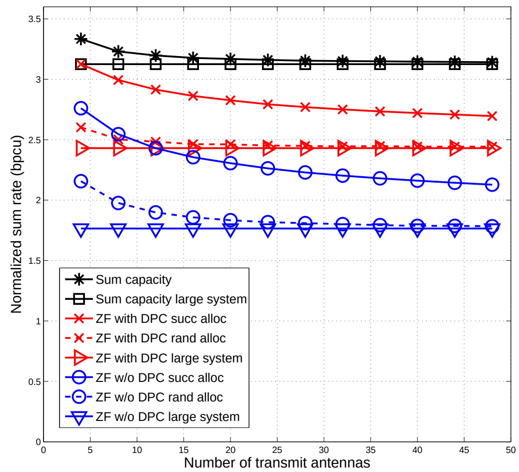


Fig. 1. Comparison of normalized average sum rates compared to large system sum rates in MISO systems ($K = 1$) with $\alpha = \frac{1}{2}$, $K = 1$, SNR= 10dB.

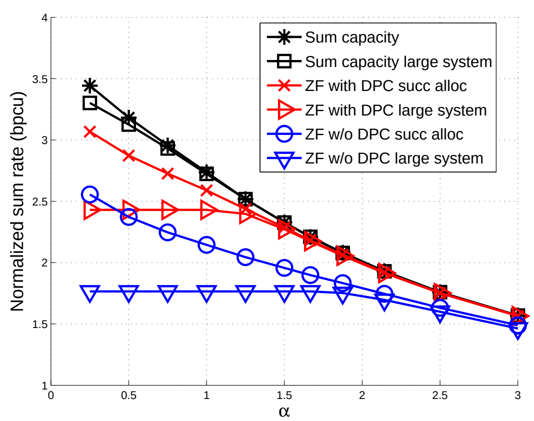


Fig. 2. Comparison of normalized average sum rates compared to large system sum rates in MISO systems with SNR= 10dB, $N_{\mathrm{T}} = 15$.

rates obtained by simulations. For the sum capacity the large system sum rate serves as very good approximation for the average sum rates even with only a few number of transmit antennas and users. Similar observations can be made, when spatial zero-forcing with and without DPC are considered with a random user allocation ("rand alloc"). Compared to the successive user allocation ("succ alloc") from [7] for the DPC and from [8] for the linear case, a gap can be observed. When α increases, i.e., there are more degrees of freedom at the transmitter compared to the number of KN_{R} data streams, this gap diminishes, as shown in Figure 2, where the ergodic and large system sum rates are plotted versus α. For the simulations an SNR of 10dB is used and the number of transmit antennas for computing the ergodic rates is equal to $N_{\mathrm{T}} = 15$. Note that for $\alpha \leq 1$ the large system sum rates of the suboptimal schemes are independent of α, as in (22) the constraint on θ and in (24) the constraint on γ are independent of α then.

In Figure 3 the ergodic sum rates normalized to the number of transmit antennas and averaged over 1000 Gaussian channel realizations are plotted versus the number of transmit antennas at an SNR of 10 dB. The ratio α of transmit antennas to

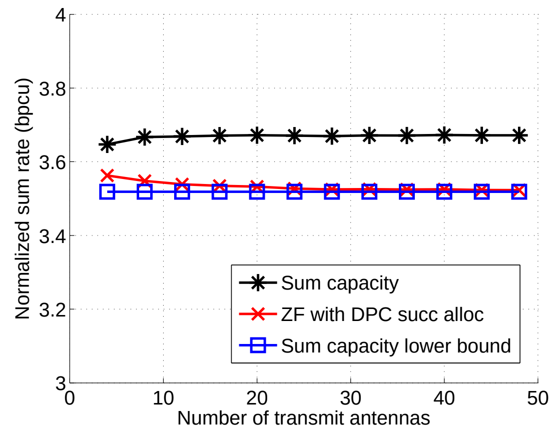


Fig. 3. Comparison of ergodic sum rates with large system sum rate in a system with $K = 5$ users, $\alpha = 2$, SNR= 10dB, $\delta = 0.05$ and $\varDelta = 0.001$.

receive antennas is set to $\alpha = 2$. The average sum rate of the successive resource allocation and spatial zero-forcing scheme ("ZF with DPC succ alloc") is plotted. The large system sum rate obtained as described above with $\delta = 0.05$ is plotted as a line with circles, where a sampling distance of $\varDelta = 0.001$ has been used to solve (19). The large system sum rate obtained this way serves as a relatively tight lower bound for the sum capacity and as a very good approximation for the ergodic sum rate achievable with spatial zero-forcing and DPC using a successive user allocation also for finite system parameters, which becomes exact for systems with $N_{\mathrm{T}} \geq 16$. In Figure 4 the ergodic and large system sum rates are plotted versus the the ratio α, where $N_{\mathrm{T}} = 15$ has been used for the computation

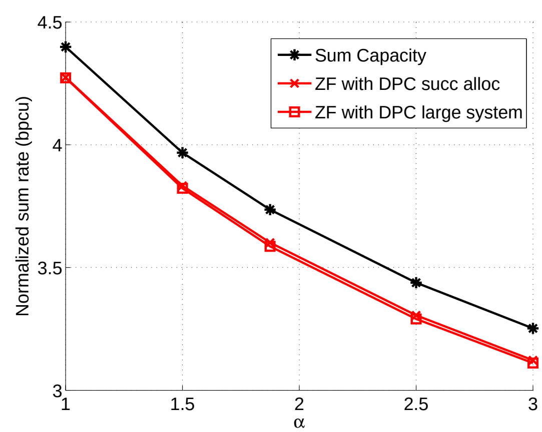


Fig. 4. Comparison of ergodic sum rates with large system sum rate in a system with $K = 5$ users, $N_{\mathrm{T}} = 15$, SNR= 10dB, $\delta = 0.05$ and $\varDelta = 0.001$.

of the ergodic rates and the remaining parameters have been unchanged compared to Figure 3. Finally, in Figure 5 the asymptotic lower bound is plotted versus $1/\delta$. With increasing $1/\delta$ the bound becomes more tight, until from $\delta = 1/20$ upwards only slight improvements can be observed, a useful fact

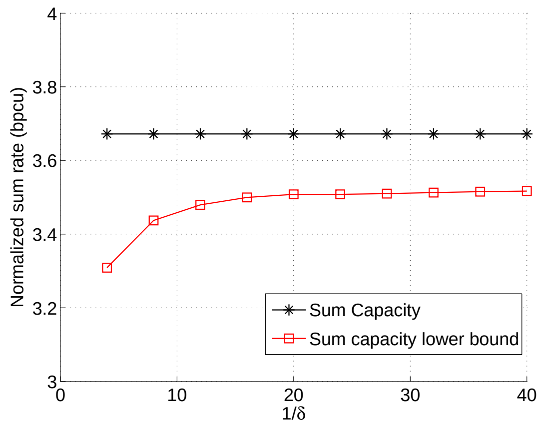


Fig. 5. Accuracy of large system lower bound for sum capacity in the MIMO broadcast channel versus $\frac{1}{\delta}$ in a system with $K = 5$ users, $\alpha = 2$, SNR= 10dB, $\varDelta = 0.001$.

for practical choices of δ. Note that to obtain the asymptotic expression at $1/\delta = 4$ the implicit equation (19) needs not to be solved at all, as all required distributions $f^{(\infty)}_{\boldsymbol{C}_j}$ can be obtained from (18).

VII. Conclusions

In this paper we have presented expressions for large system sum rates achievable by algorithms relying on perfect channel state information in the MIMO broadcast channel, when the number of transmit and receive antennas both go to infinity at a finite fixed ratio. For non-cooperating receive antennas, i.e., MISO systems, we have derived an asymptotic expression for sum capacity. Furthermore the performance of two sub-optimal schemes, which rely on spatial zero-forcing beamforming, has been quantified in the large system limit and shown to estimate the losses in sum rate compared to sum capacity quite well. As for MIMO systems with cooperating receive antennas, the asymptotic sum capacity turned out to be hard to determine, we have derived an expression for the sum rate achievable with spatial zero-forcing and DPC in the large system limit, which serves as a lower bound for sum capacity. For the problem of weighted sum rate maximization, the same method and tools as in this paper can be used to derive asymptotic weighted sum rates. As the expressions get more involved in this case, we have restricted to sum rate maximization here.

Appendix A
Proof of Lemma 1

In order to proof Lemma 1, it will be shown that an equal power allocation satisfies the Karush-Kuhn-Tucker conditions of Problem (5) in the specified large system limit. Those conditions read as

$$\begin{aligned}
&\boldsymbol{h}_n^{\mathrm{H}}\left(\boldsymbol{I}_{N_{\mathrm{T}}}+N_{\mathrm{T}}\left[\sum_{m=1}^{KN_{\mathrm{R}}}\boldsymbol{h}_m\boldsymbol{h}_m^{\mathrm{H}}w_m\right]\right)^{-1}\boldsymbol{h}_n w_n=\mu w_n,\\
&\qquad \forall n=1,\dots,KN_{\mathrm{R}},\\
&\mu\left(\sum_{n=1}^{KN_{\mathrm{R}}}w_n-P_{\mathrm{Tx}}\right)=0,\quad \mu\geq 0,\\
&w_n\geq 0,\forall n=1,\dots,KN_{\mathrm{R}},
\end{aligned}\tag{25}$$

where μ is the Lagrangian multiplier and $\hat{\boldsymbol{H}}^{\mathrm{H}}\boldsymbol{W}\hat{\boldsymbol{H}} = N_{\mathrm{T}}\sum_{m=1}^{K}\boldsymbol{h}_m\boldsymbol{h}_m^{\mathrm{H}}w_m$ with $\mathrm{diag}\left(w_1,\dots,w_{KN_{\mathrm{R}}}\right) = \boldsymbol{W}$, and $[\boldsymbol{h}_1,\dots,\boldsymbol{h}_{KN_{\mathrm{R}}}] = \frac{1}{\sqrt{N_{\mathrm{T}}}}\hat{\boldsymbol{H}}^{\mathrm{H}}$. Thus, the entries in the vectors $\boldsymbol{h}_n$ are i.i.d. Gaussian with zero mean and variance $\frac{1}{N_{\mathrm{T}}}$. With a uniform power allocation $w_n = \frac{P_{\mathrm{Tx}}}{KN_{\mathrm{R}}}$, the (in)equalities in the last two lines of (25) are clearly fulfilled apart from $\mu \geq 0$ and the remaining KKTs read as

$$\boldsymbol{h}_n^{\mathrm{H}}\left(\boldsymbol{I}_{N_{\mathrm{T}}}+\frac{N_{\mathrm{T}}P_{\mathrm{Tx}}}{N_{\mathrm{R}}K}\left[\sum_{m=1}^{K}\boldsymbol{h}_m\boldsymbol{h}_m^{\mathrm{H}}\right]\right)^{-1}\boldsymbol{h}_n=\mu,$$

Now the matrix inversion lemma is applied to these conditions so that the inverse matrices are independent of the vector $\boldsymbol{h}_n$ and

$$\boldsymbol{h}_n^{\mathrm{H}}\left(\boldsymbol{I}_{N_{\mathrm{T}}}+\frac{P_{\mathrm{Tx}}\alpha}{K}\sum_{m=1}^{KN_{\mathrm{R}}}\boldsymbol{h}_m\boldsymbol{h}_m^{\mathrm{H}}\right)^{-1}\boldsymbol{h}_n=\frac{a_n}{1+\frac{P_{\mathrm{Tx}}\alpha}{K}a_n},\tag{26}$$

where $a_n = \boldsymbol{h}_n^{\mathrm{H}}\left(\boldsymbol{I}_{N_{\mathrm{T}}}+\frac{P_{\mathrm{Tx}}\alpha}{K}\boldsymbol{B}_n\right)^{-1}\boldsymbol{h}_n$, and $\boldsymbol{B}_n = \sum_{\substack{m=1\\m\neq n}}^{KN_{\mathrm{R}}}\boldsymbol{h}_m\boldsymbol{h}_m^{\mathrm{H}}$. With the matrix $\boldsymbol{I}_{N_{\mathrm{T}}}+\frac{P_{\mathrm{Tx}}\alpha}{K}\boldsymbol{B}_n$ independent of $\boldsymbol{h}_n$ and $\boldsymbol{h}_n$ containing i.i.d. entries, Corollary 1 from [23] can be applied, so that a_n converges for $N_{\mathrm{T}}\to\infty$ according to $\frac{a_n}{N_{\mathrm{T}}} - \frac{1}{N_{\mathrm{T}}}\mathrm{tr}\left(\boldsymbol{I}_{N_{\mathrm{T}}}+\frac{P_{\mathrm{Tx}}\alpha}{K}\boldsymbol{B}_n\right)^{-1} \underset{N_{\mathrm{T}}\to\infty}{\longrightarrow} 0$. $\frac{1}{N_{\mathrm{T}}}\mathrm{tr}\left(\boldsymbol{I}_{N_{\mathrm{T}}}+\frac{P_{\mathrm{Tx}}\alpha}{K}\boldsymbol{B}_n\right)^{-1}$ asymptotically converges to $\eta_{\boldsymbol{B}_n}\left(\frac{P_{\mathrm{Tx}}\alpha}{K}\right)$, the η-transform of the matrix $\boldsymbol{B}_n$ [20, Def. 2.11]. Following [20, Theorem 2.39], this transform and consequently a_n stems from the implicit equation

$$\frac{K-1}{\alpha}=\frac{1-\frac{a_n}{N_{\mathrm{T}}}}{1-\frac{1}{1+a_n\frac{P_{\mathrm{Tx}}\alpha}{K}}}.\tag{27}$$

Obviously, the implicit equation (27) leads to the same solution for all $n = 1,\dots,KN_{\mathrm{R}}$, which is why together with (26) the KKTs in the first line of (25) are fulfilled by an equal power allocation in the large system limit. Furthermore (27) leads to a positive solution for a_n and consequently for μ, so that the inequality $\mu \geq 0$ is also fulfilled. As (25) belong to a concave optimization problem, i.e., sum rate maximization in the MISO dual uplink channel, the KKT point is necessary and sufficient.

Appendix B
Proof of Proposition 1

As the rank of $\hat{\boldsymbol{P}}_j = \hat{\boldsymbol{V}}_j\hat{\boldsymbol{V}}_j^{\mathrm{H}}$ is given by $N_{\mathrm{T}} - (j-1)\delta m$ and therefore reduced by δm with each subchannel

$$\mathrm{span}\left\{\hat{\boldsymbol{P}}_j\right\}=\mathrm{span}\left\{\hat{\boldsymbol{V}}_j\right\}\subset\mathrm{span}\left\{\hat{\boldsymbol{P}}_n\right\},\quad \forall n<j.\tag{28}$$

Of special interest is the case $n = \ell_j := j - K$. From (28) it follows that $\hat{\boldsymbol{V}}_j = \hat{\boldsymbol{P}}_{\ell_j}\hat{\boldsymbol{V}}_j$ and consequently

$$\boldsymbol{C}_j = \hat{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{V}}_j = \hat{\boldsymbol{V}}_j^{\mathrm{H}}\hat{\boldsymbol{P}}_{\ell_j}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{P}}_{\ell_j}\hat{\boldsymbol{V}}_j.$$

A reduced eigenvalue decomposition of the matrix $\hat{\boldsymbol{P}}_{\ell_j}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{P}}_{\ell_j}$ can be stated as

$$\hat{\boldsymbol{P}}_{\ell_j}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{P}}_{\ell_j} = \boldsymbol{V}_{1,\ell_j}\boldsymbol{\Sigma}_{1,\ell_j}\boldsymbol{V}_{1,\ell_j}^{\mathrm{H}} + \boldsymbol{V}_{\mathrm{r},\ell_j}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}\boldsymbol{V}_{\mathrm{r},\ell_j}^{\mathrm{H}}, \tag{29}$$

where $\boldsymbol{\Sigma}_{1,\ell_j} \in \mathbb{C}^{\delta m\times\delta m}$ is a diagonal matrix containing the δm strongest eigenvalues of the matrix $\hat{\boldsymbol{P}}_{\ell_j}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{P}}_{\ell_j}$ that are the subchannel gains from step $(\ell_j-1)\delta m+1$ to step $\ell_j\delta m$. $\boldsymbol{V}_{1,\ell_j} \in \mathbb{C}^{N_{\mathrm{T}}\times\delta m}$ contains the corresponding eigenvectors. Due to the multiplication with projection matrices, $\hat{\boldsymbol{P}}_{\ell_j}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{P}}_{\ell_j}$ contains at least $(\ell_j-1)\delta m$ zero eigenvalues. Omitting these zero eigenvalues, the remaining $N_{\mathrm{T}} - \ell_j\delta m$ eigenvalues besides the δm strongest ones are subsumed in the matrix $\boldsymbol{\Sigma}_{\mathrm{r},\ell_j} \in \mathbb{C}^{[N_{\mathrm{T}}-\ell_j\delta m]\times[N_{\mathrm{T}}-\ell_j\delta m]}$ and $\boldsymbol{V}_{\mathrm{r},\ell_j} \in \mathbb{C}^{N_{\mathrm{T}}\times[N_{\mathrm{T}}-\ell_j\delta m]}$ contains the corresponding eigenvectors.

As $\boldsymbol{V}_{1,\ell_j}$ contains the transmit vectors for the $(\ell_j-1)\delta m+1$th to the $\ell_j\delta m$th data streams and $\hat{\boldsymbol{V}}_j$ is a basis of the null-space of all transmit vectors from the previously allocated subchannels, i.e., also of the transmit vectors of the subchannels $(\ell_j-1)\delta m+1$ to $\ell_j\delta m$, $\boldsymbol{V}_{1,\ell_j}$ lies in $\mathrm{null}\left\{\hat{\boldsymbol{V}}_j\right\}$ and therefore

$$\boldsymbol{C}_j = \hat{\boldsymbol{V}}_j^{\mathrm{H}}\hat{\boldsymbol{P}}_{\ell_j}\boldsymbol{H}_{\hat{\pi}(j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(j)}\hat{\boldsymbol{P}}_{\ell_j}\hat{\boldsymbol{V}}_j = \hat{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{V}_{\mathrm{r},\ell_j}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}\boldsymbol{V}_{\mathrm{r},\ell_j}^{\mathrm{H}}\hat{\boldsymbol{V}}_j^{\mathrm{H}}.$$

Furthermore the span of $\boldsymbol{V}_{\mathrm{r},\ell_j}$ is composed as

$$\mathrm{span}\left\{\boldsymbol{V}_{\mathrm{r},\ell_j}\right\} = \mathrm{span}\left\{\hat{\boldsymbol{P}}_{\ell_j} - \boldsymbol{V}_{1,\ell_j}\boldsymbol{V}_{1,\ell_j}^{\mathrm{H}}\right\} = \mathrm{span}\left\{\hat{\boldsymbol{P}}_{\ell_j+1}\right\}. \tag{30}$$

so that $\boldsymbol{V}_{\mathrm{r},\ell_j}$ is a basis of $\hat{\boldsymbol{P}}_{\ell_j+1}$. By using (28) we know that $\mathrm{span}\left\{\hat{\boldsymbol{V}}_j\right\} \subset \mathrm{span}\left\{\hat{\boldsymbol{V}}_{\ell_j+1}\right\}$ and therefore the matrix $\hat{\boldsymbol{V}}_j$ can be stated as $\hat{\boldsymbol{V}}_j = \boldsymbol{V}_{\mathrm{r},\ell_j}\tilde{\boldsymbol{V}}_j$, where $\tilde{\boldsymbol{V}}_j \in \mathbb{C}^{[N_{\mathrm{T}}-\ell_j\delta m]\times[N_{\mathrm{T}}-(j-1)\delta m]}$ represents $\hat{\boldsymbol{V}}_j$ in the basis $\boldsymbol{V}_{\mathrm{r},\ell_j}$. Hence, $\boldsymbol{C}_j$ can be decomposed as $\boldsymbol{C}_j = \tilde{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}\tilde{\boldsymbol{V}}_j$. To show the asymptotic convergence of the eigenvalues of $\boldsymbol{C}_j$, we will first analyze the matrix $\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ in the large system limit and then derive the a.e.d. for $\boldsymbol{C}_j$ from this matrix. For this purpose first the following Lemma will be required.

Lemma 2: The matrix $\tilde{\boldsymbol{V}}_j$ is uniformly distributed over the manifold of $[N_{\mathrm{T}}-\ell_j\delta m]\times[N_{\mathrm{T}}-(j-1)\delta m]$ complex matrices with $\tilde{\boldsymbol{V}}_j^{\mathrm{H}}\tilde{\boldsymbol{V}}_j = \boldsymbol{I}_{N_{\mathrm{T}}-(j-1)\delta m}$

Proof: see Appendix C ∎

Lemma 2 implies that $\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}$ is unitarily invariant. As additionally the diagonal matrix $\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ is independent of $\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}$, both matrices are asymptotically free [20, Definition 2.19], which can be shown with the proof of the theorem in [24]. From Theorem 2.68 and Example 2.51 in [20] it can be concluded that the a.e.d. of the matrix $\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ exists, as long as the eigenvalues of $\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ asymptotically converge. For $\ell_j = 1,\ldots,K$ this has been shown in Section IV-B, as $\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ contains the δm strongest eigenvalues of the matrices $\boldsymbol{C}_{\ell_j} = \hat{\boldsymbol{V}}_{\ell_j}\boldsymbol{H}_{\hat{\pi}(\ell_j)}^{\mathrm{H}}\boldsymbol{H}_{\hat{\pi}(\ell_j)}\hat{\boldsymbol{V}}_{\ell_j}$ and their a.e.d.s are given by (18). Therefore the eigenvalues of the matrices $\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ asymptotically converge for $j = K+1,\ldots,2K$. Once it has been shown that in this case also the eigenvalues of the matrices $\boldsymbol{C}_j$ asymptotically converge, it can be proven by induction that the a.e.d.s also exits for the matrices $\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ with $\ell_j > 2K$.

However, those a.e.d.s cannot be derived explicitly anymore. Instead Example 2.51 from [20] is used to state an implicit equation for the η- transform of the matrix $\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ according to

$$\eta_{\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}}(\gamma) = \eta_{\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}}\left(\gamma + \gamma\frac{\beta_j}{\eta_{\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}}(\gamma)}\right), \tag{31}$$

where

$$\beta_j = \frac{N_{\mathrm{T}}-(j-1)\delta m}{N_{\mathrm{T}}-\ell_j\delta m} = \frac{1-(j-1)\delta\xi}{1-\ell_j\delta\xi}. \tag{32}$$

Following Lemma 2.28 from [20] the η transform (and therefore the a.e.d) of the matrix $\boldsymbol{C}_j$ exists and is related to the η-transform of $\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ according to

$$\eta_{\boldsymbol{C}_j}(\gamma) = 1 - \frac{1}{\beta_j} + \frac{1}{\beta_j}\eta_{\tilde{\boldsymbol{V}}_j\tilde{\boldsymbol{V}}_j^{\mathrm{H}}\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}}(\gamma). \tag{33}$$

By inserting (33) into (31), the η-transform $\eta_{\boldsymbol{C}_j}(\gamma)$ is given implicitly by

$$\beta_j\eta_{\boldsymbol{C}_j}(\gamma) - \beta_j + 1 = \eta_{\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}}\left(\frac{\gamma\beta_j\eta_{\boldsymbol{C}_j}(\gamma)}{\beta_j\eta_{\boldsymbol{C}_j}(\gamma) - \beta_j + 1}\right). \tag{34}$$

The diagonal elements of the matrix $\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ are given by the eigenvalues of the matrix $\boldsymbol{C}_{\ell_j}$ except the largest δm ones, which correspond to the subchannel gains from step $(\ell_j - 1)\delta m+1$ to step $\ell_j\delta m$ [c.f. (29)]. The asymptotic eigenvalue distribution of the matrix $\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ is therefore given by the a.e.d. of the matrix $\boldsymbol{C}_{\ell_j}$ truncated at $\hat{\lambda}_{\ell_j}$, where $\hat{\lambda}_{\ell_j}$ is defined in (16), and normalized by $\frac{N_{\mathrm{T}}-(\ell_j-1)\delta m}{N_{\mathrm{T}}-\ell_j\delta m}$. Thus, the η-transform of the matrix $\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}$ can be written as

$$\eta_{\boldsymbol{\Sigma}_{\mathrm{r},\ell_j}}(\gamma) = \frac{N_{\mathrm{T}}-(\ell_j-1)\delta m}{N_{\mathrm{T}}-\ell_j\delta m}\int_0^{\hat{\lambda}_{\ell_j}}\frac{f_{\boldsymbol{C}_{\ell_j}}^{(\infty)}(x)}{1+\gamma x}\,\mathrm{d}\,x, \tag{35}$$

Inserting (35) into (34) and using the relationship between the η- and the Stieltjes transform [20, Eq. (2.48)] leads to

$$\int_0^{\hat{\lambda}_{\ell_j}}\frac{f_{\boldsymbol{C}_{\ell_j}}^{(\infty)}(x)}{1-\beta_j+(x-z)m_{\boldsymbol{C}_j}(z)}\,\mathrm{d}\,x = \frac{N_{\mathrm{T}}-\ell_j\delta m}{N_{\mathrm{T}}-(\ell_j-1)\delta m}.$$

Using $\ell_j = j-K$ leads to the desired result in (19).

Appendix C
Proof of Lemma 2

In the following it will be shown that the matrix $\tilde{\boldsymbol{V}}_j$ is uniformly distributed over the manifold of $[N_{\mathrm{T}}-\ell_j\delta m]\times[N_{\mathrm{T}}-(j-1)\delta m]$ complex matrices with $\tilde{\boldsymbol{V}}_j^{\mathrm{H}}\tilde{\boldsymbol{V}}_j = \boldsymbol{I}_{N_{\mathrm{T}}-(j-1)\delta m}$. This is the case, if the subspace spanned by $\hat{\boldsymbol{V}}_j$ is uniformly distributed within the subspace spanned by $\boldsymbol{V}_{\mathrm{r},\ell_j}$, which is equal to $\mathrm{span}\left\{\hat{\boldsymbol{P}}_{\ell_j+1}\right\}$ [c.f. (30)]. This statement is in turn true, when with each step a group of δm subchannels is allocated to one user, the subspace $\mathrm{span}\left\{\hat{\boldsymbol{P}}_j\right\}$ is uniformly distributed within $\mathrm{span}\left\{\hat{\boldsymbol{P}}_{j-1}\right\}$, for $j = 1, \ldots, \frac{1}{\delta}$. Concerning $j = 2$, the basis of $\mathrm{span}\left\{\hat{\boldsymbol{P}}_2\right\}$ is composed by the eigenvectors corresponding to the $N_\mathrm{T} - \delta m$ smallest eigenvalues of the matrix $\boldsymbol{H}^\mathrm{H}_{\hat{\pi}(1)}\boldsymbol{H}_{\hat{\pi}(1)}$. As the matrix $\boldsymbol{H}_{\hat{\pi}(1)}$ contains Gaussian i.i.d. entries and its Grammian $\boldsymbol{H}^\mathrm{H}_{\hat{\pi}(1)}\boldsymbol{H}_{\hat{\pi}(1)}$ is therefore unitarily invariant, the matrix of its eigenvectors is Haar distributed [20, Lemma 2.6]. This is why $\mathrm{span}\left\{\hat{\boldsymbol{P}}_2\right\}$ is uniformly distributed within $\mathrm{span}\{\hat{\boldsymbol{P}}_1\} = \mathbb{C}^{N_\mathrm{T}}$. The eigenvectors corresponding to the non-zero eigenvalues of the matrix $\hat{\boldsymbol{P}}_2\boldsymbol{H}^\mathrm{H}_{\hat{\pi}(2)}\boldsymbol{H}_{\hat{\pi}(2)}\hat{\boldsymbol{P}}_2$ can be decomposed as $\hat{\boldsymbol{V}}_2\hat{\boldsymbol{U}}_2$, where $\hat{\boldsymbol{U}}_2 \in \mathbb{C}^{[N_\mathrm{T}-\delta m]\times[N_\mathrm{T}-\delta m]}$ are the eigenvectors of the matrix $\boldsymbol{C}_2 = \hat{\boldsymbol{V}}_2^\mathrm{H}\boldsymbol{H}^\mathrm{H}_{\hat{\pi}(2)}\boldsymbol{H}_{\hat{\pi}(2)}\hat{\boldsymbol{V}}_2$. As $\mathrm{span}\left\{\hat{\boldsymbol{P}}_2\right\}$ and therefore its basis $\hat{\boldsymbol{V}}_2$ is uniformly distributed within $\mathrm{span}\{\hat{\boldsymbol{P}}_1\}$, $\boldsymbol{C}_2$ is also unitarily invariant and the eigenspace corresponding to its $N_\mathrm{T} - \delta m$ smallest eigenvalues is uniformly distributed in $\mathbb{C}^{N_\mathrm{T}-\delta m}$. As this eigenspace defines a basis for $\mathrm{span}\left\{\hat{\boldsymbol{P}}_3\right\}$ in $\mathrm{span}\left\{\hat{\boldsymbol{P}}_2\right\}$, the former span is uniformly distributed within $\mathrm{span}\{\hat{\boldsymbol{P}}_2\}$. Continuing this way for $j = 3, \ldots, \frac{1}{K}$ concludes the proof.

Appendix D
Proof of Proposition 2

In this section, the channel gains $\hat{\lambda}_{i,\text{ZF DPC}} = \boldsymbol{e}^\mathsf{T}_{r(i)}\hat{\boldsymbol{H}}^\mathrm{H}_{\pi(i)}\boldsymbol{P}_i\hat{\boldsymbol{H}}_{\pi(i)}\boldsymbol{e}_{r(i)} := N_\mathrm{T}\boldsymbol{h}^\mathrm{H}_{\pi(i),r(i)}\boldsymbol{P}_i\boldsymbol{h}^\mathrm{H}_{\pi(i),r(i)}$ from (20) are analyzed for an infinite number of transmit and receive antennas. For the purpose of large system analysis, the vectors $\boldsymbol{h}_{\pi(i),r(i)} := \frac{1}{\sqrt{N_\mathrm{T}}}\hat{\boldsymbol{H}}_{\pi(i)}\boldsymbol{e}_{r(i)}$ have been introduced so that they contain Gaussian i.i.d. entries with zero mean and variance $\frac{1}{N_\mathrm{T}}$. With the assumption of $\boldsymbol{P}_i$ being independent of $\boldsymbol{h}_{\pi(i),r(i)}$, which follows from the random user selection, Corollary 1 from [23] can be applied and the channel gains $\frac{\hat{\lambda}_{i,\text{ZF DPC}}}{N_\mathrm{T}} = \boldsymbol{h}^\mathrm{H}_{\pi(i),r(i)}\boldsymbol{P}_i\boldsymbol{h}_{\pi(i),r(i)}$ in the large system limit converge according to

$$\boldsymbol{h}^\mathrm{H}_{\pi(i),r(i)}\boldsymbol{P}_i\boldsymbol{h}_{\pi(i),r(i)} - \frac{1}{N_\mathrm{T}}\mathrm{tr}\left(\boldsymbol{P}_i\right) \underset{N_\mathrm{T}\to\infty}{\longrightarrow} 0.$$

$\frac{1}{N_\mathrm{T}}\mathrm{tr}\left(\boldsymbol{P}_i\right)$ computes according to

$$\frac{1}{N_\mathrm{T}}\mathrm{tr}\left(\boldsymbol{P}_i\right) = \frac{1}{N_\mathrm{T}}\left[\mathrm{tr}\left(\boldsymbol{I}_{N_\mathrm{T}}\right) - \sum_{j=1}^{i-1}\boldsymbol{t}_j^\mathrm{H}\boldsymbol{t}_j\right] = 1 - \frac{(i-1)}{N_\mathrm{T}},$$

where the last equality follows from the fact that the vectors $\boldsymbol{t}_j$ are orthonormal [c.f. (21)]. The large system channel gains therefore converge to an asymptotic deterministic limit independent of the current channel realization. Thus, as in the MIMO case in Section IV-B, water-filling can be applied over these asymptotic gains and the large system limit of a lower bound for sum rate achievable with spatial zero-forcing and DPC can therefore be computed according to

$$\frac{R^\mathrm{lb}_\text{ZF DPC}}{N_\mathrm{T}} - \frac{1}{N_\mathrm{T}}\sum_{i=1}^{i_\text{max}}\log_2\left(\hat{\eta}\left(1 - \frac{i}{N_\mathrm{T}}\right)\right) \underset{N_\mathrm{T}\to\infty, N_\mathrm{R}\to\infty}{\longrightarrow} 0 \tag{36}$$

with the modified water-level $\hat{\eta} = \eta N_\mathrm{T} = \frac{N_\mathrm{T}}{i_\text{max}}\left(P_\text{Tx} + \sum_{i=1}^{i_\text{max}}\frac{1}{1-\frac{i}{N_\mathrm{T}}}\right)$. i_max is chosen according to water-filling so that first the implicit equation $\hat{\eta} = \frac{1}{1-\frac{i_\text{max}}{N_\mathrm{T}}}$ is solved. In case this equation leads to a solution $i_\text{max} \leq \min\left(N_\mathrm{T}, KN_\mathrm{R}\right)$, i_max is given this way, otherwise by $i_\text{max} = \min\left(N_\mathrm{T}, KN_\mathrm{R}\right)$. Note that due to the zero-forcing constraints in case $\alpha < K$, i.e., $KN_\mathrm{R} > N_\mathrm{T}$, not all users are served, whereas with the optimum algorithm all users receive non-zero power (see Lemma 1). Replacing the infinite sums in (36) and the water-level by integrals and introducing the finite variable $\theta = \frac{i_\text{max}}{N_\mathrm{T}}$, the asymptotic lower bound $R^\mathrm{lb}_\text{ZF DPC}$ is given by

$$\frac{R^\mathrm{lb}_\text{ZF DPC}}{N_\mathrm{T}} \underset{\substack{N_\mathrm{T}\to\infty, N_\mathrm{R}\to\infty\\ N_\mathrm{T}=\alpha N_\mathrm{R}}}{\longrightarrow} \theta\log_2\left(P_\text{Tx} + \int_0^\theta \frac{1}{1-\rho'}\,\mathrm{d}\,\rho'\right) - \\ - \theta\log_2\left(\theta\right) + \int_0^\theta \log_2\left(1-\rho'\right)\mathrm{d}\,\rho'.$$

Evaluating the integrals leads to the system of equations (22).

Appendix E
Proof of Proposition 3

The derivation of (24) is based on the fact that the gains of the scalar subchannels $\lambda_{i,\text{ZF}} = \left(\boldsymbol{e}_i^T\left(\boldsymbol{H}_\text{comp}\boldsymbol{H}^\mathrm{H}_\text{comp}\right)^{-1}\boldsymbol{e}_i\right)^{-1}$ all converge to the same asymptotic limit

$$\frac{1}{N_\text{stop}}\left(\boldsymbol{e}_i^T\left(\boldsymbol{H}_\text{comp}\boldsymbol{H}^\mathrm{H}_\text{comp}\right)^{-1}\boldsymbol{e}_i\right)^{-1} \underset{\substack{N_\mathrm{T}, N_\mathrm{R}\to\infty\\ N_\text{stop}=\gamma N_\mathrm{T}, N_\mathrm{T}=\alpha N_\mathrm{R}}}{\longrightarrow} \left(\frac{1}{\gamma} - 1\right) \tag{37}$$

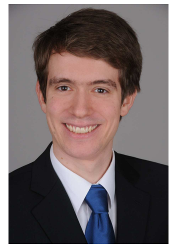
where $\gamma = \frac{N_\text{stop}}{N_\mathrm{T}} \leq \min\left(1, \frac{K}{\alpha}\right)$ denotes the normalized number of subchannels that receive nonzero powers. The large system limit in (37) can be obtained as described after Equations (10) and (11) in [15]. As all channel gains are asymptotically equal and deterministic, the optimum asymptotic power allocation is given by $p_{i,\text{ZF}} = \frac{P_\text{Tx}}{N_\text{stop}}$. By finally finding the optimum $N_\text{stop} = \gamma N_\mathrm{T}$ leading to the maximum sum rate, the asymptotic bound in (24) can be found. The assumption of $\boldsymbol{H}_\text{comp}$ containing i.i.d. entries is only valid for an arbitrary, i.e., not optimized selection of active users. That is because an optimum user selection implies that the active users are chosen so that the rows in $\boldsymbol{H}_\text{comp}$ are as orthogonal to each other as possible and therefore not independent. Thus, Equation (24) states a lower bound for the achievable sum rate in the large system limit. In case it is asymptotically optimum to serve all users in the system or a random user selection is applied instead of a greedy one, this assumption is valid, which is why in these cases the bound becomes tight.

References

[1] N. Jindal, W. Rhee, S. A. Vishwanath, S. Jafar, and A. Goldsmith, "Sum Power Iterative Water-filling for Multi-antenna Gaussian Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, April 2005.

[2] W. Yu, "Sum-Capacity Computation for the Gaussian Vector Broadcast Channel," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 754–759, February 2006.

[3] R. Hunger, D. Schmidt, and W. Utschick, "Sum-Capacity and MMSE for the MIMO Broadcast Channel without Eigenvalue Decompositions," in *International Symposium on Information Theory*, June 2007.
[4] G. Caire and S. Shamai, "On the Achievable Throughput of Multi-antenna Gaussian Broadcast Channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.
[5] M. Costa, "Writing on Dirty Paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
[6] P. Tejera, W. Utschick, G. Bauch, and J. A. Nossek, "Subchannel Allocation in Multiuser Multiple Input Multiple Output Systems," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4721–4733, October 2006.
[7] Z. Tu and R. Blum, "Multiuser Diversity for a Dirty Paper Approach," *IEEE Communications Letters*, vol. 7, no. 8, pp. 370–372, August 2003.
[8] G. Dimić and N. Sidoropoulos, "On Downlink Beamforming with Greedy User Selection," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3857–3868, October 2005.
[9] T. Yoo and A. Goldsmith, "On the Optimality of Multiantenna Broadcast Scheduling Using Zero-Forcing Beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, March 2006.
[10] C. Guthy, W. Utschick, and G. Dietl, "Low Complexity Linear Zero-Forcing for the MIMO Broadcast Channel," *IEEE Journal of Selected Topics in Signal Processing, special issue on Managing Complexity in Multiuser MIMO Systems*, vol. 3, no. 6, pp. 1106–1117, December 2009.
[11] A. Lozano and A. Tulino, "Capacity of Multiple-Transmit Multiple-Receive Antenna Architectures," *IEEE Transactions on Information Theory*, vol. 48, no. 12, pp. 3117–3128, December 2002.
[12] P. Kazakopoulos, P. Mertikopolous, A. Moustakas, and G. Caire, "Distribution of MIMO Mutual Information: A Large Deviations Approach," in *Proc. of IEEE Information Theory Workshop on Networking and Information Theory (ITW)*, 2009, pp. 306–310.
[13] V. Nguyen and J. Evans, "Multiuser Transmit Beamforming via Regularized Channel Inversion: A Large System Analysis," in *Proc. of Global Telecommunications Conference (GLOBECOM)*, 2008.
[14] C. Vaze and M. Varanasi, "A large-system Analysis of the Imperfect-CSIT Gaussian Broadcast Channel with a DPC-based Transmission Strategy," in *Proc. of 44th Annual Conference on Information Sciences and Systems (CISS)*, 2010.
[15] H. Huh, A. Tulino, and G. Caire, "Network MIMO Large-System Analysis and the Impact of CSIT Estimation," in *Proc. of 44th Annual Conference on Information Sciences and Systems (CISS)*, 2010.
[16] S.-H. Moon, H. Huh, Y.-T. Kim, G. Caire, and I. Lee, "Weighted Sum Rate of Multi-Cell MIMO Downlink Channels in the Large System Limit," in *Proc. of International Conference on Communications (ICC)*, 2010.
[17] H. Huh, G. Caire, S.-H. Moon, and I. Lee, "Multi-cell MIMO Downlink with Fairness Criteria: the Large System Limit," in *Proc. of International Symposium on Information Theory (ISIT)*, 2010, pp. 2058 – 2062.
[18] R. Zakhour and S. Hanly, "Base Station Cooperation on the Downlink: Large System Analysis," *IEEE Transactions on Information Theory*, vol. 58, no. 4, pp. 2064–2078, April 2012.
[19] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, Achievable Rates, and Sum-Rate Capacity of Gaussian MIMO Broadcast Channels," *IEEE Transactions on Information Theory*, vol. 49, pp. 2658–2668, 2003.
[20] A. Tulino and S. Verdu, *Random Matrix Theory and Wireless Communications*, ser. Foundations and Trends in Communications and Information Theory. Now Publishers, 2004, no. 1.
[21] F. Hildebrand, *Introduction to Numerical Analysis*, ser. International Series in Pure and Applied Mathematics. McGrawHill Book Company, 1956.
[22] M. Bazaraa, H. Sherali, and C. Shetty, *Nonlinear Programming – Theory and Applications*, 3rd ed. Wiley Interscience, 2006.
[23] J. Evans and D. Tse, "Large System Performance of Linear Multiuser Receivers in Multipath Fading Channels," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 2059–2078, September 2000.
[24] R. Speicher, "Free Convolution and the Random Sum of Matrices," *Publ. Res. Inst. Math. Sc.*, vol. 29, pp. 731–744, 1993.



Christian Guthy was born in Munich, Germany in 1979. He received the B.Sc., Dipl.-Ing., and Dr.-Ing. degrees from the Technische Universität München (TUM), Munich, Germany in 2004, 2005, and 2012, respectively.

From 2005 to 2011 he has been with the Associate Institute for Signal Processing at TUM. Since 2011 he is working as an development engineer for digital signal processing in industry. His main research interests include design and analysis of low complexity signal processing algorithms for next generation wireless communication systems with focus on multiantenna and multicarrier systems.

Wolgang Utschick (SM'06) was born on May 6, 1964 in Germany. He completed several industrial education programs before he received the high-school degree as well as the diploma and doctoral degrees, both with honors, in Electrical Engineering from the Technische UniversitÌÛt MÌÎnchen (TUM) in 1993 and 1998. In this period, he held a scholarship of the Bavarian Ministry of Education for exceptional students. From 1998 to 2002, he co-directed the Signal Processing Group of the Institute of Circuit Theory and Signal Processing, TUM. Since 2000, he has been consulting in 3GPP standardization in the field of multi-element antenna systems. In 2002, he was appointed Professor and Director of the Associate Institute for Signal Processing at the TUM. In 2007, Wolfgang received the ITG Award of the German Society for Information Technology (ITG). He is a senior member of the German VDE/ITG and was appointed member of the ITG Expert Committee for Information and System Theory in 2009. He is currently also serving as a chairman of the national DFG Focus Program 'Communications in Interference Limited Networks' (COIN). Wolfgang is an Associate Editor for the IEEE Transactions on Signal Processing and an editor and founder of the Springer Book Series 'Foundations in Signal Processing, Communications and Networking'. Since 2010, he is appointed member of the Technical Committee of Signal Processing for Communications and Networking in the IEEE Signal Processing Society.

Michael L. Honig received the B.S. degree in electrical engineering from Stanford University in 1977, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1978 and 1981, respectively. He subsequently joined Bell Laboratories in Holmdel, NJ, where he worked on local area networks and voiceband data transmission. In 1983 he joined the Systems Principles Research Division at Bellcore, where he worked on Digital Subscriber Lines and wireless communications. Since the Fall of 1994, he has been with Northwestern University where he is a Professor in the Electrical and Computer Engineering Department. He has held visiting scholar positions at the Naval Research Laboratory (San Diego), the University of California, Berkeley, the University of Sydney, Princeton University, and the Technical University of Munich. He has also worked as a free-lance trombonist.
Dr. Honig has served as an editor for the IEEE Transactions on Information Theory (1998-2000) and the IEEE Transactions on Communications (1990-1995), and as guest editor for the Journal on Selected Topics in Signal Processing, European Transactions on Telecommunications, and Wireless Personal Communications. He has also served as a member of the Digital Signal Processing Technical Committee for the IEEE Signal Processing Society, and as a member of the Board of Governors for the Information Theory Society (1997-2002). He is a Fellow of IEEE, the recipient of a Humboldt Research Award for Senior U.S. Scientists, and the co-recipient of the 2002 IEEE Communications Society and Information Theory Society Joint Paper Award and the 2010 IEEE Marconi Prize Paper Award.