On Overloaded Vector Precoding for Single-User MIMO Channels

Rodrigo de Miguel, Vesna Gardašević, Ralf R. Müller, *Senior Member, IEEE*, and Finn F. Knudsen

***Abstract*—We address the possibility of overloaded vector precoding in single user MIMO channels, *i.e.* the number of data streams is larger than the minimum of the number of antennas at transmitter and receiver side. We find that the convex vector precoding introduced in [1] allows for overloading while, with a certain probability, keeping the received signal free of interference. We find that the probability that overloading is not possible decays exponentially with the size of the system as long as the number of data streams is less than twice the minimum number of antennas. We give an explicit formula to calculate this probability for any antenna configuration in presence of correlated Rayleigh fading. Although overloading comes with the need for increased transmitted power, we show by means of the replica method that overloading up to 22% yields better spectral and power efficiency than without overload and spatial matched filter processing at the receiver.**

***Index Terms*—MIMO, channel inversion, zero-forcing, non-linear vector precoding, singular channels, asymptotic analysis, replica method.**

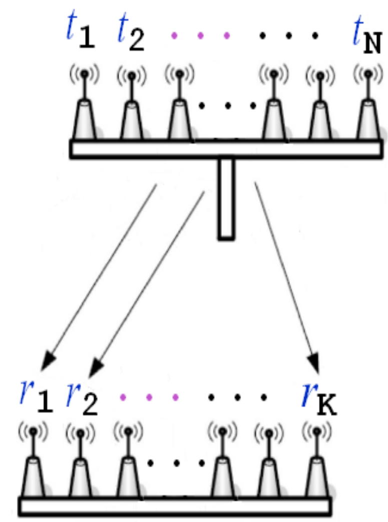


Fig. 1. A single user MIMO channel with N transmitting antennas and K collocated receiving antennas.

I. Introduction

In multiple-input/multiple-output (MIMO) channels information is conveyed simultaneously from a group of transmitting antennas to a group of receiving antennas. As these transmissions are not orthogonal, yet they occur over the same physical medium and bandwidth, crosstalk becomes unavoidable. As a result signal processing is needed at the receiver and/or transmitter side of the channel if significant data rates are to be achieved. In single user MIMO channels, depicted in Fig. 1, both the transmitting and the receiving antennas are collocated and they can jointly generate, preprocess, and post-process the data streams.

However, in the context of low cost receivers with limited processing power, it might be advantageous to shift most of the signal processing to the transmitter side. One technique which might be employed by the transmitter in order to keep the receivers from doing any signal processing is channel inversion before transmission (provided that the transmitter has complete channel state information). If the number of transmit antennas is larger than the number of receive antennas, direct channel inversion is not posssible. However, this problem can be overcome modifying the channel inversion process and applying a spatial matched filter at the receiver.

Regardless of whether a spatial matched filter is applied at the receiver or not, channel inversion at the transmitter comes with the need for more transmitted power. One popular technique which may be used to contain the transmit power while inverting the channel is non-linear vector precoding (henceforth vector precoding) [1]–[5]. Most such methods are based upon a redundant lattice representation of the information to be transmitted [2]–[5]. Although the optimum algorithm for finding the lattice point leading to the smallest transmitted energy is NP-hard, non-linear precoding based on redundant lattice representation has become feasible in practice with the help of sphere decoding [6] and lattice basis reduction algorithms [7], [8]. However, all these lattice-based algorithms are limited to channel matrices with full rank, *e.g.* the number of data streams to be not greater than the minimum of the number of antenna elements.

In [1], one of the authors proposed *convex precoding*, a different precoding technique not based on lattice extensions of the symbol constellations. Convex precoding was originally proposed to allow for convex optimization procedures to find the transmitted data vector with the minimum transmitted energy. Now, we have discovered that convex precoding al-

Manuscript received February 6, 2009; revised July 7, 2009; accepted September 15, 2009. The associate editor coordinating the review of this paper and approving it for publication was L. Lampe.

R. de Miguel, V. Gardašević, and R. R. Müller are with the Department of Electronics and Telecommunications, The Norwegian University of Science and Technology, Trondheim, Norway (e-mail: {demiguel, vesna.gardasevic, ralf}@iet.ntnu.no).

F. F. Knudsen is with the Department of Mathematical Sciences, The Norwegian University of Science and Technology, Trondheim, Norway (e-mail: knudsen@math.ntnu.no).

This material is based upon work partly supported by the Research Council of Norway under Grant No. 171133/V30. This paper was presented in part at the 2008 IEEE International Zurich Seminar on Communications, Zurich, Switzerland, the 2008 IEEE International Symposium on Applied Sciences in Biomedical and Communication Technologies, Aalborg, Denmark, and the 2008 IEEE International Symposium on Information Theory and its Applications, Auckland, New Zealand.

Digital Object Identifier 10.1109/TWC.2010.02.090179

lows, in principle, for vector precoding in cases where the channel matrix is rank deficient (the number of data streams is larger than the minimum of the number of antenna elements) while neither creating crosstalk nor the need for any particular signal processing at the receiver side. We call this the ability of convex precoding to deal with *overload*. While convex precoding falls slightly behind lattice-based precoding in terms of uncoded transmitted power, the possibility for overload makes it an attractive candidate for MIMO systems where the number of receive antennas is less than the number of transmit antennas, or in cases where the MIMO channel shows rank deficiencies due to key-hole effects. Furthermore, convex precoding was recently shown to outperform lattice-based precoding at low-signal-to-noise ratios if embedded into strong outer forward error-correction coding [17], [18].

In [1], the possibility for overload was only found in the large-system limit, *i.e.* the number of antennas at both ends of the link grow towards infinity, and with the help of the non-rigorous replica method. This left open the question of whether the possibility for overloading is more than an artefact of the large-system limit and the replica-symmetry assumed in [1] for sake of analytical tractability. In this paper we show that overload, with high probability, is indeed possible even in finite sized systems. Moreover, we explicitly calculate this probability by rigorous mathematical means as a function of the number of antennas for a MIMO channel with arbitrary (even not necessarily Kronecker-type [9]) correlations.

Since there is a fundamental trade-off between spectral efficiency and power efficiency on additive noise channels [11] and overloading increases the spectral efficiency, overloading must come at some increased need for transmitted power. Comparing overloaded precoding to non-overloaded precoding with spatial matched filtering (SMF) at the receiver, we find in this work that for overloads of up to 22%, the transmitted energy stays lower than for SMF while the spectral efficiency is higher. This result suggests that, also in practice, overloading is a sensible way to go under certain circumstances.

This paper is organized as follows. The vector precoding technique is presented in Section II. In Section III we find the probability that interference-free transmision in overloaded MIMO channels is possible. Section IV gives large-system results for the transmitted energy with convex vector precoding with and without SMF. In order to address SMF, we present a non-trivial extension of the results presented in [1]. Finally, numerical results and quantitative comparisons are presented in Section V. In order to not affect readability, we relegate some required technical derivations and proofs to the Appendices.

II. Vector Precoding

The single-user MIMO channel may be represented by the following vector equality:

$$\boldsymbol{r}[\mu] = \frac{\boldsymbol{H}\boldsymbol{t}[\mu]}{\frac{1}{N}\sqrt{\boldsymbol{t}^\dagger[\mu]\boldsymbol{t}[\mu]}} + \boldsymbol{n}[\mu], \tag{1}$$

where $\boldsymbol{t}[\mu] \in \mathbb{C}^{N\times 1}$ (up to a normalization that compensates for an eventually increased energy per symbol due to precoding) contains the signals transmitted at the N elements of the transmitting antenna array at discrete time μ. The column vector $\boldsymbol{r}[\mu] \in \mathbb{C}^{K\times 1}$ contains the signal received at the K elements of the receiving antenna array and $\boldsymbol{n}[\mu] \in \mathbb{C}^{K\times 1}$ is a random vector containing additive noise components. The matrix $\boldsymbol{H} \in \mathbb{C}^{K\times N}$ accounts for the propagation on the MIMO channel. In the following, we drop the discrete time index μ as we consider channels and signals without memory.

A. Standard Zero-Forcing

In order to reduce signal processing at the receiver site to a minimum, the transmitted vector $\boldsymbol{t}$ is formed by a linear transformation of the vector of PSK modulated information symbols $\boldsymbol{x} \in \{z \in \mathbb{C} : |z| = 1\}^{K\times 1}$ such that there is no crosstalk in the received vector $\boldsymbol{r}$. This goal is achieved by the linear transformation

$$\boldsymbol{t} = \lim_{\epsilon\to 0} \boldsymbol{H}^\dagger \left(\boldsymbol{H}\boldsymbol{H}^\dagger + \epsilon\mathbf{I}\right)^{-1} \boldsymbol{x} \tag{2}$$

where $\cdot^\dagger$ denotes conjugate transpose and $\mathbf{I}$ denotes the identity matrix. In that case, we simply get

$$\boldsymbol{r} = \boldsymbol{x} + \boldsymbol{n} \tag{3}$$

and all signal impairment is solely due to additive noise. The avoidance of crosstalk comes at the cost of an increased transmitted energy per symbol which is given by

$$E_s = \frac{1}{K}\lim_{\epsilon\to 0} \boldsymbol{x}^\dagger \left(\boldsymbol{H}\boldsymbol{H}^\dagger + \epsilon\mathbf{I}\right)^{-1} \boldsymbol{H}\boldsymbol{H}^\dagger \left(\boldsymbol{H}\boldsymbol{H}^\dagger + \epsilon\mathbf{I}\right)^{-1} \boldsymbol{x}. \tag{4}$$

If $\boldsymbol{H}\boldsymbol{H}^\dagger$ is invertible, the energy per symbol is finite for all symbol vectors $\boldsymbol{x}$ and can be simplified to read

$$E_s^{\text{inv}} = \frac{\boldsymbol{x}^\dagger \left(\boldsymbol{H}\boldsymbol{H}^\dagger\right)^{-1} \boldsymbol{x}}{K}. \tag{5}$$

However, the energy per symbol can still be finite even if $\boldsymbol{H}\boldsymbol{H}^\dagger$ is singular. This is, if and only if the symbol vector $\boldsymbol{x}$ lies within the span of $\boldsymbol{H}\boldsymbol{H}^\dagger$. What is more, no crosstalk will occur in that case, since only those components of $\boldsymbol{x}$ that lie in the nullspace of $\boldsymbol{H}\boldsymbol{H}^\dagger$ would get distorted.

B. Spatial Matched Filtering

If the matrix $\boldsymbol{H}\boldsymbol{H}^\dagger$ is singular, standard zero-forcing (SZF) works only for particular data symbols and not in general. However, if $\boldsymbol{H}\boldsymbol{H}^\dagger$ is singular, it might well be that $\boldsymbol{H}^\dagger\boldsymbol{H}$ is still invertible. Actually, for i.i.d. Rayleigh fading channels this even happens with probability one. This gives motivation to address a precoding scheme where the finiteness of the energy per symbol hinges on the existence of the inverse of $\boldsymbol{H}^\dagger\boldsymbol{H}$:

Consider now the transmitted vector to be formed by the transformation

$$\boldsymbol{t} = \lim_{\epsilon\to 0} \left(\boldsymbol{H}^\dagger\boldsymbol{H} + \epsilon\mathbf{I}\right)^{-1} \boldsymbol{x} \tag{6}$$

instead of transformation (2). Note that in contrast to Section II-A, here $\boldsymbol{x} \in \mathbb{C}^{N\times 1}$. The energy per symbol is given by

$$E_s = \frac{1}{N}\lim_{\epsilon\to 0} \boldsymbol{x}^\dagger \left(\boldsymbol{H}^\dagger\boldsymbol{H} + \epsilon\mathbf{I}\right)^{-2} \boldsymbol{x} \tag{7}$$

in this case. The limit exists whenever $\boldsymbol{x} \in \operatorname{span}(\boldsymbol{H}^\dagger\boldsymbol{H})$ even if $\boldsymbol{H}^\dagger\boldsymbol{H}$ is not invertible.

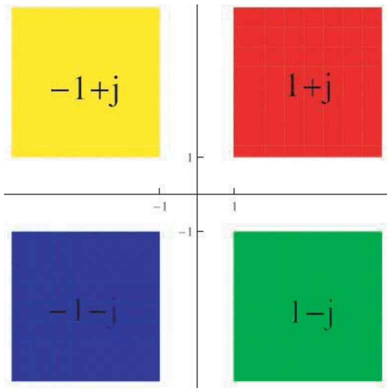


Fig. 2. Convex alphabet relaxation for QPSK.

The pre-processing in (6) does not avoid crosstalk at the receiver without post-processing. However, post-processing is reduced to a small margin: in order to get an estimate for the data vector, we have to calculate

$$\hat{\boldsymbol{x}} = \boldsymbol{H}^\dagger \boldsymbol{r}. \tag{8}$$

Although this requires the multiplication of a matrix with a vector, it does not require more complicated linear algebra like matrix inversions in linear MMSE (minimum mean-squared error) post-processing or singular value decompositions.

C. Redundant Signal Representation

No matter whether one uses standard zero-forcing or spatial matched filtering, the linear pre-processing comes with the need for more transmitted energy per symbol. In order to mitigate this effect and reduce the required transmit power, the symbol alphabet is relaxed into redundant sets, *i.e.* each point of a signal constellation is replaced by a set of points, before the linear pre-processing. Then, those representations of the data symbols are chosen which minimize the transmitted energy per symbol. In order to avoid degradation in terms of error rate performance, the sets are chosen in such way that the minimum distances between the sets are not lower than the minimum distance within the original signal constellation. The most popular way for such relaxations is to use disjoint sub-lattices of the complex integer lattice $\mathbb{Z}+\mathrm{j}\mathbb{Z}$ based on the Tomlinson-Harashima precoding (THP) strategy [12], [13], see *e.g.* [4], [14] and references therein.

Consider the alternative relaxation of the QPSK (quaternary phase-shift keying) constellation shown in Fig. 2. It is defined by the rule that the amplitudes for both quadrature components are allowed to be larger than for standard QPSK, but not smaller. This preserves the minimum distance and, as a nice side effect, makes the optimization procedure to find the signal representation with the smallest transmitted energy a convex optimization problem since, in contrast to lattice relaxations, the sets in the relaxed constellation are convex.

In order to mathematically formalize the optimization procedure, let s_i denote the QPSK symbol to be communicated via the i^{th} antenna element and let $\mathcal{A}_{s_i}$ denote the set into which this symbol is relaxed. For convex precoding of QPSK, this means

$$\begin{aligned} \mathcal{A}_{1+\mathrm{j}} &= -\mathcal{A}_{-1-\mathrm{j}} = \{\xi : \Re\xi \geq 1 \wedge \Im\xi \geq 1\}, \\ \mathcal{A}_{1-\mathrm{j}} &= -\mathcal{A}_{-1+\mathrm{j}} = \{\xi : \Re\xi \geq 1 \wedge \Im\xi \leq -1\}. \end{aligned} \tag{9}$$

Furthermore define the Cartesian product of sets

$$\mathcal{A}_{\boldsymbol{s}} = \prod_i \mathcal{A}_{s_i}. \tag{10}$$

Then, the vector $\boldsymbol{x}$ that results in minimum transmitted energy after linear pre-processing is found by

$$\boldsymbol{x}^{\text{SZF}} = \lim_{\epsilon\to 0} \operatorname*{argmin}_{\tilde{\boldsymbol{x}}\in\mathcal{A}_{\boldsymbol{s}}} \tilde{\boldsymbol{x}}^\dagger \left(\boldsymbol{H}\boldsymbol{H}^\dagger + \epsilon\mathbf{I}\right)^{-1} \tilde{\boldsymbol{x}} \tag{11}$$

and

$$\boldsymbol{x}^{\text{SMF}} = \lim_{\epsilon\to 0} \operatorname*{argmin}_{\tilde{\boldsymbol{x}}\in\mathcal{A}_{\boldsymbol{s}}} \tilde{\boldsymbol{x}}^\dagger \left(\boldsymbol{H}^\dagger\boldsymbol{H} + \epsilon\mathbf{I}\right)^{-2} \tilde{\boldsymbol{x}} \tag{12}$$

for standard zero-forcing and spatial matched filtering, respectively.

If the relaxed sets are lattices, then (11) and (12) have exponential complexity and become prohibitively expensive when the number of antenna elements is large [6]. This holds even when lattice reduction algorithms are used [15]. However, if the relaxed sets are convex as in Fig. 2, then efficient polynomial time algorithms might be used to find the optimal vector $\boldsymbol{x}$ [16]. Moreover, it has recently been shown that, when compared to lattice precoding, convex precoding schemes are competitive in terms of the achievable power-bandwidth trade-off when strong outer forward error correction coding is applied [17], [18].

Another advantage of convex precoding is that, as opposed to lattice precoding, it allows for overloading. If a higher dimensional vector is defined in a completely discrete space (*e.g.* a relaxed THP lattice), it has zero probability of lying in a random lower dimensional space; as a result relaxations on a lattice do not allow for overloading, *i.e* the relaxed vector cannot have higher dimension than the space spanned by the channel if interference-free transmission is to ever occur. However, relaxations on a continuous space allow for the possibility of tuning the components in the transmitted vector until it lies in the random lower dimensional space defined by the channel. In Section III we shall derive a rigorous formula for the probability of successful overloading when continuous sets as in Fig. 2 are used to relax a QPSK constellation; that is we shall derive the probability of having the set $\mathcal{A}_{\boldsymbol{s}}$ from eqs. (9) and (10) intersect with the space spanned by a lower dimensional random channel $\boldsymbol{H}$.

III. Failure Probability in Singular Channels

If the matrix $\boldsymbol{H}\boldsymbol{H}^\dagger$ or the matrix $\boldsymbol{H}^\dagger\boldsymbol{H}$ is not invertible in case of standard zero-forcing or spatial matched filtering, respectively, we call the channel *singular*. For singular channels, the limits and the minimizations in (11) and (12) do not commute. Furthermore, it is not sure whether the limits actually exist. As discussed in the previous section, the existence of the limits requires that the vector $\boldsymbol{x}$ lies within the span of $\boldsymbol{H}\boldsymbol{H}^\dagger$ and $\boldsymbol{H}^\dagger\boldsymbol{H}$, respectively. Thus, the limit exists if and only if the respective span intersects with the relaxed set $\mathcal{A}_{\boldsymbol{s}}$.

Both the respective span and the relaxed set $\mathcal{A}_{\boldsymbol{s}}$ are random variables. There will be a certain probability that they do not intersect leading to failure, *i.e.* the intended data cannot be transmitted due to unfavorable channel conditions. In this context, however, unfavorable channel conditions must be seen

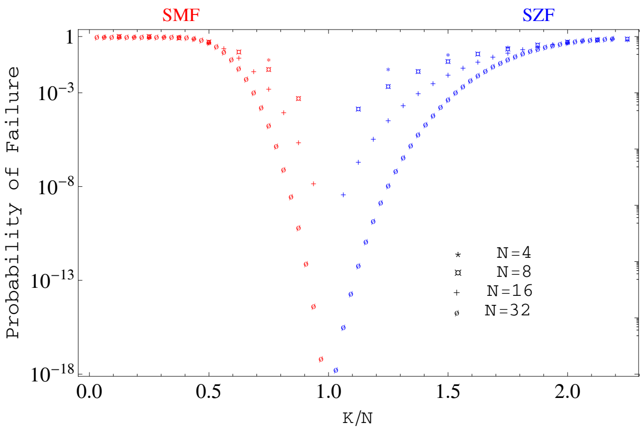


Fig. 3. Probability that convex precoding fails to find a solution vs. the ratio of the number of antennas at the receiver to that at the transmitter.

with respect to the data to be transmitted. It may well be that, with the same channel realization, other data do not lead to failure.

It is the purpose of this section to calculate the probability of failure. While for non-singular channels this probability is obviously 0, for a large class of singular channels it is given by the following theorem:

THEOREM 1 *Let $\boldsymbol{H}$ be an $K \times N$ random matrix whose distribution is absolutely continuous on $\mathbb{C}^{K\times N}$ with respect to Lebesgue measure. Let $\boldsymbol{s} \in \mathbb{C}^{N\times 1}$ be a random vector with i.i.d. entries uniformly distributed over the QPSK alphabet and let $\mathcal{A}_{\boldsymbol{s}}$ be defined by (9) and (10). Then,*

$$\Pr\left(\mathcal{A}_{\boldsymbol{s}} \cap \operatorname{span}(\boldsymbol{H}\boldsymbol{H}^\dagger) = \emptyset\right) = 1 - 2^{1-2K} \sum_{\ell=0}^{2N-1} \binom{2K-1}{\ell}. \tag{13}$$

Theorem 1 is proven in Appendix A. Note that the conditions for Theorem 1 are very weak. They include the Kronecker model [9] with full-rank correlation matrices and even MIMO channel models with correlations that do not fit into the Kronecker model framework as proposed in[10].

Replacing $\boldsymbol{H}$ by its conjugate transpose in Theorem 1 leads to the failure probability for spatial matched filtering:

COROLLARY 1 *Let $\boldsymbol{H}$ be a $K \times N$ random matrix whose distribution is absolutely continuous on $\mathbb{C}^{K\times N}$ with respect to Lebesgue measure. Let $\boldsymbol{s} \in \mathbb{C}^{K\times 1}$ be a random vector with i.i.d. entries uniformly distributed over the QPSK alphabet and let $\mathcal{A}_{\boldsymbol{s}}$ be defined by (9) and (10). Then,*

$$\Pr\left(\mathcal{A}_{\boldsymbol{s}} \cap \operatorname{span}(\boldsymbol{H}^\dagger\boldsymbol{H}) = \emptyset\right) = 1 - 2^{1-2N} \sum_{\ell=0}^{2K-1} \binom{2N-1}{\ell}. \tag{14}$$

Figure 3 shows the probability that the proposed convex precoding scheme fails to find a vector $\boldsymbol{x}$ which lies in the space spanned by the channel matrix. While for full rank matrices the probability of failure is zero, as the null space of the matrix increases finding a solution becomes less and less probable. However, significant overloads (50-80%) are possible while keeping the probability of failure relatively small. It should be noted that when K and N grow to infinity the probability collapses to a step function: interference-free transmission is always possible if the matrix is at least half rank and impossible if the rank is lower than half.

IV. The Many Antenna Limit

While the failure probability could be calculated exactly for any number of antenna elements, we provide only asymptotic results for the transmitted energy per symbol when the number of antenna elements grows large. Furthermore, we shall assume i.i.d. Rayleigh fading in order to keep the resulting mathematics at least somewhat contained. Hereby, the key parameter to describe the system setting will be the ratio of the number of antenna elements at the receiver to the corresponding number at the transmitter $\alpha = K/N$. Due to the mathematical similarity of uncorrelated MIMO systems with direct-sequence code-division multiple-access communications [20], the parameter α is called the *load* of the system.

In order to evaluate the performance in the many antenna limit, we make use of the analysis and the results of [1]. There it was shown[1] that for any bi-unitarily invariant matrix $\boldsymbol{J}$, we have

$$E_{\mathrm{s}} = \min_{\boldsymbol{x}\in\mathcal{A}_{\boldsymbol{s}}} \frac{\boldsymbol{x}^\dagger \boldsymbol{J}\boldsymbol{x}}{\dim(\boldsymbol{J})} \tag{15}$$

$$\longrightarrow q\frac{\partial}{\partial\chi}\chi R_{\boldsymbol{J}}(-\chi), \tag{16}$$

where the parameters q and χ are given by the following pair of coupled self-consistent equations

$$q = \sum_{s_i} \Pr(s_i) \int_{\mathbb{C}} \left| \operatorname*{argmin}_{\xi\in\mathcal{A}_{s_i}} \left| \frac{\sqrt{qR'_{\boldsymbol{J}}(-\chi)}}{R_{\boldsymbol{J}}(-\chi)} - \frac{\xi}{z} \right| \right|^2 \mathrm{D}z, \tag{17}$$

$$\chi = \sum_{s_i} \frac{\Pr(s_i)}{\sqrt{qR'_{\boldsymbol{J}}(-\chi)}} \Re \int_{\mathbb{C}} \operatorname*{argmin}_{\xi\in\mathcal{A}_{s_i}} \left| \frac{\sqrt{qR'_{\boldsymbol{J}}(-\chi)}}{R_{\boldsymbol{J}}(-\chi)} - \frac{\xi}{z} \right| z^*\mathrm{D}z \tag{18}$$

where $R_{\boldsymbol{J}}(\cdot)$ is the R-transform of the asymptotic eigenvalue distribution of the kernel matrix $\boldsymbol{J}$, $R'_{\boldsymbol{J}}(-\chi)$ denotes the first derivative of $R_{\boldsymbol{J}}(t)$ evaluated at $t = -\chi$, and $\mathrm{D}z = \exp(-|z|^2)/\pi \,\mathrm{d}\Re z \,\mathrm{d}\Im z$ denotes the complex Gaussian measure.

The R-transform for standard zero-forcing in i.i.d. Rayleigh fading was given in [1] and reads

$$R^{\mathrm{SZF}}(w) = \lim_{\epsilon\to 0} R_{(\boldsymbol{H}\boldsymbol{H}^\dagger+\epsilon\mathbf{I})^{-1}}(w) \tag{19}$$

$$= \frac{2}{1-\alpha+\sqrt{(1-\alpha)^2-4\alpha w}}. \tag{20}$$

For spatial matched filtering, the R-transform is not reported in literature. Its derivation is quite cumbersome and delegated to Appendix B. Note that the limits $\epsilon \to 0$ in (19) and (31) exist no matter whether $\boldsymbol{H}\boldsymbol{H}^\dagger$ and $\boldsymbol{H}^\dagger\boldsymbol{H}$ are invertible or not.

V. Results

Figure 4 shows, as a function of α, the energy per trans-

[1]The derivation of this result in [1] makes use of some tools from statistical mechanics and requires the technical assumption of replica symmetry. While replica symmetry is a questionable assumption for lattice-based relaxations, it is fully justified for any convex relaxation.

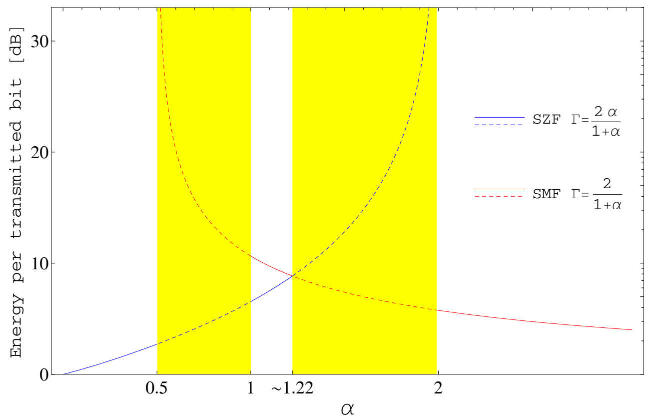


Fig. 4. Energy per transmitted bit vs. ratio of receiver to transmitter antennas, for SZF (blue curve) and SMF (red curve). In the shaded regions there is a tradeoff between spectral efficiency Γ and energy per transmitted bit.

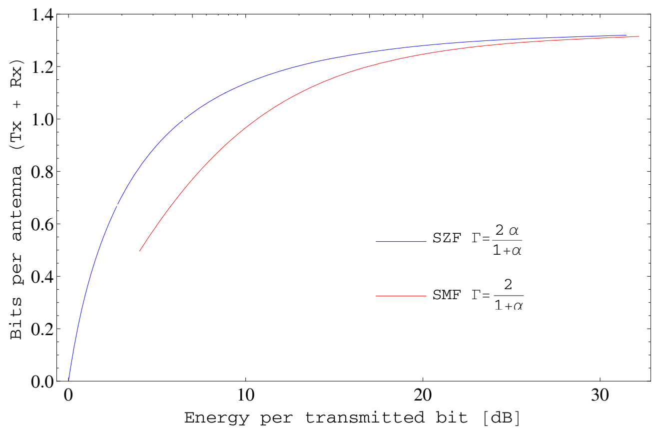


Fig. 5. Uncoded spectral efficiency vs. transmitted energy per bit. A given level of efficiency might be achieved at a lower cost using SZF.

mitted bit (15) for the SMF and SZF schemes proposed in Section II. As anticipated in Section III, in the large system limit a vector $\boldsymbol{x}$ cannot be found in the span when the rank of the matrix is less than 1/2. As a result when the ratio of antennas at the transmitter to that of the receiver is smaller than 0.5 SMF fails; analogously, only SMF might be employed when $\alpha > 2$. Figure 3 shows that this is nearly always the case also for finite systems.

In order to compare both schemes, we introduce the uncoded spectral efficiency Γ, which is the number of bits per antenna in the system (both transmitting and receiving antennas). While for SMF Γ equals $2/(1+\alpha)$, for SZF Γ equals $2\alpha/(1+\alpha)$.

When the ratio α of antennas at the receiver to that at the transmitter is between 1 and 1.22, greater Γ is achieved by the SZF scheme at a lower cost in energy per transmitted bit. However, as we can see in Fig. 3, in this range SZF has a slight probability of failure in finite systems; nevertheless, this probability is small and perhaps tolerable in systems with subsequent error control coding.

In the two remaining regions, namely $0.5 < \alpha < 1$ and $1.22 < \alpha < 2$, a tradeoff situation occurs between SZF and SMF. While one of the two schemes performs better with respect to energy efficiency, the other one has higher spectral efficiency. However, one must be careful when assessing this tradeoff under certain antenna configurations, namely when α is too close to 0.5 (only SZF is viable in realistic finite channels) or too close to 2 (only SMF is viable); see Fig. 3.

Depending on the priorities and resources of transmitter and receiver, other tradeoff situations might be of interest. For instance, looking at Fig. 4 we can see that, if the receiver has only slightly more than twice as many antennas as the transmitter, then some rate might be gained by turning off roughly half of the receiving antennas, such that $\alpha \in (1, 1.22)$, and switching to SZF mode; this gain in rate would come, however, at a moderate probability of failure and also a moderate additional energy penalty per bit for the transmitter.

For a given total number of antennas, using SZF with $\alpha = \alpha_0$ yields the same spectral efficiency as using SMF with $\alpha = 1/\alpha_0$. If the system designer has freedom to arrange the antenna configuration before deciding to use SZF or SMF, then it is useful to know that a given level of spectral efficiency can be achieved at lower energy cost using SZF. This relation between spectral efficiency and transmitted energy is shown in Fig. 5.

VI. Conclusion

We have demonstrated that convex precoding is an attractive method of choice for the downlink of future wireless communication systems. It performs well compared with other precoding methods like lattice precoding or linear precoding and it does not require regularization efforts. Furthermore, it can cope with singular channels (within certain limits) and makes deployed systems more flexible with respect to the change of antenna hardware and more robust against adverse channel conditions like keyholes.

Acknowledgments

The authors would like to thank Alexei Rudakov for a fruitful discussion about Theorem 1, Stanislaw J. Szarek for later hinting us towards reference [19], Nikolai Ushakov for checking the statistical notation of an earlier version, and Daniel Ryan for discussions on the complexity of lattice reduction methods for vector precoding.

Appendix A: Proof of Theorem 1

In 1962 J. G. Wendel solved the problem of determining the probability that when K points are scattered at random on the surface of the unit sphere in N-dimensional space, they all lie on some hemisphere. If $\boldsymbol{x}$ is a point on the sphere, let $-\boldsymbol{x}$ be the antipodal point. Wendel's ingenious idea was to extend the sample space to the Cartesian product of the K-fold $N{-}1$-sphere and the set $\{1, -1\}^K$. Then he considered the event that for a random sequence of points $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_K)$ and a random sequence $(\epsilon_1, \epsilon_2, \ldots, \epsilon_K)$, the points $(\epsilon_1\boldsymbol{x}_1, \epsilon_2\boldsymbol{x}_2, \ldots, \epsilon_K\boldsymbol{x}_K)$ lie on some hemisphere. It turns out that there is a fixed number $2f(N{-}1, K{-}1)$ of ϵ-sequences that lead to success for almost all sequences of points. This number was known to Ludwig Schläfli around 1850. Wendel rediscovered Schläfli's result and H. S. M. Coxeter told him about it. See Theorem 2 for the definition of $f(N, K)$, and [19] or [25] for a proof. In this appendix we will state and prove a complex version of Wendel's theorem.

Geometric Preliminaries

For any complex matrix $\boldsymbol{M}$ we can associate a real matrix $\boldsymbol{M}_r$, which we call the real version of $\boldsymbol{M}$. We define this as

$$\boldsymbol{M}_r = \begin{pmatrix} \Re\boldsymbol{M} & -\Im\boldsymbol{M} \\ \Im\boldsymbol{M} & \Re\boldsymbol{M} \end{pmatrix}.$$

It is well known that the real version satisfies the following relations

$$(\boldsymbol{L}\boldsymbol{M})_r = \boldsymbol{L}_r\boldsymbol{M}_r, \tag{21}$$

$$(\boldsymbol{M}^\dagger)_r = (\boldsymbol{M}_r)^\dagger, \tag{22}$$

$$\det(\boldsymbol{M}_r) = |\det\boldsymbol{M}|^2. \tag{23}$$

DEFINITION 1 *Let K and N be arbitrary positive integers. However, the interesting cases are when $K > N$. For any real $K\times N$-matrix M, we have a* constellation $\mathcal{C}(\boldsymbol{M})$ *of K hyperplanes through the origin in $\mathbb{R}^N$. An $N\times 1$-vector $\boldsymbol{w}$ belongs to $\mathcal{C}(\boldsymbol{M})$ if at least one of the components of the vector $\boldsymbol{M}\boldsymbol{w}$ vanishes. The complement $\overline{\mathcal{C}}(\boldsymbol{M}) = \mathbb{R}^N \setminus \mathcal{C}(\boldsymbol{M})$ consists of vectors $\boldsymbol{w}$ with the property that all the components of $\boldsymbol{M}\boldsymbol{w}$ are non-zero. The* good *constellations are the ones with the property that the intersection of any n of the hyperplanes with $n \leq N$ has dimension $N - n$.*

DEFINITION 2 *We say that a real matrix is* general *if all its maximal minors are non-zero, and* completely general *if all its square submatrices have full rank. A complex matrix will be called* general *or* completely general *if its real version is.*

Note that the constellation $\mathcal{C}(\boldsymbol{M})$ is good if and only if the matrix $\boldsymbol{M}$ is general. The following result was first obtained by Ludwig Schläfli [26]. More elaborated proofs can be found in any of the references [19] or [27].

THEOREM 2 *If $\boldsymbol{M}$ is a general $K\times N$-matrix, the number of connected components of $\overline{\mathcal{C}}(\boldsymbol{M})$ is $2f(N-1, K-1)$, where*

$$f(N,K) = \binom{K}{0} + \binom{K}{1} + \cdots + \binom{K}{N}. \tag{24}$$

Proof: By restricting the problem to one of the hyperplanes in $\mathcal{C}(\boldsymbol{M})$, one sees that for $K > N$, $f(N-1, K-1) + f(N, K-1) = f(N, K)$. For $K = N$ we have $f(K, K) = 2^K$, so the result follows by induction. ■

DEFINITION 3 *We will say that a complex number $z = x + \mathrm{j}y$ is* proper, *if $xy \neq 0$. These are the numbers that belong to any of the four open quadrants. A complex vector is proper if each component is proper.*

DEFINITION 4 *For a* complex *$K\times N$-matrix $\boldsymbol{M}$, the* constellation $\mathcal{C}(\boldsymbol{M})$ *and its complement $\overline{\mathcal{C}}(\boldsymbol{M}) = \mathbb{C}^K \setminus \mathcal{C}(\boldsymbol{M})$ are defined a little differently. In the complex case, $\overline{\mathcal{C}}(\boldsymbol{M})$ consists of the complex vectors $\boldsymbol{w}$ with the property that all the components of $\boldsymbol{M}\boldsymbol{w}$ are proper complex numbers.*

LEMMA 1 *If $\phi : \mathbb{C}^N \to \mathbb{R}^{2N}$ is the map defined by*

$$\phi(\boldsymbol{w}) = \begin{pmatrix} \Re\boldsymbol{w} \\ \Im\boldsymbol{w} \end{pmatrix},$$

we have

$$\phi(\mathcal{C}(\boldsymbol{M})) = \mathcal{C}(\boldsymbol{M}_r).$$

Proof: This follows immediately from the relations above. ■

COROLLARY 2 *If $\boldsymbol{M}$ is a complex $K\times N$-matrix such that $\boldsymbol{M}_r$ is general, the number of connected components of $\overline{\mathcal{C}}(\boldsymbol{M})$ is $2f(2N-1, 2K-1)$.* ■

Absolutely Continuously Distributed Matrices

Even if all maximal minors of a complex matrix are non-zero, the real version might not be general. We will show, however, that a large class of complex matrices are completely general.

DEFINITION 5 *We will say that a random complex $K\times N$-matrix $\boldsymbol{M}$ is* absolutely continuously distributed (a.c.d.)*, if its distribution on $\mathbb{C}^{KN}$ is absolutely continuous with respect to Lebesgue measure.*

LEMMA 2 *Let $\boldsymbol{M}$ be a matrix with variable entries $z_{i,j} = x_{i,j} + \mathrm{j}y_{i,j}$. Then for every square submatrix $\boldsymbol{D}$ of $\boldsymbol{M}_r$, the polynomial* $\det(\boldsymbol{D})$ *is of degree* $\dim(\boldsymbol{D})$.

Proof: Let $\boldsymbol{C}$ be a $j\times j$-submatrix of $\boldsymbol{M}_r$. The statement is clear for $j = 1$. We use induction with respect to j. If $\boldsymbol{C} = \boldsymbol{D}_r$, for some complex submatrix of $\boldsymbol{M}$, the statement follows from relation (23). If not, $\boldsymbol{C}$ has a row and a column consisting of variables not occurring only once in $\boldsymbol{C}$. Expanding the determinant along this row or column, the coefficients are by induction hypothesis all polynomials of degree $j - 1$, and since the row or column elements are all different the lemma follows. ■

LEMMA 3 *Any complex a.c.d. matrix $\boldsymbol{M}$ is completely general with probability 1.*

Proof: By Lemma 2, the probability that $\boldsymbol{M}$ is not general is the measure of a vanishing set of a finite set of non-zero polynomials. Such a set has Lebesgue measure equal to zero. By absolute continuity, this probability is also zero. ■

The Proof

Let GU be the group of Gaussian units $\mathrm{GU} = \{1, \mathrm{j}, -1, -\mathrm{j}\}$. For every proper complex number z there is a unique Gaussian unit which we denote by $g(z)$, such that $g^*(z)z$ belongs to the first quadrant in the complex plane.

LEMMA 4 *The assignment g is multiplicative in the sense that if ϵ is a Gaussian unit and z any proper complex number, then $g(\epsilon z) = \epsilon g(z)$.* ■

Let $\mathrm{GU}(K)$, be the K-fold product of GU. We will consider the left action of this group on the set of $K\times N$-matrices, for any K. We also denote by g the map from proper vectors in $\mathbb{C}^K$ to $\mathrm{GU}(K)$.

DEFINITION 6 *For any* $\boldsymbol{\epsilon} \in \mathrm{GU}(K)$*, we denote by* $\mathcal{A}_{\boldsymbol{\epsilon}} \subseteq \mathbb{C}^K$*, the set of proper vectors* $\boldsymbol{z}$ *such that* $g(\boldsymbol{z}) = \boldsymbol{\epsilon}$*.*

DEFINITION 7 *We will say that a* $K \times N$*-matrix* $\boldsymbol{M}$ *is a* winner *if there exists a vector* $\boldsymbol{w}$ *such that* $\boldsymbol{M}\boldsymbol{w} \in \mathcal{A}_{\mathbf{1}}$*. We denote by* $\mathcal{W}_{\mathbf{1}}$ *the set of winning matrices. More generally let* $\mathcal{W}_{\boldsymbol{\epsilon}} = \{\boldsymbol{M} : \exists \boldsymbol{w}, \boldsymbol{M}\boldsymbol{w} \in \mathcal{A}_{\boldsymbol{\epsilon}}\}$

DEFINITION 8 *We denote by* $\mathcal{W}' \subseteq \mathrm{GU}(K) \times \mathbb{C}^{K\times N}$ *the set of all pairs* $(\boldsymbol{\epsilon}, \boldsymbol{M})$ *such that* $\boldsymbol{\epsilon}\boldsymbol{M}$ *is a winner. For a fixed matrix* $\boldsymbol{M}$*, we denote by* $\mathcal{W}'_{\boldsymbol{M}}$ *the subset of pairs in* $\mathcal{W}'$ *whose second coordinate is* $\boldsymbol{M}$*.*

LEMMA 5 *For any complex* $K \times N$*-matrix* $\boldsymbol{M}$ *and any vector* $\boldsymbol{w} \in \overline{\mathcal{C}}(\boldsymbol{M})$*, we have*

$$g(g^*(\boldsymbol{M}\boldsymbol{w})\boldsymbol{M}\boldsymbol{w}) = \mathbf{1}.$$

In other words $g^*(\boldsymbol{M}\boldsymbol{w})\boldsymbol{M}$ *is a winner.* ■

Note that for any $K \times N$-matrix $\boldsymbol{M}$ it follows that if $\overline{\mathcal{C}}(\boldsymbol{M}) \neq \emptyset$, there is an $\boldsymbol{\epsilon} \in \mathrm{GU}(K)$ such that the matrix $\boldsymbol{\epsilon}\boldsymbol{M}$ is a winner. Moreover $\boldsymbol{M} \in \mathcal{W}_{\boldsymbol{\epsilon}}$ if and only if $\boldsymbol{\epsilon}^*\boldsymbol{M} \in \mathcal{W}_{\mathbf{1}}$.

LEMMA 6 *If* $\boldsymbol{M}$ *is a complex* $K \times N$*-matrix such that* $\boldsymbol{M}_r$ *is general, then the cardinality of the set* $\mathcal{W}'_{\boldsymbol{M}}$ *is given by* $|\mathcal{W}'_{\boldsymbol{M}}| = 2f(2N-1, 2K-1)$*.*

Proof: Consider the map $\overline{\mathcal{C}}(\boldsymbol{M}) \to \mathrm{GU}(K)$ defined by $\boldsymbol{w} \mapsto g(\boldsymbol{M}\boldsymbol{w})$. Since for a fixed $\boldsymbol{M}$ the map is linear and the open quadrants in the complex plane are convex, vectors in different components of $\overline{\mathcal{C}}(\boldsymbol{M})$ map to different elements in $\mathrm{GU}(K)$. Therefore the lemma follows from Corollary 2. ■

THEOREM 3 *If* $\boldsymbol{H}$ *is a random complex* $K \times N$*-matrix and* $\boldsymbol{\epsilon}$ *is a random element of* $\mathrm{GU}(K)$ *such that* $\boldsymbol{H}$ *and* $\boldsymbol{\epsilon}$ *are independent,* $\boldsymbol{H}$ *is general with probability 1 and* $\boldsymbol{\epsilon}$ *is uniformly distributed, then* $\Pr((\boldsymbol{\epsilon}, \boldsymbol{H}) \in \mathcal{W}') = q(N, K)$ *where*

$$q(N, K) = \frac{2f(2N-1, 2K-1)}{2^{2K}}. \tag{25}$$

Proof: Let $\boldsymbol{I}'$ be the indicator of $\mathcal{W}'$. Then $\Pr((\boldsymbol{\epsilon}, \boldsymbol{H}) \in \mathcal{W}') = \mathsf{E}(\boldsymbol{I}')$. By our assumptions, the conditional $\mathsf{E}(\boldsymbol{I}'|\boldsymbol{H}) = |\mathcal{W}'_{\boldsymbol{H}}|/|\mathrm{GU}(K)|$. It follows from Lemma 6 and our assumptions that the conditional $\mathsf{E}(\boldsymbol{I}'|\boldsymbol{H}) = q(N, K)$ with certainty. Hence $\mathsf{E}(\boldsymbol{I}') = q(N, K)$. ■

Note that the two statements $(\boldsymbol{\epsilon}^*, \boldsymbol{H}) \in \mathcal{W}'$ and $\mathrm{colspan}(\boldsymbol{H}) \cap \mathcal{A}_{\boldsymbol{\epsilon}} \neq \emptyset$ are equivalent.

DEFINITION 9 *We will say that a random complex* $K \times N$*-matrix* $\boldsymbol{M}$ *is a Wendel-matrix if it is general with probability 1 and its distribution is invariant under the action of* $\mathrm{GU}(K)$*.*

THEOREM 4 *Let* $\boldsymbol{M}$ *be a random complex* $K \times N$*-matrix. Then if* $\boldsymbol{M}$ *is a Wendel-matrix, the probability* $\Pr(\boldsymbol{M} \in \mathcal{W}) = q(N, K)$*.*

Proof: Let $\boldsymbol{I}$ be the indicator of the set $\mathcal{W}$ and let $\boldsymbol{I}'$ be the indicator of the set $\mathcal{W}'$. Then $\Pr(\boldsymbol{M} \in \mathcal{W}) = \mathsf{E}(\boldsymbol{I})$. Since the distribution of $\boldsymbol{M}$ is invariant under the action of $\mathrm{GU}(K)$, $\mathsf{E}(\boldsymbol{I}') = \mathsf{E}(\boldsymbol{I})$. ■

APPENDIX B: ASYMPTOTICS OF RANDOM MATRICES

Let $\mathrm{P}_{\boldsymbol{M}}(x)$ denote the eigenvalue distribution of the matrix $\boldsymbol{M}$. Let

$$m_{\boldsymbol{M}}(s) = \int \frac{\mathrm{dP}_{\boldsymbol{M}}(x)}{x - s}, \tag{31}$$

which is known as the Stieltjes transform. Then, the R-transform of $\mathrm{P}_{\boldsymbol{M}}(x)$ is

$$R_{\boldsymbol{M}}(w) = m_{\boldsymbol{M}}^{\mathrm{inv}}(-w) - \frac{1}{w}, \tag{32}$$

with $m^{\mathrm{inv}}(\cdot)$ denoting the inverse function of $m(\cdot)$.

As $K = \alpha N \to \infty$ (for α finite and positive) the matrix $\boldsymbol{H}^\dagger\boldsymbol{H}$ has a limiting eigenvalue distribution, namely the Marchenko Pastur distribution [21]–[23]: (33)
For matrix inversion purposes we shall neglect the contributions from the null eigenvalues. We then fully characterize this eigenvalue distribution by its Stieltjes transform, which is given by (34)

The eigenvalue distribution of $\left(\boldsymbol{H}^\dagger\boldsymbol{H} + \epsilon\mathbf{I}\right)^{-2}$ may also be characterized by its Stieltjes transform, which may be written in terms of the Stieltjes transform of $\boldsymbol{H}^\dagger\boldsymbol{H}$ as (35).

The R-transform of the eigendistribution may now be obtained explicitly using eq. (32), resulting in the somewhat tedious expression (36) with (37-45).

REFERENCES

[1] R. R. Müller, D. Guo, and A. Moustakas, "Vector precoding for wireless MIMO systems and its replica analysis," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 530-540, Apr. 2008.

[2] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*. John Wiley and Sons, 2002.

[3] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3. no. 4, pp. 1305-1316, 2004.

[4] B. M. Hochwald, C. Peel, and A. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537-544, Mar. 2005.

[5] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit procesing in MIMO communication systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 537-544, Aug. 2005.

[6] J. Jalden and B. Ottersten, "On the complexity of sphere decoding in digital communications," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1474-1484, Apr. 2005.

[7] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Trans. Commun.*, vol. 52, no. 12, pp. 2057-2060, Dec. 2004.

[8] M. Taherzadeh, A. Mobasher, and A. K. Kandhani, "Communication over MIMO broadcast channels using lattice-basis reduction," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4567-4582, Dec. 2007.

[9] C. Chuan, D. M. C. Tse, J. Kahn, and R. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," *IEEE Trans. Inf.. Theory*, vol. 48, no. 3, pp. 637–650, Mar. 2002.

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

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.

[12] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *IEE Electron. Lett.*, vol. 7, pp. 138-139, Mar. 1971.

[13] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. COM-20, pp. 774-780, Aug. 1972.

[14] R. F. H. Fischer, C. Windpassinger, A. Lampe, and J. B. Huber, "Space-time transmission using Tomlinson-Harashima precoding," in *Proc. 4th Int. ITG Conf.*, Berlin, Germany, Jan. 2002.

[15] D. Seethaler, J. Jalden, C. Studer, and H. Boelcskei, "Tail behavior of sphere-decoding complexity in random lattices," in *Proc. IEEE International Symp. Inf. Theory (ISIT)*, Seoul, Korea, June/July 2009.

$$\mathrm{dP}_{\boldsymbol{H}^\dagger\boldsymbol{H}}(x) = (1-\alpha)^+\,\delta(x)\mathrm{d}x + \frac{\alpha}{2\pi x}\sqrt{\left(x-\left(1-1/\sqrt{\alpha}\right)^2\right)^+\left(x-\left(1+1/\sqrt{\alpha}\right)^2\right)^+}\,\mathrm{d}x. \tag{28}$$

$$m_{\boldsymbol{H}^\dagger\boldsymbol{H}}(s) = \frac{\alpha}{2s}\left(1-\frac{1}{\alpha}-s+\sqrt{\frac{1}{\alpha^2}+(s-1)^2-\frac{2}{\alpha}(1+s)}\right). \tag{29}$$

$$m_{\left(\boldsymbol{H}^\dagger\boldsymbol{H}+\epsilon\mathbf{I}\right)^{-2}}(s) = \frac{1}{2s}\left(s^{-1/2}m_{\boldsymbol{H}^\dagger\boldsymbol{H}}\left(-s^{-1/2}\right)-s^{-1/2}m_{\boldsymbol{H}^\dagger\boldsymbol{H}}\left(s^{-1/2}\right)-2\right). \tag{30}$$

$$\begin{aligned} R^{\mathrm{SMF}}(w) &= \lim_{\epsilon\to 0} R_{\left(\boldsymbol{H}^\dagger\boldsymbol{H}+\epsilon\mathbf{I}\right)^{-2}}(w) \\ &= 2\left(\rho-\sqrt{\kappa+\lambda}-\sqrt{\kappa-\lambda+\rho^2-\delta-\frac{\epsilon+8\rho^3-24\mu}{4\sqrt{\kappa+\lambda}}}\right)^{-1}-\frac{1}{w} \end{aligned} \tag{31}$$

$$\epsilon = \frac{64\alpha^2 w^3}{1+\alpha+w}, \tag{32}$$

$$\delta = \frac{2\alpha w^2}{3\tau}(5+14\alpha+5\alpha^2), \tag{33}$$

$$\rho = \frac{w(1+7\alpha+7\alpha^2+\alpha^3+\alpha w)}{2\tau}, \tag{34}$$

$$\lambda = \frac{2\alpha^2 w^4\left((\alpha-1)^4+24\alpha w(\alpha+1)\right)+2^{1/3}\psi^{2/3}}{3\cdot 2^{2/3}\tau\psi^{1/3}}, \tag{35}$$

$$\mu = \frac{\alpha w^3}{6\tau^2}\left(5+\alpha(14+5\alpha)\right)\left(1+\alpha\left(7+\alpha(7+\alpha)+w\right)\right), \tag{36}$$

$$\kappa = \frac{w^2}{12}\frac{(\alpha^2-1)^2\left(3+\alpha(2+3\alpha)\right)-2\alpha w(1+\alpha)\left(17+\alpha(38+17\alpha)\right)+3\alpha^2w^2}{\tau^2}, \tag{37}$$

$$\tau = (1+\alpha)(1+\alpha+w), \tag{38}$$

$$\psi = 2\alpha^3 w^6\left(72\alpha w(\alpha-1)^2(\alpha+1)+54\alpha^2w^2-(\alpha-1)^6\right)+12\sqrt{3}\sqrt{\alpha^7 w^{13}}\eta, \tag{39}$$

$$\begin{aligned}\eta = \alpha w(\alpha-1)^4(31+\alpha(66+31\alpha))-2(\alpha-1)^8(\alpha+1)+ \\ +27\alpha^3w^3-8\alpha^2w^2(1+\alpha)(7+\alpha)(1+7\alpha).\end{aligned} \tag{40}$$

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

[17] B. M. Zaidel, R. R. Müller, R. de Miguel, and A. L. Moustakas, "On spectral efficiency of vector precoding for gaussian MIMO broadcast channels," in *Proc. IEEE International Symp. Spread Spectrum Techniques Applications (ISSSTA)*, Bologna, Italy, Aug. 2008.

[18] B. M. Zaidel, R. R. Müller, R. de Miguel, and A. L. Moustakas, "On replica symmetry breaking in vector precoding for the Gaussian MIMO broadcast channel," in *Proc. 46th Annual Allerton Conf. Commun., Control Computing*, Monticello, IL, USA, Sept. 2008.

[19] J. G. Wendel, "A problem in geometric probability," *Mathematica Scandinavica*, vol. 11, pp. 109-111, 1962.

[20] S. Verdú, *Multiuser Detection*. Cambridge University Press, 1998.

[21] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free Random Variables*. American Mathematical Society, 1992.

[22] R. R. Müller, "Applications of large random matrices in communications engineering," in *Proc. International Conf. Advances Internet, Processing, Syst., Interdisciplinary Results (IPSI)*, Sveti Stefan, Montenegro, Oct. 2003.

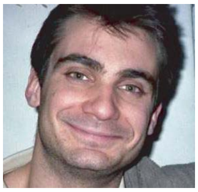
[23] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations Trends Commun. Inf. Theory*, vol. 1, June 2004.

[24] C. B. Peel, B. M. Hochwald, and A. L. Swindleburst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195-202, Jan. 2005.

[25] B. Bollobás, *The Art of Mathematics: Coffee Time in Memphis*. Cambridge University Press, 2006.

[26] L. Schläfli, *Gesammelte Mathematische Abhandlungen*, 209–212. Birkhäuser, 1950.

[27] P. Orlik and H. Terao, "Arrangements of hyperplanes," *Grundlehren der Mathematischen Wissenschaften*, vol. 300, 1992.



Rodrigo de Miguel was born in Madrid, Spain, in 1978. In 2001 he received a Bachelor of Science in Physics from the University of Puget Sound and was inducted a member of Sigma Pi Sigma, The Physics Honor Society (American Institute of Physics). He was granted a Chemical Physics fellowship by the Institute for Physical Science and Technology at the University of Maryland, where he passed the Ph.D. Qualifying Examination in 2002. In 2003 he was awarded a Marie Curie Fellowship by the European Commission to analyze some physical factors contributing to the inefficiency of fuel cells. In 2005 Rodrigo de Miguel joined the Department of Electronics and Telecommunications at the Norwegian University of Science and Technology, where in 2009 he received a PhD for his research at the interphase between statistical physics and wireless communications. In 2006 he was a visiting scholar at the Department of Physics in Bar Ilan University (Tel-Aviv).

Vesna Gardašević received the B.Sc. in electrical engineering from the University of Banja Luka, Bosnia and Herzegovina and the M.Sc.in electrical engineering from the University of Belgrade, Serbia and Montenegro, in 1998 and 2005, respectively. From 1998 to 2006 she worked at the Mobile Divison of Telecom Srpske, Bosnia and Herzegovina. She is currently enrolled in the PhD program at Norwegian University of Science and Technology (NTNU). Her research interests include multiuser communication theory, vector precoding, belief networks, neural networks.

Ralf R. Müller (S'96-M'03-SM'05) was born in Schwabach, Germany, 1970. He received the Dipl.-Ing. and Dr.-Ing. degree with distinction from University of Erlangen-Nuremberg in 1996 and 1999, respectively. From 2000 to 2004, he directed a research group at Vienna Telecommunications Research Center in Vienna, Austria and taught as an adjunct professor at Vienna University of Technology. Since 2005 he has been a full professor at the Department of Electronics and Telecommunications at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. He held visiting appointments at Princeton University, US, Institute Eurecom, France, University of Melbourne, Australia, University of Oulu, Finland, National University of Singapore, Babes-Bolyai University, Cluj-Napoca, Romania, and Kyoto University, Japan. Dr. Müller received the Leonard G. Abraham Prize (jointly with Sergio Verdú) for the paper "Design and analysis of low-complexity interference mitigation on vector channels" from the IEEE Communications Society. He was presented awards for his dissertation "Power and bandwidth efficiency of multiuser systems with random spreading" by the Vodafone Foundation for Mobile Communications and the German Information Technology Society (ITG). Moreover, he received the ITG award for the paper "A random matrix model for communication via antenna arrays," as well as the Philipp-Reis Award (jointly with Robert Fischer). Dr. Müller served as an associate editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2003 to 2006.

Finn F. Knudsen is currently Professor of mathematics at the Norwegian University of Science and Technology (NTNU) in Trondheim Norway. He was born in Connecticut, USA in 1942, and got his Ph.D. in mathematics at Harvard University in 1972. He has extensive knowledge in many fields of mathematics including Logic, Set Theory, Measure Theory, Probability, Algebra, Topology and Geometry. His spesial interest lies in the field of Algebraic Geometry.