Chalmers Publication Library

How costly is it to learn fading channels?

(article starts on next page)

How Costly is it to Learn Fading Channels?

Erwin Riegler[1], Günther Koliander[1], Wei Yang[2], and Giuseppe Durisi[2]

[1]Institute of Telecommunications, Vienna University of Technology, 1040 Vienna, Austria
[2]Department of Signals and Systems, Chalmers University of Technology, 41296 Gothenburg, Sweden

Abstract—Recent results in communication theory suggest that substantial throughput gains in wireless fading networks can be achieved by exploiting network coordination (e.g., CoMP, network MIMO, interference alignment). However, these results are often based on the simplifying assumption that each node in the network has perfect channel knowledge and ignore the channel-estimation overhead.

In this tutorial paper, we take a fresh look at the problem of learning fading channels. By focusing on simple channel models, we will illustrate how to quantify rigorously the throughput loss due to channel-estimation overhead. Specifically, by exploiting that in the absence of a priori channel knowledge at the receiver, the noiseless received signal is a nonlinear function of the transmitted signals and the propagation channel, we will show how to unveil the geometric structure underlying the channel input output relation, and how to use this geometry to characterize capacity at high SNR. We will also demonstrate that this approach is useful to determine the largest rate achievable at finite SNR and finite blocklength.

I. INTRODUCTION

The surge in the use of broadband services in the public and private sectors poses high demands on future wireless communication networks, from the core (backhaul) to the periphery (cellular base stations). Network equipment operators predict that network throughput will have to increase by two to three orders of magnitude by 2020 to match future demands. In fact, it is envisioned that future networks must support capacity densities of several Gb/s per square kilometer [1]. Under this scenario, the availability of effective interference management schemes becomes crucial.

Unfortunately, some of the recently proposed interference-management techniques such as CoMP [2], network MIMO [3], and interference alignment [4] result in excessive control-information overhead [5], [6]. In some scenarios, this overhead has been shown to effectively decrease the net throughput, i.e., the throughput after the control-information overhead is accounted for.

The reason behind this poor performance lies in the assumption of perfect channel state information (CSI) under which these interference management techniques are developed. This results in an alarming disconnect between theory and practice within the field of interference management: schemes that are optimal under the assumption that perfect CSI is available yield sometimes only modest throughput gains in field trials or no gains at all [2], [7].

In this tutorial paper, we take a fresh look at the problem of learning fading channels. By means of simple channel models, we illustrate how geometric insights can be used to quantify rigorously the throughput loss due to channel-estimation overhead.

We will focus on block-fading MIMO models and assume no *a priori* CSI at the receiver and present a novel method to compute the number of degrees of freedom, i.e., the asymptotic ratio between capacity and the logarithm of SNR as SNR grows large. The strength of this method is that it can be applied to more accurate channel models than the classic *constant* block-fading model [8]–[10]. For example, it can be applied to models that allow for generic temporal correlation of the channel coefficients inside the blocks or to models where the number of random parameters that characterizes the fading process within each block is larger than one [11], [12]. For these models, the available tools for the characterization of the number of degrees of freedom yield lower bounds that are not tight.

Finally, we shall address the problem of characterizing the largest rate achievable for a given SNR, a given blocklength, and a given frame error rate. The geometry underlying the channel input-output relation in the absence of additive noise will turn out to be pivotal for deriving tight bounds on the largest achievable rate.

II. DEGREES OF FREEDOM

A. The constant block-fading model

The classic constant block-fading channel model with T transmit and R receive antennas is perhaps the simplest model to capture channel variations in MIMO fading links. It results in the following channel input-output relations within a given block of N channel uses [10]:

$$\boldsymbol{y}_r = \sqrt{\frac{\rho}{T}} \sum_{t \in [1:T]} s_{r,t}\, \boldsymbol{x}_t + \boldsymbol{w}_r\,, \quad r \in [1:R]\,. \qquad (1)$$

Here, $\boldsymbol{x}_t \in \mathbb{C}^N$ is the signal vector originating from the tth transmit antenna over the block of N channel uses; $\boldsymbol{y}_r \in \mathbb{C}^N$ is the signal vector at the rth receive antenna; $s_{r,t} \sim \mathcal{CN}(0,1)$ is the channel coefficient between receive antenna r and transmit antenna t; $\boldsymbol{w}_r \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_N)$ is the noise vector at the rth receive antenna; finally, $\rho \in \mathbb{R}^+$ is the SNR. The channel coefficients $s_{r,t}$ are assumed to be mutually independent and independent across $r \in [1:R]$ and $t \in [1:T]$, and to change in an independent fashion from block to block ("block-memoryless" assumption). The transmitted signal vectors $\boldsymbol{x}_t$ are assumed to be independent of the channel coefficients $s_{r,t}$ and the noise vectors $\boldsymbol{w}_r$. For

later use, we note that the input-output relations (1) can be written as

$$Y = \sqrt{\frac{\rho}{T}}\bar{Y} + W \qquad (2)$$

with $W \triangleq [w_1, \ldots, w_R] \in \mathbb{C}^{N \times R}$, $Y \triangleq [y_1, \ldots, y_R] \in \mathbb{C}^{N \times R}$, and

$$\bar{Y} \triangleq XS \qquad (3)$$

where $X \triangleq [x_1, \ldots, x_T] \in \mathbb{C}^{N \times T}$ and $S \triangleq [s_1, \ldots, s_R] \in \mathbb{C}^{T \times R}$ with columns $s_r \triangleq (s_{r,1}, \ldots, s_{r,T})^{\mathrm{T}}$.

Because of the block-memoryless assumption, the coding theorem in [13, Sec. 7.3] implies that the capacity of the channel (1) is given by

$$C(\rho) = \frac{1}{N} \sup I(X; Y). \qquad (4)$$

Here, $I(X; Y)$ denotes mutual information [14, p. 251] and the supremum is taken over all input distributions on $\mathbb{C}^{TN}$ that satisfy the average-power constraint

$$\mathbb{E}[\mathrm{Tr}\{XX^{\mathrm{H}}\}] \leq TN. \qquad (5)$$

The number of degrees of freedom χ is defined as follows:

$$\chi \triangleq \lim_{\rho \to \infty} \frac{C(\rho)}{\log \rho}.$$

B. Grassmannian structure

For the constant block-fading model (1), the number of degrees of freedom can be guessed through the following calculation that exploits the geometry underlying the input-output relation. Consider for simplicity the case where the number of transmit antennas equals the number of receive antennas, i.e., $T = R = M$. The number of degrees of freedom can be thought of as the number of parameters that can be learned about X from the observation of $\bar{Y}$. We see from (3) that, because S is unknown, the only information the receiver can acquire about X is the M-dimensional linear space spanned by its N-dimensional columns. In fact, the column vectors of the received signal $\bar{Y}$ span the same M-dimensional subspace as the column vectors of X. In other words, the random channel S does not affect the subspace Ω_X spanned by the columns of X. The set of the subspaces Ω_X that can be distinguished at the receiver constitutes the Grassmannian manifold $\mathcal{G}_{N,M}$, whose dimension is given by $M(N - M)$. This number corresponds to the degrees of freedom available over a block of N channel uses, which implies that $\chi = M(1 - M/N)$. Indeed, using the Grassmannian structure just discussed, Zheng & Tse [10] proved that, for the case of arbitrary T and R, the number of degrees of freedom is given by (see also [15] for a more compact proof)

$$\chi = M\left(1 - \frac{M}{N}\right), \quad \text{with } M = \min\{T, R, \lfloor N/2 \rfloor\}. \qquad (6)$$

It turns out that a simple scheme based on the transmission of M^2 pilot symbols and on the linear minimum-mean-square error (LMMSE) estimation of S at the receiver achieves χ [10, Sec. V]. One key feature of this scheme is that it does not require the full knowledge of the Grassmannian

geometry underlying (3). It exploits only the dimension of the information-carrying object, i.e., of the Grassmannian manifold. One limitation lies in the linear estimation step, which turns out to be suboptimal for channel models that are more sophisticated than the constant block-fading model, as we shall see in Section II-D.

C. A local one-to-one mapping

We next present a pilot-based method that overcomes the shortcomings of the scheme proposed in [10, Sec. V] by avoiding the LMMSE estimation step. The idea behind this method is to choose the number of pilots so as to guarantee that the map between $\bar{Y}$ and the pair (X, S) is one-to-one locally. In other words, the number of pilots is chosen so that the Jacobian matrix corresponding to the transformation $(X, S) \to \bar{Y}$ has full rank.

We now illustrate this method for the constant block-fading model. Its application to more sophisticated channel models is discussed in Section II-D. For concreteness, we shall focus on the special case $T = 2, R = 4, N = 4$. Using (6), we get $M = 2$, which implies $\chi = 2(1 - 1/2) = 1$. Next, we decompose the input matrix X as $X = X_{\mathrm{P}} + X_{\mathrm{D}}$ with

$$X_{\mathrm{D}} \triangleq \begin{pmatrix} x_{1,1} & 0 \\ x_{2,1} & 0 \\ 0 & x_{3,2} \\ 0 & x_{4,2} \end{pmatrix}; \qquad X_{\mathrm{P}} \triangleq \begin{pmatrix} 0 & x_{1,2} \\ 0 & x_{2,2} \\ x_{3,1} & 0 \\ x_{4,1} & 0 \end{pmatrix}. \qquad (7)$$

We can think of the nonzero entries in X_{D} as the data symbols, and of the nonzero entries in X_{P} as the pilot symbols. The decomposition in (7) yields a total number of four data symbols per blocklength, which agrees with $\chi = 1$. Interpreting the pilot symbols as fixed parameters, the noiseless input-output relation (3) gives us $R \times N = 16$ equations (the number of entries of $\bar{Y}$) to reconstruct $RT + T(N - T) = 12$ unknowns (the entries of S and the nonzero entries of X_{D}). If we want to reconstruct the data symbols in the absence of noise, the noiseless input-output relation (3) must be, at least locally, one-to-one almost everywhere. The inverse function theorem [16, Th. 9.24] tells us that a mapping $\mathbb{C}^{12} \to \mathbb{C}^{16}$ is locally one-to-one almost everywhere provided that the Jacobian matrix has full rank almost everywhere. Noting that

$$\frac{\partial \bar{y}_{\bar{r},n}}{s_{r,t}} = \delta_{r,\bar{r}} x_{n,t} \qquad \frac{\partial \bar{y}_{r,\bar{n}}}{x_{n,t}} = \delta_{n,\bar{n}} s_{r,t}$$

where

$$\delta_{a,b} \triangleq \begin{cases} 1 & \text{if } a = b \\ 0 & \text{else} \end{cases}$$

we can compute the Jacobian matrix J of the noiseless input-output relation (3) with fixed pilot symbols and obtain

$$J = \begin{pmatrix} X & 0 & 0 & 0 & A_1 \\ 0 & X & 0 & 0 & A_2 \\ 0 & 0 & X & 0 & A_3 \\ 0 & 0 & 0 & X & A_4 \end{pmatrix}$$

where each matrix

$$\boldsymbol{A}_r \triangleq \begin{pmatrix} s_{r,1} & 0 & 0 & 0 \\ 0 & s_{r,1} & 0 & 0 \\ 0 & 0 & s_{r,2} & 0 \\ 0 & 0 & 0 & s_{r,2} \end{pmatrix}$$

contains the partial derivatives of $\bar{\boldsymbol{y}}_r$ with respect to the nonzero data symbols in $\boldsymbol{X}_{\mathrm{D}}$. To show that $\boldsymbol{J}$ has full rank, it is sufficient to prove that the determinant of the submatrix

$$\boldsymbol{K} \triangleq [\boldsymbol{J}]_{[1:16]\setminus\{1,6,11,16\}}$$

which is obtained from $\boldsymbol{J}$ by removing the rows number $1, 6, 11,$ and 16, does not vanish. Using the Laplace expansion [17, Sec. 0.3.1], it is possible to find a particular choice of $\boldsymbol{S}$ and $\boldsymbol{X}_{\mathrm{D}}$ for which the determinant of $\boldsymbol{K}$ is nonzero. But the determinant of $\boldsymbol{K}$ is an analytic function. Hence, either it vanishes identically or does not vanish almost everywhere. This implies that the pilot-based transmission strategy (7) results in a noiseless input-output relation that is locally one-to-one almost everywhere.

It turns out that these geometric arguments suffice to establish that the number of degrees of freedom is at least as large as (6). The key step is to write $I(\boldsymbol{Y};\boldsymbol{X}) = h(\boldsymbol{Y}) - h(\boldsymbol{Y} \mid \boldsymbol{X})$ and lower-bound $h(\boldsymbol{Y})$ using a change of variables. It is precisely this change-of-variables step that requires the noiseless input-output relation to be locally one-to-one almost everywhere [11], [12].

D. More accurate channel models

In MIMO orthogonal frequency-division-multiplexing systems, the fading gains associated to the channel between a given transmit antenna t and a given receive antenna r within a block can be compactly represented as an N-dimensional vector whose average squared inverse Fourier transform corresponds to the power-delay profile of the (t,r) channel. This suggests to modify (1) as follows:

$$\boldsymbol{y}_r = \sqrt{\frac{\rho}{T}} \sum_{t \in [1:T]} s_{r,t}\, \boldsymbol{z}_{r,t} \odot \boldsymbol{x}_t + \boldsymbol{w}_r\,, \quad r \in [1:R]. \quad (8)$$

Here, $\odot$ denotes the Kronecker product and $\boldsymbol{z}_{r,t} \in \mathbb{C}^N$ are deterministic vectors accounting for the different power-delay profiles corresponding to different (t,r) channels. Note that (1) is a special case of (8) with $\boldsymbol{z}_{r,t} = (1,\ldots,1)^{\mathrm{T}}$.

An even more general channel model can be obtained by replacing the scalars $s_{r,t}$ with Q-dimensional random vectors with independent and identically distributed entries, and by replacing the deterministic vectors $\boldsymbol{z}_{r,t}$ by $Q \times N$ deterministic matrices. The corresponding input-output relation describes the situation where the random fading gains within a block are controlled by $Q \geq 1$ random parameters [11], [18] (the model in (8) corresponds to the case $Q = 1$).

Note that (8) can no longer be rewritten in a simple matrix form similar to (2). Hence, the Grassmannian dimension-counting argument described in Section II-B cannot be applied in this setup.

By using the pilot-based scheme combined with LMMSE estimation described in [10], one can show that the number

of degrees of freedom is at least as large as in (6). It turns out, perhaps surprisingly, that this lower bound is not tight. The application of the method described in Section II-C reveals that for almost all matrices

$$\boldsymbol{Z} \triangleq \begin{pmatrix} \boldsymbol{z}_{11} & \cdots & \boldsymbol{z}_{1T} \\ \vdots & \cdots & \vdots \\ \boldsymbol{z}_{R1} & \cdots & \boldsymbol{z}_{RT} \end{pmatrix}$$

a higher number of degrees of freedom than for the constant block fading model can be achieved. Specifically, coming back to the $T = 2, R = 4, N = 4$ example, one can show [12] that the number of degrees of freedom is $\chi = 3/2$ for almost all matrices $\boldsymbol{Z}$, compared to $\chi = 1$ degrees of freedom for the constant block-fading model. This means that the number of degrees of freedom "jumps" from 1 to $3/2$ when we substitute an all-one $\boldsymbol{Z}$ matrix by a generic (i.e., randomly generated out of a continuous distribution) $\boldsymbol{Z}$ matrix. The reason for this increase in the number of degrees of freedom is as follows: a lower number of pilot symbols are needed to make the Jacobian matrix corresponding to the map between the noiseless output signal and the pair input signal and channel gains full rank. This holds because the choice of $\boldsymbol{Z}$ generic guarantees linear independence between the column vectors of the Jacobian matrix.

To summarize, we have shown in this section that a fairly minor modification of the input-output relation of the constant block-fading model, aimed at making the model more accurate, yields a profound change in the geometry, and an increase in the number of degrees of freedom. This shows once more (see also [19]–[21]) that the number of degrees of freedom is extremely sensitive to small perturbations in the statistics of the propagation channel. Hence, one should be careful in using the number of degrees of freedom as a metric to determine the cost of acquiring CSI. More reliable non-asymptotic metrics, some of which discussed in the next two sections, are called for.

III. BEYOND DEGREES OF FREEDOM

The first step towards obtaining high-SNR capacity characterizations that are more accurate than the number of degrees of freedom is to determine the second term in the high-SNR capacity expansion. For the constant block-fading model (1), one can show that

$$C(\rho) = \chi \log \rho + c + o(1), \quad \rho \to \infty \quad (9)$$

where χ is given in (6), c is a constant that does not depend on ρ, and $o(1)$ denotes a function of ρ that vanishes as $\rho \to \infty$. The computation of c in (9) is more involved than that of the number of degrees of freedom χ. For the case $N \geq T + R$, this constant is characterized in [10, Eq. (24)]. The case $N < T + R$ has been recently solved in [15, Eq. (9)].

We next discuss the structure of the input distribution that achieves both χ and c in (9). Using mutual information manipulations exploiting the isotropic nature of $\boldsymbol{S}$ and of $\boldsymbol{W}$, one can show that the capacity-achieving input matrix must be of the form $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}$ where $\boldsymbol{U} \in \mathbb{C}^{N \times T}$ is an isotropically distributed truncated unitary matrix, and

$\boldsymbol{D} = \mathrm{diag}(d_1, \ldots, d_T) \in \mathbb{C}^{T \times T}$ is a diagonal matrix [8, Th. 2].

The geometry unveiled in Section II-B suggests that, at high SNR, the information bearing objects are the subspaces $\Omega_{\boldsymbol{X}}$. As no subspace should be preferred, one is tempted to conclude that the uniform distribution over the Grassmannian is capacity achieving at high SNR. This choice yields $\boldsymbol{D} = \sqrt{N}\boldsymbol{I}_T$, and the corresponding probability distribution on $\boldsymbol{X}$ is sometimes referred to as *unitary space-time modulation (USTM)* [22]–[24]. It turns out that USTM achieves the first two terms in the high-SNR expansion (9) only for the case $N \geq T + R$ [15, Eq. (9)]. The reason is as follows. In [10, p. 369], it is shown that the probability distribution of the entries $\{d_1, \ldots, d_T\}$ must be chosen so as to maximize[1]

$$h(\boldsymbol{QDS}) + (N - T - R)\mathbb{E}[\log \det \boldsymbol{D}^2] \qquad (10)$$

where $\boldsymbol{Q}$ is a $T \times T$ isotropically distributed unitary matrix independent of both $\boldsymbol{D}$ and $\boldsymbol{S}$. For the case $N \geq T + R$, the average-power constraint (5) implies that

$$h(\boldsymbol{QDS}) \leq TR \log(\pi e N)$$
$$\mathbb{E}[\log \det \boldsymbol{D}^2] \leq T \log N.$$

Both inequalities hold with equality if one chooses $\boldsymbol{D} = \sqrt{N}\boldsymbol{I}_T$.

For the case $N < T + R$, the second term in (10) becomes negative and USTM is no longer optimal. Through algebraic manipulations one can show that [15]

$$h(\boldsymbol{QDS}) + (N - T - R)\mathbb{E}[\log \det \boldsymbol{D}^2] = h(\boldsymbol{G}) + \mathrm{const}.$$

Here, $\boldsymbol{G} \in \mathbb{C}^{T \times (N-T)}$ is a random matrix with singular values distributed as the singular values of $\boldsymbol{DS}$ and with isotropically distributed singular vectors. By choosing the joint probability density function (pdf) of the nonzero entries of $\{d_1^2, \ldots, d_T^2\}$ to be the same as the joint pdf of the ordered eigenvalues of a positive-definite $T \times T$ random matrix that has a complex matrix-variate Beta distribution, one can make the entries of the matrix $\boldsymbol{G}$ to be independent and Gaussian-distributed. This maximizes $h(\boldsymbol{G})$. Therefore, the probability distribution just described, which is referred to in [15] as *Beta-variate space-time modulation (BSTM)*, is optimal for the case $N < T + R$.

To conclude, we note that, once the second term in the high-SNR capacity expansion is characterized, non-asymptotic capacity bounds that are tight over a large range of SNR values can be obtained [25]. For the more sophisticated models described in Section II-D, a high-SNR capacity characterization beyond the number of degrees of freedom is currently not available.

IV. FINITE BLOCK-LENGTH REGIME

The ergodic capacity (4) is a relevant performance metric only when sufficiently long codewords are used. Under stringent delay constraints, however, a more appropriate

---

performance metric is the maximal achievable rate $R^*(n, \epsilon)$ for a given blocklength n and frame error probability ϵ.

In this section, we shall consider the scenario where the random fading channel stays constant for the duration of each transmit codeword (i.e., $n = N$). This leads to the so-called quasi-static fading model originally proposed in [26]. For quasi-static fading channels, the ergodic capacity is zero for many fading distributions of practical interest such as the Rayleigh distribution. In this case, the *outage* capacity C_ϵ, which is obtained by letting $n \to \infty$ in $R^*(n, \epsilon)$ for a fixed $\epsilon > 0$, is the appropriate asymptotic metric. The outage capacity of MIMO quasi-static fading channels is given by [27]

$$C_\epsilon = \sup\left\{ \xi : P_{\mathrm{out}}(\xi) \leq \epsilon \right\} \qquad (11)$$

where $P_{\mathrm{out}}(\xi)$ denotes the outage probability corresponding to a given rate ξ

$$P_{\mathrm{out}}(\xi) = \inf_{\substack{\boldsymbol{Q}:\boldsymbol{Q} \geq 0 \\ \mathrm{Tr}\{\boldsymbol{Q}\} \leq T}} \mathbb{P}\left[\log \det\left(\boldsymbol{I}_r + \sqrt{\frac{\rho}{T}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{QS} \right) \leq \xi \right]. (12)$$

In [28], it is shown that, under mild conditions on the probability distribution of the channel-gain matrix $\boldsymbol{S}$, the maximal achievable rate $R^*(n, \epsilon)$ can be lower-bounded as follows:[2]

$$R^*(n, \epsilon) \geq C_\epsilon - O\left(\frac{\log n}{n} \right). \qquad (13)$$

This implies that the convergence of $R^*(n, \epsilon)$ to C_ϵ is fast (compare (13) with [30, Eq. (1)], where the convergence rate of $R^*(n, \epsilon)$ to C for the AWGN channel is reported).

We now demonstrate that a simple coding scheme that exploits the geometry unveiled in Section II-B achieves this lower bound. For simplicity, we focus on the case $T = R$, and assume $n \geq T + R$. We further assume that the fading distribution yields a matrix $\boldsymbol{S}$ that has full rank almost surely. The considerations reported in Section II-B and Section III suggest that, in this case, it is optimal to convey information using the subspace $\Omega_{\boldsymbol{X}}$ spanned by the columns of $\boldsymbol{X}$, and to employ a decoder that bases its decision on the subspace $\Omega_{\boldsymbol{Y}}$ spanned by the received signal $\boldsymbol{Y}$. Since both $\Omega_{\boldsymbol{X}}$ and $\Omega_{\boldsymbol{Y}}$ belong to the Grassmannian $\mathcal{G}_{n,T}$, we can measure their "distance" $d(\Omega_{\boldsymbol{X}}, \Omega_{\boldsymbol{Y}})$. Among the various notions of distance available between points on the Grassmannian [24], [31], the following definition turns out to be appropriate[3]

$$d(\Omega_{\boldsymbol{X}}, \Omega_{\boldsymbol{Y}}) = \prod_{i=1}^{T} \sin^2 \theta_i. \qquad (14)$$

Here, $\pi/2 \geq \theta_1 \geq \cdots \geq \theta_T \geq 0$ are the principle angles (see, e.g., [31, p. 2450]) between $\Omega_{\boldsymbol{X}}$ and $\Omega_{\boldsymbol{Y}}$.

Consider now a decoder that for a given received signal $\boldsymbol{Y}$, declares $\boldsymbol{X}_i$ to be the transmitted codeword if $\boldsymbol{X}_i$ is the only codeword for which $d(\Omega_{\boldsymbol{X}_i}, \Omega_{\boldsymbol{Y}}) \leq \gamma$, where γ is

---

[1]We shall assume $T \leq R$. As shown in [10], using more than R transmit antennas does not increase capacity.

[2]In the special case of $T = 1$, i.e., single-input multiple-output channel, the lower bound (13) is tight [29].

[3]Strictly speaking, $d(\cdot, \cdot)$ in (14) is not a distance in $\mathcal{G}_{n,T}$ because $d(\Omega_1, \Omega_2) = 0$ does not imply $\Omega_1 = \Omega_2$.

a pre-defined threshold. If no codewords or more than one codeword meet this condition, the decoder declares an error. Surprisingly, this simple decoding scheme, when equipped with the appropriate threshold γ, turns out to be sufficient to achieve (13). To motivate this fact, we observe that given $\boldsymbol{X}$ and $\boldsymbol{S}$, we have that [28]

$$d(\Omega_{\boldsymbol{X}}, \Omega_{\boldsymbol{Y}}) \to \frac{1}{\det\left(\boldsymbol{I}_T + \sqrt{\rho/T}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{Q}_{\boldsymbol{X}}\boldsymbol{S}\right)}, \quad n \to \infty$$

where $\boldsymbol{Q}_{\boldsymbol{X}} = n^{-1}\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X}$. Hence,

$$\mathbb{P}[d(\Omega_{\boldsymbol{X}}, \Omega_{\boldsymbol{Y}}) \geq \gamma] \approx \mathbb{P}\left[\det\left(\boldsymbol{I}_T + \sqrt{\frac{\rho}{T}}\boldsymbol{S}^{\mathrm{H}}\boldsymbol{Q}_{\boldsymbol{X}}\boldsymbol{S}\right) \leq \gamma^{-1}\right].$$
(15)

In view of (12) and (15), we may choose $\gamma \approx e^{-C_\epsilon}$ and take the $\boldsymbol{Q}_{\boldsymbol{X}}$ that attains the infimum in (12). Then, (11) implies that

$$\mathbb{P}[d(\Omega_{\boldsymbol{X}}, \Omega_{\boldsymbol{Y}}) \geq e^{-C_\epsilon}] \approx \epsilon.$$

Hence, with probability $(1-\epsilon)$, the subspace $\Omega_{\boldsymbol{Y}}$ lies inside a ball of radius e^{-C_ϵ} centered on $\Omega_{\boldsymbol{X}}$. Let $\mathcal{B}(\gamma)$ be the ball of radius γ in $\mathcal{G}_{n,T}$. Coarsely speaking, the number of codewords that can be transmitted without violating the frame-error constraint ϵ is then given by the ratio between the volume of the Grassmannian and the volume of $\mathcal{B}(\gamma)$. One can show that [28]

$$\frac{\mathrm{Volume}(\mathcal{G}_{n,T})}{\mathrm{Volume}(\mathcal{B}(e^{-C_\epsilon}))} = e^{nC_\epsilon - O(\log n)}.$$
(16)

Taking the logarithm of the right-hand side of (16) and then dividing by n one obtains (13). A rigorous derivation of this result can be found in [28]. Note that the decoder just described does not need the knowledge of the distribution of the fading gains. Hence, it is universal.

REFERENCES

[1] Z. Roth, M. Goldhamer, N. Chayat, A. Burr, M. Dohler, N. Bartzoudis, C. Walker, Y. Leibe, C. Oestges, M. Brzozowy, and I. Bucaille, "Vision and architecture supporting wireless GBit/sec/km2 capacity density deployments," in *Future Network and Mobile Summit, 2010*, Florence, Italy, Jun. 2010, pp. 1–7.

[2] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[3] H. Huh, S.-H. Moon, Y.-T. Kim, I. Lee, and G. Caire, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: A large-system limit analysis," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7771–7786, Dec. 2011.

[4] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[5] O. El Ayach, A. Lozano, and R. Heath, "On the overhead of inter-ference alignment: Training, feedback, and cooperation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 4192–4203, Nov. 2012.

[6] A. Lozano, R. W. Heath Jr., and J. G. Andrews, "Fundamental limits of cooperation," Apr. 2012. [Online]. Available: http://arxiv.org/abs/1204.0011v1

[7] A. Barbieri, P. Gaal, S. Geirhofer, T. Ji, D. Malladi, Y. Wei, and F. Xue, "Coordinated downlink multi-point communications in heterogeneous cellular networks," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, Feb. 2012, pp. 7 –16.

[8] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[9] B. M. Hochwald and T. L. Marzetta, "Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

[10] L. Zheng and D. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[11] G. Koliander, E. Riegler, G. Durisi, V. I. Morgenshtern, and F. Hlawatsch, "A lower bound on the noncoherent capacity pre-log for the MIMO channel with temporally correlated fading," in *Proc. Allerton Conf. Commun. Contr. Comp.*, Monticello, IL, Oct. 2012, pp. 1198–1205.

[12] G. Koliander, E. Riegler, G. Durisi, and F. Hlawatsch, "Generic correlation increases noncoherent MIMO capacity," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT 2013)*, Istanbul, Turkey, Jul. 2013, to appear.

[13] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY: Wiley, 2006.

[15] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 117–132, Feb. 2013.

[16] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York, NY: McGraw-Hill, 1976.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[18] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.

[19] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proc. Allerton Conf. Commun. Contr. Comp.*, Monticello, IL, Oct. 2003, pp. 410–419.

[20] ——, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[21] G. Durisi, V. I. Morgenshtern, and H. Bölcskei, "On the sensitivity of continuous-time noncoherent fading channel capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6372–6391, Oct. 2012.

[22] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1962–1973, Sep. 2000.

[23] B. M. Hassibi and T. L. Marzetta, "Multi-antennas and isotropically random unitary inputs: the received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1473–1484, Jun. 2002.

[24] A. Ashikhmin and R. Calderbank, "Grassmannian packings from operator Reed–Muller codes," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5689–5714, Nov. 2010.

[25] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus channel knowledge at finite block-length," in *Proc. IEEE Inf. Theory Workshop (ITW 2012)*, Lausanne, Switzerland, Sep. 2012.

[26] L. H. Ozarow, S. S. (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 359–378, May 1994.

[27] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[28] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static MIMO fading channels at finite blocklength," 2013, in preparation.

[29] ——, "Quasi-static SIMO fading channels at finite blocklength," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT 2013)*, Istanbul, Turkey, Jul. 2013, to appear.

[30] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[31] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2450–2454, Sep. 2002.