Geometrical relations between space time block code designs and complexity reduction

Oliver Henkel (henkel@hhi.fraunhofer.de)
Fraunhofer German-Sino Lab for Mobile Communications – MCI
Einsteinufer 37, 10587 Berlin, Germany

***Abstract*— In this work, the geometric relation between space time block code design for the coherent channel and its non-coherent counterpart is exploited to get an analogue of the information theoretic inequality $I(X;S) \leq I((X,H);S)$ in terms of diversity. It provides a lower bound on the performance of non-coherent codes when used in coherent scenarios. This leads in turn to a code design decomposition result splitting coherent code design into two complexity reduced sub tasks. Moreover a geometrical criterion for high performance space time code design is derived.**

I. Introduction

In MIMO (Multiple Input Multiple Output) systems space time coding schemes have been proven to be an appropriate tool to exploit the spatial diversity gains. Two distinct scenarios are common, whether the channel coefficients are known (coherent scenario) [1], to the receiver or not (non-coherent scenario) [2]. Prominent coherent codes are the well known Alamouti scheme [3] and general orthogonal designs [4]. A more flexible coding scheme are the so-called linear dispersion codes. They have been introduced in [5] and were further investigated in [6]. A full rate high performing example is the recently discovered Golden code [7]. Genuine non-coherent codes have been proposed in [8], but most of the research efforts in the literature focus on differential schemes, introduced in [9], since differential codes usually provide higher data rates than comparable non differential codes. High performing examples have been constructed in [10], [11],[12],[13]. However, in both (coherent and non-coherent) cases most research effort has been undertaken for space time block codes with quadratic 2-by-2, resp. n_t-by-n_t code matrices (n_t denotes the number of transmit antennas). Although linear dispersion codes are not restricted to quadratic shape of the design matrices the block length is not a free design parameter when the number of transmit antennas is held fixed (compare the asymptotic guidelines in [6]).

In general [14] the signal matrices are of rectangular shape of size $T \times n_t$ with unitary columns. The corresponding coding spaces for the coherent and non-coherent scenario are the complex Stiefel and Grassmann manifolds respectively. Typically the number n_t of transmit antennas is a small number due to hardware limitations, while the block length T can be chosen rather large, upper bounded only by the coherence length of the channel. Inspired from [15], in [16] a general analysis of packings in the Stiefel and Grassmann manifold revealed, that the achievable squared minimal distance (i.e. the squared diameter of the decision regions for decoding) grows proportionally with the block length T, more precisely the following proposition holds [16]):

Journal reference: IEEE Trans. Inform. Theory **52**, no 12 (2006), 5324–5335.

Proposition I.1

For any $T \geq 2n_t$ set $D_{n_t,T} = n_t(2T - n_t)$ (coherent channel), resp. $D_{n_t,T} = 2n_t(T - n_t)$ (non-coherent channel). Then for any prescribed rate $R = \frac{1}{T}\log|\mathcal{C}|$ there exist space time block codes $\mathcal{C}$ with rate R and minimal distance d_0 satisfying

$$d_0 \geq C\sqrt{\rho\frac{T}{n_t}}\left(\frac{1}{2}\right)^{\frac{TR}{D_{n_t,T}}} \quad , \text{ (provided } \rho \geq 1\text{)} \qquad (1)$$

for some constant $C > 0$ depending on the channel knowledge at the receiver. Since the rightmost term is monotonically increasing as a function of T, the receiver performance increases proportionally to $\frac{T}{n_t}$.

Having the common literature (see above) in mind, this result comes rather unexpected and further research effort seems promising. However, explicit code constructions have already been achieved in [17]. Moreover, the Proposition I.1 becomes even more important when considering space frequency code design: The schemes [18], [19] indicate, that the relevant coding spaces are certain subsets of (large dimensional) Stiefel and Grassmann manifolds. Thus considering these coding spaces in general may be of considerable importance for space frequency code designs. Explicit space frequency constructions can be found in e.g. [20].

In the present work it will be shown, how general space time block code designs can be decomposed into two 'smaller' pieces with reduced design complexity (Theorem III.5), both already in the focus of current research. The achieved result can be seen as complementary to that of Kammoun and Belfiore [21], who presented a coding scheme for non-coherent channel space time block codes in terms of coherent channel ones, compare Remark III.7 for further implications.

The key observation is the quite intuitive but technically not obvious diversity monotonicity (Proposition III.4), which states that the performance of each non-coherent channel space time block code grows when considered as a coherent channel

code. This turns out to be due to some higher resolution of the coherent channel receiver, reflecting the information theoretic relation between the system designs.

Further insight on the performance is obtained by local analysis of diversity, leading to the overall picture of space time coding as a constrained sphere packing problem. It reveals additional structures not obvious from the traditional point of view, proposing high performance design criteria (Conclusion IV.2, IV.4) and adding a further estimate (Proposition IV.5) to the diversity embedding. By the way all results are obtained in the spirit of geometrical methods in space time coding theory.

The remainder of this work is organized as follows. Section II introduces the basic models for the channels and coding spaces (with emphasis on their geometrical structures), fixes notation and conventions used throughout this work. Section III defines diversities for the coherent/non-coherent channel cases as our fundamental performance measure and analyses their interrelations, culminating in the embedding and decomposition results mentioned above. Section IV focuses on the local analysis of diversity and the connection to the sphere packing problem, exploring its consequences. Finally, the main results will be summarized for concise reference together with remaining open questions.

II. Channel model and coding spaces

In this section the basic channel model will be presented, leading to the Stiefel and Grassmann manifolds as coding spaces. These spaces will be introduced with emphasis on their topological metric structures induced by the maximum likelihood receivers. The geometric relation between the coding spaces is precisely expressed by the principal fiber structure, which is also introduced here. Although the geometric terms used in this work will be defined (as far as it seems necessary to understand the concepts), the reader who prefers rigorous definitions is invited to consult standard text books e.g. [22], [23] (manifolds), [23], [24], [25] (homogeneous spaces, Lie groups), and/or [23], [25] (principal fibers). For the particular case of the (complex) Stiefel and Grassmann manifolds an introduction to their real counterparts aimed at non-specialists is [26].

A. Channel model

We consider the Rayleigh flat fading MIMO (multiple input multiple output) channel without channel knowledge at the transmitter and maximum likelihood decoding at the receiver as described in [14] (with normed expected power $\sum_j \mathsf{E}|s_{ij}|^2 = 1$ per time step, $i = 1, \dots, T$, E denotes expectation):

$$\begin{gathered} X = \sqrt{\rho}\, SH + W\,, \\ S = (s_{ij}) \in \mathbb{C}^{T\times n_t}, \quad H \in \mathbb{C}^{n_t \times n_r}, \quad X, W \in \mathbb{C}^{T\times n_r} \end{gathered} \tag{2}$$

whereas T denotes the coherence time of the channel (respectively the block length of the signals), n_t, n_r denote the number of transmit, resp. receiver antennas, $W \sim$ i.i.d. $\mathcal{CN}(0,1)$ is the noise, $H \sim$ i.i.d. $\mathcal{CN}(0,1)$ the channel matrix and S, X denote the transmitted, resp. received signal with SNR (signal to noise ration) ρ. The (ergodic) channel capacity is defined by the supremum of the mutual information

$$C = \frac{1}{T} \sup_{p(S)} I((X,H);S)\,, \text{ resp. } C = \frac{1}{T} \sup_{p(S)} I(X;S) \tag{3}$$

for the coherent (resp. non-coherent) channel, and we define the rate R of the code $\mathcal{C}$ by

$$R := \frac{1}{T} \log|\mathcal{C}| \tag{4}$$

The normalization by $1/T$ is merely a convention to have the block length T as a free design parameter of the code, such that codes with distinct block length are comparable.

B. Coding spaces

Hochwald and Marzetta [14, Theorem 1 and 2] have shown, that signals S of the form $S = \frac{1}{\sqrt{n_t}}\Phi\mathcal{E}$ are optimal with respect to the channel capacity (due to the central limit theorem tending to C_0 defined below, when $T \longrightarrow \infty$), if the receiver does not know the channel. More precisely one has $T \geq n_t$, $\mathcal{E} = \operatorname{diag}(\epsilon_i) \in \mathbb{C}^{n_t \times n_t}$ with ϵ_i non-negative, $\mathsf{E}\epsilon_i^2 = T$ stochastic independent from Φ, obeying $\Phi^\dagger\Phi = \mathbf{1}_{n_t}$ ($n_t \times n_t$-unit matrix), Φ therefore being canonically an element of the complex Stiefel manifold $V_{n_t,T}^{\mathbb{C}}$ defined below. In [2] Hochwald/Marzetta, and more generally Zheng/Tse in [27, Lemma 8] have shown that the optimal energy allocation $\mathcal{E}$ of the antennas equals (asymptotically in ρ) $\mathcal{E} = \sqrt{T}\mathbf{1}$, thus

$$S = \sqrt{\frac{T}{n_t}}\,\Phi \tag{5}$$

for $n_t \geq n_r$, $T \geq 2n_r$. The signal $S = \sqrt{\frac{T}{n_t}}\,\Phi$ then carries the total energy $\|S\|_F^2 = T$, thus the transmitter sends with unit power per time step. In this case the mutual information (ergodic in the channel realizations) $I(X;S)$ depends only on the subspace in $\mathbb{C}^T$ spanned by the columns of Φ, not on Φ itself [27]. This is reflected by the fact, that scalings and linear combinations of the columns of Φ are indistinguishable for the detector, when the channel is non-coherent. Therefore these transformations cannot carry any information and we end up with signals $S \in \sqrt{\frac{T}{n_t}}\, G_{n_t,T}^{\mathbb{C}}$, $G_{n_t,T}^{\mathbb{C}}$ denoting the complex Grassmann manifold of n_t-dimensional linear subspaces of $\mathbb{C}^T$.

For the coherent channel the capacity has been calculated by Telatar [28] to

$$C_0 := \mathsf{E} \log\det\left(\mathbf{1} + \frac{\rho}{n_t} H^\dagger H\right) \tag{6}$$

Assuming the same energy allocation $\mathcal{E} = \sqrt{T}\mathbf{1}$ one can justify, that now the asymptotically optimal signal space consists of signals $S \in \sqrt{\frac{T}{n_t}}\, V_{n_t,T}^{\mathbb{C}}$.

We focus on both signal designs in this article, sometimes called unitary space time modulation in the literature (introduced in [2]).

C. Coherent channel: The Stiefel manifold $V_{n_t,T}^{\mathbb{C}}$

The (complex) Stiefel manifold defined by

$$V_{n_t,T}^{\mathbb{C}} := \{\Phi \in \mathbb{C}^{T\times n_t} \mid \Phi^\dagger\Phi = \mathbf{1}\} \tag{7}$$

is diffeomorphic to a coset space with respect to the unitary group $U(T)$ of T-by-T unitary matrices:

$$V_{n_t,T}^{\mathbb{C}} \cong U(T)\,/\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & U(T-n_t)\end{pmatrix}\ ,\ \ \Phi \cong (\Phi,\Phi^\perp)\begin{pmatrix}\mathbf{1}\\ \mathbf{0}\end{pmatrix} \tag{8}$$

whereas $\cong$ means 'diffeomorphic to'. From this equivalence we obtain

$$\begin{aligned}\bar{D}_{n_t,T} &:= \dim_{\mathbb{R}} V_{n_t,T}^{\mathbb{C}} = \dim_{\mathbb{R}} U(T) - \dim_{\mathbb{R}} U(T-n_t)\\ &= n_t(2T-n_t)\end{aligned} \tag{9}$$

for free. Since the elements of the Stiefel manifolds are n_t-dimensional orthonormal bases, they are called n_t-*frames*. Geometrically the coset representation of $V_{n_t,T}^{\mathbb{C}}$ is interpreted as a so-called *homogeneous space*

$$\left(U(T) \xrightarrow{\pi_V^U} V_{n_t,T}^{\mathbb{C}};\, U(T-n_t)\right) \tag{10}$$

This means that each $\Phi \in V_{n_t,T}^{\mathbb{C}}$ is the image of a projection π_V^U from some unitary T-by-T matrix U (in the coset representation π_V^U is simply the projection on the first n_t columns of $U = (\Phi,\Phi^\perp)$) and for each $\Phi,\Psi \in V_{n_t,T}^{\mathbb{C}}$ there exist an unitary $U_{\Phi\Psi}$ with $\Psi = U_{\Phi\Psi}\Phi$. The latter property is obviously fulfilled and called a *transitive left action* of the group $U(T)$ on $V_{n_t,T}^{\mathbb{C}}$ (the defining property for $V_{n_t,T}^{\mathbb{C}}$ being a homogeneous space), while the former property means that π_V^U is invariant with respect to the *right action* of $U(T-n_t)$ on $V_{n_t,T}^{\mathbb{C}}$.

As a linear algebraic convention used in this work, eigenvalues λ_i and singular values σ_i of matrices will be arranged in decreasing order, thus $\lambda_1 \geq \cdots \geq \lambda_{n_t}$, and $\sigma_1 \geq \cdots \geq \sigma_{n_t}$.

A code $\mathcal{C}^V$ for the coherent channel model is given by a discrete set $\mathcal{C}^V = \{\Phi_i\} \subset V_{n_t,T}^{\mathbb{C}}$. At the receiver the maximum likelihood decision reads (see [2])

$$\Phi_{\mathrm{ML}} = \arg\min_{\forall\Phi\in\mathcal{C}} \left\| X - \sqrt{\rho\tfrac{T}{n_t}}\,\Phi H\right\|_{\mathrm{F}} \tag{11}$$

whereas $X = \sqrt{\rho\frac{T}{n_t}}\,\Psi H + W$ is the received signal. Throwing away the noise term allows a formulation of a code design criterion in the signal space $V_{n_t,T}^{\mathbb{C}}$, induced from the ML receiver: The maximization of the pairwise distances $\bar{d}$, given by

$$\bar{d}(\Phi,\Psi) := \|\bar{\Delta}\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^{n_t}\sigma_i^2(\bar{\Delta})} = \|\bar{\sigma}\| \tag{12}$$

where we have set

$$\bar{\Delta} := \Phi - \Psi \tag{13}$$

$$\bar{\sigma} := (\bar{\sigma}_1,\ldots,\bar{\sigma}_{n_t})\,,\ \ \bar{\sigma}_i := \sigma_i(\bar{\Delta}) \in [0,2] \tag{14}$$

Thus coding corresponds to a packing problem on the metric space $\left(V_{n_t,T}^{\mathbb{C}},\bar{d}\right)$[a]. Note, that by

$$\sigma_i(U(\Phi-\Psi)v) = \sigma_i(\Phi-\Psi)\,,\ \ \forall_{U\in U(T),\,v\in U(n_t)} \tag{15}$$

[a] We will see in section III, that this is only an approximation of the design criterion, but the importance of the packing gain will become clear in section IV

the metric $\bar{d}$ remains invariant under left or right multiplication of its arguments with unitary matrices (also denoted as *left invariance* resp. *right invariance*):

$$\bar{d}(U\Phi v, U\Psi v) = \bar{d}(\Phi,\Psi) \tag{16}$$

This property is one motivation for the geometric picture of the Stiefel manifold as a homogeneous space with its corresponding left and right actions. Furthermore for each singular value $\bar{\sigma}_i$ holds

$$\begin{aligned}\tfrac{1}{2}\bar{\sigma}_i^2 &= \tfrac{1}{2}\sigma_i^2(\bar{\Delta}) = \tfrac{1}{2}\lambda_i(\bar{\Delta}^\dagger\bar{\Delta})\\ &= \lambda_i(\mathbf{1} - {}^{\mathcal{H}}(\Phi^\dagger\Psi)) = 1 - \lambda_{n_t-i+1}({}^{\mathcal{H}}(\Phi^\dagger\Psi))\end{aligned} \tag{17}$$

whereas ${}^{\mathcal{H}}M := 1/2(M+M^\dagger)$ denotes the hermitian part.

D. Non-Coherent channel: The Grassmann manifold $G_{n_t,T}^{\mathbb{C}}$

The (complex) Grassmann manifold $G_{n_t,T}^{\mathbb{C}}$ is the set of all n_t-dimensional (complex) linear subspaces of $\mathbb{C}^T$:

$$G_{n_t,T}^{\mathbb{C}} := \{\langle\Phi\rangle \mid \Phi \in V_{n_t,T}^{\mathbb{C}}\} \tag{18}$$

whereas $\langle\Phi\rangle$ denotes the column space of Φ. Since $\Phi \longmapsto \langle\Phi\rangle$ is a projection invariant under all n_t-by-n_t unitary basis transformations we get the coset representation

$$G_{n_t,T}^{\mathbb{C}} \cong U(T)\,\Big/\begin{pmatrix}U(n_t) & \mathbf{0}\\ \mathbf{0} & U(T-n_t)\end{pmatrix}\ ,\ \ \langle\Phi\rangle \cong \Phi\Phi^{1^{-1}} \tag{19}$$

($\Phi^1 := (\mathbf{1},\mathbf{0})\Phi$) and

$$\begin{aligned}\underline{D}_{n_t,T} &:= \dim_{\mathbb{R}} G_{n_t,T}^{\mathbb{C}}\\ &= \dim_{\mathbb{R}} U(T) - \dim_{\mathbb{R}} U(n_t) - \dim_{\mathbb{R}} U(T-n_t)\\ &= 2n_t(T-n_t)\end{aligned} \tag{20}$$

Note that the coordinate representation $\langle\Phi\rangle \cong \Phi\Phi^{1^{-1}}$ holds only locally in general (since it requires Φ^1 to have full rank), but it turns out, that this representation covers all but a set of measure zero and we abandon this distinction between local and global properties in the sequel and drop the distinction between $G_{n_t,T}^{\mathbb{C}}$ and its coordinate domain.

Again we have a geometrical reformulation in terms of the homogeneous space

$$\left(U(T) \xrightarrow{\pi_G^U} G_{n_t,T}^{\mathbb{C}};\, U(n_t)\times U(T-n_t)\right) \tag{21}$$

whereas the transitive left action now reads $\langle\Psi\rangle = U_{\langle\Phi\rangle\langle\Psi\rangle}\langle\Phi\rangle := \left\langle U_{\langle\Phi\rangle\langle\Psi\rangle}\Phi\right\rangle$ (e.g. choose $U_{\langle\Phi\rangle\langle\Psi\rangle} = U_{\Phi\Psi}$). The projection π_G^U is now invariant with respect to the combined right action of $U(n_t)\times U(T-n_t)$, because not only the orthogonal complement of the columns in Φ has been neglected, but also the particular choice of the spanning n_t-frame: Each $(\Phi,\Phi^\perp)\begin{pmatrix}u & \mathbf{0}\\ \mathbf{0} & v\end{pmatrix}$ represents the same space $\langle\Phi\rangle$ for arbitrary $u \in U(n_t)$, $v \in U(T-n_t)$.

To simplify matters let us assume $n_t \leq T/2$ whenever we are in contact with the Grassmann manifold. This is no restriction, since for $n_t \geq T/2$ we can always switch to the orthogonal complement of the subspaces under consideration. Given now two elements $\langle\Phi\rangle, \langle\Psi\rangle \in G_{n_t,T}^{\mathbb{C}}$ then there exist n_t principal angles $0 \leq \vartheta_1 \leq \cdots \leq \vartheta_{n_t} \leq \pi/2$ between $\langle\Phi\rangle$ and $\langle\Psi\rangle$. They are defined successively by the critical

values $\arccos|<v_i, w_i>|$, $i = 1, \dots, n_t$ (in increasing order), of $(v, w) \longmapsto \arccos|<v, w>|$ where the unit vectors v, w vary over $\{v_1, \dots, v_{i-1}\}^\perp \subset \langle\Phi\rangle$, respectively $\{w_1, \dots, w_{i-1}\}^\perp \subset \langle\Psi\rangle$, compare [29]. The components of the vector of principal angles $\vartheta := (\vartheta_1, \dots, \vartheta_{n_t})$ can be computed by the formula (any representing n_t-frame will do) [29]

$$\cos\vartheta_i = \sigma_i(\Phi^\dagger\Psi) \tag{22}$$

An important application of principal angles on some given pair $\langle\Phi\rangle, \langle\Psi\rangle$ with principal angles ϑ is, that due to the transitivity of the unitary group action there exist an unitary U, such that Ψ (say) can always be translated into $\binom{1}{0} = U\Psi$ and in $U\langle\Phi\rangle = \langle U\Phi\rangle$ one can choose a basis such that we end up with the canonical representing n_t-frames

$$\Psi_0 = \begin{pmatrix}\mathbf{1}\\ \mathbf{0}\end{pmatrix}\ ,\ \ \Phi_0 = \begin{pmatrix}(\cos\vartheta_i)\\ (\sin\vartheta_i)\\ \mathbf{0}\end{pmatrix} \tag{23}$$

(where $(\cos\vartheta_i) := \mathrm{diag}(\cos\vartheta_i)_{i=1,\dots,n_t} \in \mathbb{R}^{n_t\times n_t}$) for the translated spaces $\langle\Psi_0\rangle = U\langle\Psi\rangle$, $\langle\Phi_0\rangle = U\langle\Phi\rangle$. Note, that the demand to *choose* the appropriate basis Φ_0 in $U\langle\Phi\rangle$ is mandatory, in general there is no $U \in U(T)$ which translates the n_t-frames Ψ, Φ simultaneously into Ψ_0, Φ_0.

E. The principal fiber structure P_G^V

The natural relationship between the homogeneous spaces $V_{n_t,T}^{\mathbb{C}}$ and $G_{n_t,T}^{\mathbb{C}}$ is subsumed in the canonical *principal fiber bundle* structure

$$P_G^V := \left(V_{n_t,T}^{\mathbb{C}} \xrightarrow{\pi_G^V} G_{n_t,T}^{\mathbb{C}};\, U(n_t)\right) \tag{24}$$

which (locally) embeds $G_{n_t,T}^{\mathbb{C}}$ into $V_{n_t,T}^{\mathbb{C}}$ by choosing a representing n_t-frame Φ which spans the subspace $\langle\Phi\rangle$. However there remains the freedom of multiplication with arbitrary unitary matrices $u \in U(n_t)$ from the right (all of them have the same image under the projection π_G^V), and for practical applications it is necessary to specify a unique choice for Φ and u, given $\langle\Phi\rangle$ (simultaneously for all $\langle\Phi\rangle \in G_{n_t,T}^{\mathbb{C}}$, not only for pairs as in (23)). But locally this can always be achieved and we do not want to go into details here. The term 'principal fiber bundle' means a generalization of the term 'homogeneous space', where now the *total space* $V_{n_t,T}^{\mathbb{C}}$ no longer need to be a group and the *base space* $G_{n_t,T}^{\mathbb{C}}$ is a projection π_G^V of the total space which is invariant under a right action of $U(n_t)$. The set of all elements Φu is called a *fiber* over $\langle\Phi\rangle$.

This geometrical point of view makes clear, that we can consider codes $\mathcal{C}^G \subset G_{n_t,T}^{\mathbb{C}}$ for the non-coherent channel as discrete subsets of $V_{n_t,T}^{\mathbb{C}}$ in virtue of the local embedding of $G_{n_t,T}^{\mathbb{C}}$ into $V_{n_t,T}^{\mathbb{C}}$. But one motivation for the introduction of all these perhaps unfamiliar geometrical terms is to clarify the relationship between the coding spaces, i.e. that there is no canonical representation of $\mathcal{C}^G$ in $V_{n_t,T}^{\mathbb{C}}$. In practical applications this peculiarity is often overlooked, since common mathematics software packages already use certain conventions when representing subspaces in terms of singular value decompositions. Furthermore we will see that the unitary left and right actions on the coding spaces lead naturally to the diversity embedding results derived in the next section. These results are geometrical in nature rather than linear algebraic, but only in the geometric context it becomes clear, that they are not obvious at all, since they relate distinct metric structures. For the Stiefel manifold the relevant metric structure has already been defined in (12) and for the Grassmann manifold we will define it next.

We consider codes $\mathcal{C}^G \subset G_{n_t,T}^{\mathbb{C}}$ always as discrete subsets of $V_{n_t,T}^{\mathbb{C}}$ and the maximum likelihood criterion for the non-coherent channel receiver reads now ([2])

$$\Phi_{\mathrm{ML}} = \arg\max_{\forall_{\Phi\in\mathcal{C}}} \left\| \sqrt{\rho\tfrac{T}{n_t}}\,\Phi^\dagger X \right\|_{\mathrm{F}} \tag{25}$$

whereas $X = \sqrt{\rho\frac{T}{n_t}}\,\Psi H + W$ is the received signal. To obtain a design criterion in the signal space $G_{n_t,T}^{\mathbb{C}}$ we throw away the noise term (as in the coherent channel case) and pass from ΨH to $\Psi \in G_{n_t,T}^{\mathbb{C}} \subset V_{n_t,T}^{\mathbb{C}}$ (this operation does not change the column space of Ψ). Setting

$$\underline{\Delta} := \Phi^\dagger\Psi \tag{26}$$

$$\|\underline{\Delta}\|_{\mathrm{F}}^2 = \sum_{i=1}^{n_t}\sigma_i^2(\underline{\Delta}) \overset{(22)}{=} \sum_i \cos^2\vartheta_i = n_t - \sum_i \sin^2\vartheta_i \tag{27}$$

(note that (27) does not depend on the choice of the representing n_t-frame, thus represents really an entity on $G_{n_t,T}^{\mathbb{C}}$), and

$$\underline{\sigma} = (\underline{\sigma}_1, \dots, \underline{\sigma}_{n_t})\ ,\ \ \underline{\sigma}_i := \sigma_i(\underline{\Delta}) = \cos\vartheta_i \tag{28}$$

the ML criterion demands the maximization of the pairwise distances

$$\begin{aligned}\underline{d}(\Phi,\Psi) &:= \sqrt{n_t - \|\underline{\Delta}\|_{\mathrm{F}}^2} = \sqrt{\mathrm{tr}\left(\mathbf{1} - \underline{\Delta}^\dagger\underline{\Delta}\right)}\\ &= \sqrt{\sum_{i=1}^{n_t}\left(1-\underline{\sigma}_i^2\right)} = \sqrt{\sum_{i=1}^{n_t}\sin^2\vartheta_i}\end{aligned} \tag{29}$$

Formally $\underline{d}$ is defined on all of $V_{n_t,T}^{\mathbb{C}}$, but independent of the choice of the representing n_t-frame as already indicated. Of course, it is a metric in the strict sense only as a function on $G_{n_t,T}^{\mathbb{C}}$ (known as the 'chordal' distance, compare [30], [15]), turning again the coding problem into a packing problem in $\left(G_{n_t,T}^{\mathbb{C}}, \underline{d}\right)$. It shares the invariance properties of the coherent channel metric $\bar{d}$, but satisfies even more:

$$\underline{d}(U\Phi v, U\Psi w) = \underline{d}(\Phi,\Psi) \tag{30}$$

by

$$\sigma_i((U\Phi v)^\dagger(U\Psi w)) = \sigma_i(\Phi^\dagger\Psi)\ ,\ \ \forall_{U\in U(T),\, v,w\in U(n_t)} \tag{31}$$

III. Performance analysis: Diversity

In practical settings, where $\rho, T \ll \infty$, the receiver metrics $\underline{d}, \bar{d}$ fail to be the sole code design criteria. Denoting the pairwise error probability of mistaking one symbol for another at the receiver generically as P_{ij} one gets the union upper bound

$$\frac{1}{|\mathcal{C}|}\sum_i\sum_{j\neq i} P_{ij} \tag{32}$$

for the exact error probability. This section deals with the pairwise error probability Chernov bound, more precisely

with the *diversity*, which is essentially the reciprocal of the Chernov bound. It turns out, that the receiver metric coincides with the first order term of the diversity and the highest order term leads to the so called diversity product (further analyzed in section IV). Adopting the diversity as the major performance measure, section III-A investigates the connection between non-coherent and coherent channel designs and its consequences for code design. For convenience we fix the pair (Φ, Ψ) of code symbols throughout this section and suppress their notation as function arguments.

For the coherent channel case the pairwise error probability has been calculated in [2] to

$$P(\Phi,\Psi) = \sum_{\{\bar{\alpha}_j\}} \mathrm{Res}_{\omega=\imath\bar{\alpha}_j} \left\{ -\frac{1}{\omega+\imath/2} \prod_{\substack{i=1\\ \bar{\sigma}_i>0}}^{n_t} \left[\frac{1}{\rho\frac{T}{n_t}\,\bar{\sigma}_i^2(\omega^2+\bar{\alpha}_i^2)} \right]^{n_r} \right\} \tag{33}$$

with $\bar{\alpha}_i := \sqrt{\frac{1}{4}+\frac{1}{(\rho T/n_t)\bar{\sigma}_i^2}}$. Analogously for the non-coherent channel case holds [2]

$$P(\Phi,\Psi) = \sum_{\{\underline{\alpha}_j\}} \mathrm{Res}_{\omega=\imath\underline{\alpha}_j} \left\{ -\frac{1}{\omega+\imath/2} \prod_{\substack{i=1\\ \underline{\sigma}_i<1}}^{n_t} \left[\frac{1+(\rho\frac{T}{n_t})}{(\rho\frac{T}{n_t})^2(1-\underline{\sigma}_i^2)(\omega^2+\underline{\alpha}_i^2)} \right]^{n_r} \right\} \tag{34}$$

with $\underline{\alpha}_i := \sqrt{\frac{1}{4}+\frac{1+(\rho T/n_t)}{(\rho T/n_t)^2(1-\underline{\sigma}_i^2)}}$.

For both cases the we have the Chernov bound

$$P \le \frac{1}{2}\left(\prod_{i=1}^{n_t} \left[1+\varrho\,\sigma_i^2\right] \right)^{-n_r} \tag{35}$$

whereas (coherent channel)

$$\varrho = \bar{\varrho} := \frac{\rho T}{4n_t} \tag{36}$$

$$\sigma_i = \bar{\sigma}_i \tag{37}$$

respectively (non-coherent channel)

$$\varrho = \underline{\varrho} := \frac{\bar{\varrho}^2}{\bar{\varrho}+1/4} = \frac{(\rho\frac{T}{n_t})^2}{4(1+\rho\frac{T}{n_t})} \tag{38}$$

$$\sigma_i = \sqrt{1-\underline{\sigma}_i^2} \tag{39}$$

The term in parentheses in (35) is called (pairwise) diversity

$$\mathcal{D}iv := \prod_{i=1}^{n_t} \left[1+\varrho\,\sigma_i^2\right] \tag{40}$$

and we take it as our basic performance measure for codes. Rewriting $\mathcal{D}iv$ as a polynomial in ϱ requires the use of elementary symmetric polynomials defined by $\mathrm{sym}_0^{n_t} := 1$, $\mathrm{sym}_j^{n_t}(x_1,\dots,x_{n_t}) = \sum_{I_j\in\binom{\mathrm{n_t}}{j}} x_{I_j} = \sum_{I_j\in\binom{\mathrm{n_t}}{j}} x_{i_1}\cdots x_{i_j}$ (with $\binom{\mathrm{n_t}}{j} := \{(i_1,\dots,i_j)\in\mathbb{N}^j \,|\, 1\le i_1<\dots<i_j\le n_t\}$), $j=1,\dots,n_t$. With the abbreviation

$$s_j := \mathrm{sym}_j^{n_t}(\sigma_1^2,\dots,\sigma_{n_t}^2) = \sum_{I_j\in\binom{\mathrm{n_t}}{j}} s_{I_j}^2 \tag{41}$$

we find generically

$$\mathcal{D}iv = \sum_{i=0}^{n_t} s_i\varrho^i \tag{42}$$

The first and highest order coefficient of this polynomial are of particular importance, since they dominate the diversity in the low and high SNR regime respectively. They are called *diversity sum* and *diversity product* respectively, and are given by

$$d = \sqrt{s_1} = \|\sigma\| \tag{43}$$

$$p := \sqrt{s_{n_t}} = \sigma_1\cdots\sigma_{n_t} \tag{44}$$

The diversity sum is our familiar metric $\bar{d} = \|\bar{\sigma}\| = \|\bar{\Delta}\|_{\mathrm{F}}$ (12), resp. $\underline{d} = \|\underline{\sigma}\| = \sqrt{\mathrm{tr}(\mathbf{1}-\underline{\Delta}^\dagger\underline{\Delta})} = \|\sin\vartheta\|$ (29). The diversity product acts as a regularity criterion for the positive semidefinite matrix $\bar{\Delta}^\dagger\bar{\Delta}$, resp. $\mathbf{1}-\underline{\Delta}^\dagger\underline{\Delta}$: In the coherent channel case $\bar{p}^2 = \det\bar{\Delta}^\dagger\bar{\Delta} > 0$ is known as diversity criteria (resp. rank criteria or determinant criteria) in the literature (e.g. [1]). In the non-coherent channel case $\underline{p}^2 = \det(\mathbf{1}-\underline{\Delta}^\dagger\underline{\Delta}) = (1-\underline{\sigma}_1^2)\dots(1-\underline{\sigma}_{n_t}^2) \overset{(28)}{=} \sin^2\vartheta_1\dots\sin^2\vartheta_{n_t} > 0$ measures the positivity of the principal angles between $\langle\Phi\rangle$ and $\langle\Psi\rangle$.

All terms $\bar{s}_{I_j}$, resp. $\underline{s}_{I_j}$ in the diversity expansion possess the invariance properties induced by (15), resp. (31). Therefore the analysis in this section applies to all terms in (42) and the result can be stated in closed form for the full diversity, rather than only to its first and highest order coefficient.

Specializing (42) to the non-coherent channel case, one checks easily that the coefficients are formally defined on all of $V_{n_t,T}^{\mathbb{C}}$, but independent of the choice of the representing n_t-frame. Note that the coherent and non-coherent channel diversities are formally similar due to (42), but the constituting singular values (14), (28) reflect the underlying topological structures induced by the maximum likelihood receivers (11), (25) (resp. the metrics $\bar{d}$, $\underline{d}$). And these structures are entirely distinct.

A. Embedding properties

Now let us investigate the relation between the non-coherent and coherent channel diversity quantities. From the information theoretic inequality $I(X;S) \le I((X,H);S)$ between the corresponding mutual informations we expect such a relation satisfied by the diversity. The ranges for $\bar{\sigma}$ (14) and $\underline{\sigma}$ (28) indicate, that the coherent channel receiver may benefit from some higher 'resolution', but if and how this carries over to the diversity is not obvious and requires a rigorous proof. The investigations of this section give an affirmative answer to that conjecture.

By a slight abuse of notation let us define the 'fiber minima' of $\bar{s}_{I_j}$ with respect to the fibers of P_G^V (24) as

$$\bar{s}\mathrm{dist}_{I_j}(\Phi,\Psi) := \min_{\substack{\Phi\in\pi_G^{V-1}(\langle\Phi\rangle)\\ \Psi\in\pi_G^{V-1}(\langle\Psi\rangle)}} \bar{s}_{I_j}(\Phi,\Psi) \tag{45}$$

Then we obtain

Lemma III.1

Let $\langle\Phi\rangle, \langle\Psi\rangle \in G_{n_t,T}^{\mathbb{C}}$ separated by principal angles

$\vartheta_1, \ldots, \vartheta_{n_t}$. *Then* $\forall_{I_j \in \binom{\mathrm{n_t}}{j}}$

$$\bar{s}\mathrm{dist}_{I_j}(\Phi, \Psi) = \sqrt{2^j \prod_{i \in I_j} (1 - \cos \vartheta_{n_t - i + 1})} \tag{46}$$

holds.

Proof:
Due to left invariance of $\bar{s}_{I_j}$ we can switch to the canonical n_t-frame bases Ψ_0, Φ_0 (23) of $U\langle\Psi\rangle, U\langle\Phi\rangle$. With $\Phi_0(u) := \Phi_0 u$ ($u \in U(n_t)$), running through the fiber over $\langle\Phi_0\rangle$, $\bar{\Delta}(u) := \Phi_0(u) - \Psi_0$, and $\frac{1}{2}\bar{\Delta}(u)^\dagger \bar{\Delta}(u) = \mathbf{1} - {}^{\mathcal{H}}((\cos\vartheta_l)u)$ (recall, that $(\cos\vartheta_l) = \mathrm{diag}(\cos\vartheta_l)_{l=1\ldots n_t}$) we have

$$\begin{aligned}
\bar{s}^2_{I_j}(\Phi_0(u), \Psi_0) &= \prod_{i \in I_j} \sigma_i^2(\bar{\Delta}(u)) \\
&\overset{(17)}{=} \prod_{i \in I_j} 2\big\{1 - \lambda_{n_t - i + 1}[{}^{\mathcal{H}}((\cos\vartheta_l)u)]\big\} \\
&\overset{(*)}{\geq} 2^j \prod_{i \in I_j} \big\{1 - \sigma_{n_t - i + 1}\left((\cos\vartheta_l)u\right)\big\} \\
&= 2^j \prod_{i \in I_j} \big\{1 - \sigma_{n_t - i + 1}((\cos\vartheta_l))\big\} \\
&= 2^j \prod_{i \in I_j} (1 - \cos\vartheta_{n_t - i + 1})
\end{aligned} \tag{47}$$

where $(*)$ comes from the general inequality

$$\lambda_i\big({}^{\mathcal{H}}(A)\big) \leq \sigma_i(A) \tag{48}$$

devoted to Fan-Hoffman in [31, Prop. III.5.1]. $u = \mathbf{1}$ achieves equality in $(*)$ and this completes the proof. ∎
In particular we have the fiber distance

$$\begin{aligned}
\bar{d}\mathrm{dist}(\Phi, \Psi) &:= \min_{\substack{\Phi \in \pi_G^{V^{-1}}(\langle\Phi\rangle) \\ \Psi \in \pi_G^{V^{-1}}(\langle\Psi\rangle)}} \bar{d}(\Phi, \Psi) \\
&= \sqrt{\sum_{I_1 \in \binom{\mathrm{n_t}}{1}} \bar{s}\mathrm{dist}^2_{I_1}} = \sqrt{2\sum_{i=1}^{n_t}(1 - \cos\vartheta_i)}
\end{aligned} \tag{49}$$

and its analogon for the diversity product

$$\begin{aligned}
\bar{p}\mathrm{dist}(\Phi, \Psi) &:= \min_{\substack{\Phi \in \pi_G^{V^{-1}}(\langle\Phi\rangle) \\ \Psi \in \pi_G^{V^{-1}}(\langle\Psi\rangle)}} \bar{p}(\Phi, \Psi) \\
&= \bar{s}\mathrm{dist}_{I_{n_t}} = \sqrt{2^{n_t} \prod_{i=1}^{n_t}(1 - \cos\vartheta_i)}
\end{aligned} \tag{50}$$

We observe, that the fiber minima $\bar{s}\mathrm{dist}_{I_j}$ for each given pair (Φ, Ψ) are realized by the same choice (Φ_0, Ψ_0), which justifies the definition

$$\bar{s}\mathrm{dist}_j(\Phi, \Psi) := \sum_{I_j \in \binom{\mathrm{n_t}}{j}} \bar{s}\mathrm{dist}^2_{I_j}(\Phi, \Psi) \tag{51}$$

(thus we have in particular $\bar{d}\mathrm{dist} = \sqrt{\bar{s}\mathrm{dist}_1}$ and $\bar{p}\mathrm{dist} = \sqrt{\bar{s}\mathrm{dist}_{n_t}}$) leading to

Corollary III.2
For any pair $\langle\Phi\rangle, \langle\Psi\rangle \in G^{\mathbb{C}}_{n_t,T}$ *we have* $\forall_{j=1,\ldots,n_t}$

$$\underline{s}_j(\Phi, \Psi) \leq \bar{s}\mathrm{dist}_j(\Phi, \Psi) \leq \bar{s}_j(\Phi, \Psi) \tag{52}$$

Proof:
The second inequality holds by definition of $\bar{s}\mathrm{dist}$. So let us turn to the first inequality and denote the principal angles between $\langle\Phi\rangle$ and $\langle\Psi\rangle$ by $\vartheta = (\vartheta_1, \ldots, \vartheta_{n_t})$, $0 \leq \vartheta_i \leq \frac{\pi}{2}$. We have $\underline{s}^2_{I_j} = \prod_{i \in I_j} \sin^2\vartheta_i = \prod_{i \in I_j}(1 - \cos^2\vartheta_i)$ and $\bar{s}\mathrm{dist}^2_{I_j} = 2^j \prod_{i \in I_j}(1 - \cos\vartheta_{n_t - i + 1}) = \prod_{i \in I_j} 2(1 - \cos\vartheta_{n_t - i + 1})$. For any $0 \leq \vartheta_i \leq \frac{\pi}{2}$

$$2\cos\vartheta_i - \cos^2\vartheta_i \leq 1 \tag{53}$$

holds, thus

$$\underline{s}_{I_j} \leq \bar{s}\mathrm{dist}_{I_j^*} \tag{54}$$

with $I_j^* := (n_t - i_1 + 1, \ldots, n_t - i_j + 1)$ for each $I_j = (i_1, \ldots, i_j)$, and by

$$\underline{s}_j = \sum_{I_j} \underline{s}^2_{I_j} \overset{(54)}{\leq} \sum_{I_j} \bar{s}\mathrm{dist}^2_{I_j^*} = \sum_{I_j} \bar{s}\mathrm{dist}^2_{I_j} = \bar{s}\mathrm{dist}_j \tag{55}$$

the claim follows. ∎
Given a function $s : \mathcal{C} \times \mathcal{C} \longrightarrow \mathbb{R}$, let us define $s^{\min} := \min_{\mathcal{C}\times\mathcal{C}} s$. Then we state

Corollary III.3

$$\underline{s}_j^{\min} \leq \bar{s}\mathrm{dist}_j^{\min} \leq \bar{s}_j^{\min}, \qquad \forall_{j=1,\ldots,n_t} \tag{56}$$

(unfortunately neither there seems to be a canonical way to determine the pairs of points, which realize the minima, nor whether this could be achieved simultaneously for each of the quantities above by a single pair of points)

Proof:
Corollary III.3 is an easy consequence of Corollary III.2. For each inequality the proof goes the same, so let us take two functions $\mathrm{f} \leq \mathrm{F}$, $X \times X \longrightarrow \mathbb{R}$, X a discrete set, for brevity. Then there are two cases

1) $\mathrm{f}^{\min} = \mathrm{f}(x_0, y_0) \leq \mathrm{F}(x_0, y_0) = \mathrm{F}^{\min}$ and there is nothing to do
2) $\mathrm{f}^{\min} = \mathrm{f}(x_0, y_0)$, but $\mathrm{F}(x_0, y_0) > \mathrm{F}(x_1, y_1) = \mathrm{F}^{\min}$. But then still $\mathrm{f}^{\min} \leq \mathrm{f}(x_1, y_1) \leq \mathrm{F}(x_1, y_1) = \mathrm{F}^{\min}$ holds.

∎

On the metric level (the diversity sum) this inequalities provide a distance gain due to the channel knowledge. It increases the resolution of the detector and allows the receiver to separate points better than the non-coherent channel receiver could do, $\underline{d} \leq \bar{d}$, or equivalently the unit spheres with respect to $\bar{d}$ occupy smaller volume than the corresponding (embedded) unit spheres with respect to $\underline{d}$, thus one can pack more $\bar{d}$-spheres into $V^{\mathbb{C}}_{n_t,T}$ than $\underline{d}$-spheres. But due to the famous estimate (48) we have proven a considerable stronger result not confined to the diversity sum, but rather to any coefficient in the diversity expansion (42). Thus we are able to relate the inequalities derived so far to the diversity as a whole:

Comparing the 'effective' SNRs $\overline{\varrho}$ (36), $\underline{\varrho}$ (38) in the diversity (42) demands one additional estimate (provided $T \geq n_t$)

$$\frac{4}{\rho}\underline{\varrho} = \frac{\rho(\frac{T}{n_t})^2}{1+\rho\frac{T}{n_t}} \leq \frac{T}{n_t} \quad \Longrightarrow \quad \underline{\varrho} \leq \overline{\varrho} \tag{57}$$

thus we have

Proposition III.4

For any pair $\langle\Phi\rangle, \langle\Psi\rangle \in G^{\mathbb{C}}_{n_t,T}$

$$\underline{\mathcal{D}iv}(\Phi,\Psi) \leq \overline{\mathcal{D}iv}(\Phi,\Psi) \tag{58}$$

holds.

Proof:
The proposition follows directly from (52) and (57) ∎
So we conclude, that the coherent channel maximum likelihood receiver applied to $\mathcal{C}^G$ has at least the diversity as the non-coherent channel receiver, the diversity grows. This approves the information theoretic inequality $I(X;S) \leq I((X,H);S)$ motivating our analysis.

Having explored the relationship of the embedding $G^{\mathbb{C}}_{n_t,T} \subset V^{\mathbb{C}}_{n_t,T}$ let us come to a somewhat complementary scenario, which offers the possibility of coding complexity reduction: Consider a single fiber over $\langle\Phi\rangle$. Then, by $\Phi w = (\Phi, \Phi^{\perp})\left(\begin{smallmatrix} w \\ \mathbf{0} \end{smallmatrix}\right)$, there holds a special kind of 'vertical' left invariance, namely

$$\overline{s}_{I_j}(\Phi u, \Phi v) = \overline{s}^U_{I_j}(u,v)\,, \qquad \forall_{I_j \in \binom{n_t}{j}}\,,\ \forall_{u,v\in U(n_t)} \tag{59}$$

where the right hand side is evaluated in $U(n_t) = V^{\mathbb{C}}_{n_t,n_t}$. Analogously we define for the special case $T = n_t$: $\overline{\varrho}^U := \frac{\rho}{4}$, $\overline{s}^U_j := \sum_{I_j \in \binom{n_t}{j}} (\overline{s}^U_{I_j})^2$, $\overline{\mathcal{D}iv}^U := \sum_i \overline{s}^U_i (\overline{\varrho}^U)^i$ and we arrive at

Theorem III.5

Given codes $\mathcal{C}^G \subset G^{\mathbb{C}}_{n_t,T} \subset V^{\mathbb{C}}_{n_t,T}$ *and* $\mathcal{C}^U \subset U(n_t)$*, then the composed code* $\mathcal{C}^V \subset V^{\mathbb{C}}_{n_t,T}$ *given by*

$$\mathcal{C}^V := \mathcal{C}^G \cdot \mathcal{C}^U = \left\{\Phi u \,|\, \Phi \in \mathcal{C}^G, u \in \mathcal{C}^U\right\} \tag{60}$$

satisfies

$$\overline{s}^{\min}_j \geq \min\{\overline{s}\mathrm{dist}^{\min}_j, \overline{s}^{U\,\min}_j\}\,, \qquad \forall_{j=1,\ldots,n_t} \tag{61}$$

and

$$\overline{\mathcal{D}iv}^{\min} \geq \min\{\underline{\mathcal{D}iv}^{\min}, \widetilde{\overline{\mathcal{D}iv}}^{U\,\min}\} \tag{62}$$

holds, whereas $\widetilde{\overline{\mathcal{D}iv}}^U := \sum_i (\frac{T}{n_t})^i \overline{s}^U_i (\overline{\varrho}^U)^i$ *(thus the power constraint factor* $\sqrt{\frac{T}{n_t}}$ *sharpens the estimate).*

Proof: The theorem follows directly from Corollary III.3, Proposition III.4 and the definition of $\overline{s}\mathrm{dist}$ as a fiber minimum. ∎
Therefore the code design splits up into two parts: Codes $\mathcal{C}^G$ represent the familiar coding problem for the non-coherent channel corresponding to $G^{\mathbb{C}}_{n_t,T}$, which has smaller dimension as the general problem in $V^{\mathbb{C}}_{n_t,T}$. The code $\mathcal{C}^U$ represents a coding problem for the coherent channel in $U(n_t) = V^{\mathbb{C}}_{n_t,n_t}$, contributing the dimensions left by $V^{\mathbb{C}}_{n_t,T} \cong G^{\mathbb{C}}_{n_t,T} \times U(n_t)$ locally. So both parts represent a somewhat smaller coding problem with respect to the dimension of the signal spaces. Moreover for both parts the code design is easier to solve than in $(V^{\mathbb{C}}_{n_t,T}, \overline{d})$: In $U(n_t)$ there are many solutions (i.e. codes) in the common literature, e.g. the Alamouti scheme for $n_t = 2$, orthogonal designs for $n_t \geq 2$, quasi-orthogonal space time block codes, and many more. The Grassmannian part is also simpler (not only concerning dimensions but also) in structure, because the 'chordal' design metric $\underline{d}$ is geometrically more natural than the Euclidean distance measure $\overline{d}$ (in terms of their relation to the natural geodesic distance [16]), thus geometric methods may apply. Also packings in $(G^{\mathbb{C}}_{n_t,T}, \underline{d})$ are already in the focus of current research, e.g. [30], [15], whereas [30] also contains explicit constructions for packings in the (real) Grassmann manifold. In [32] a differential geometric connection (based on [16]) has been developed to construct space time (and space frequency) codes for for the coherent and non-coherent channel case. Further research [17] led to space time codes with reduced design complexity by utilizing III.5.

Remark III.6

A related question arises, when one considers the task of given a code $\mathcal{C}^U$, does there exist a code $\mathcal{C}^V$ with the same rate but better performance than $\mathcal{C}^U$? Concerning the diversity sum d a partial answer gives [16]: The transmit power constraint sets the requirement $\sqrt{\frac{T}{n_t}}\overline{d}^{\min} \geq \overline{d}^{U\,\min}$. Since there exist a monotonically increasing lower bound for $\overline{d}^{\min}$ when $\frac{T}{n_t}$ grows (Proposition I.1) this requirement can be certainly fulfilled. This again emphasizes the need for coding strategies in the general coding spaces $V^{\mathbb{C}}_{n_t,T}$, $G^{\mathbb{C}}_{n_t,T}$, T larger than n_t. However, it remains an open question, whether we can achieve the goal by composed codes of the form $\mathcal{C}^V = \mathcal{C}^G \cdot \mathcal{C}^U$.

Remark III.7

A conceptual simple (but computational complex) embedding of $G^{\mathbb{C}}_{n_t,T}$ into $V^{\mathbb{C}}_{n_t,T}$ is given by the parametrization of $G^{\mathbb{C}}_{n_t,T}$ with (so-called 'horizontal') tangents $X^H = \left(\begin{smallmatrix} \mathbf{0} & -B^{\dagger} \\ B & \mathbf{0} \end{smallmatrix}\right)$, $B \in \mathbb{C}^{(T-n_t)\times n_t}$ in its total space $U(T)$. In a recent article [21] it has been shown, that coding for the non-coherent channel is under certain assumptions equivalent to coding on the horizontal tangent space, with respect to the coherent channel diversity for $V^{\mathbb{C}}_{n_t,T-n_t}$. Combining that with Theorem III.5 we can roughly state this correspondence as $V^{\mathbb{C}}_{n_t,T-n_t} \subset G^{\mathbb{C}}_{n_t,T} \subset V^{\mathbb{C}}_{n_t,T}$, which gives rise to a sequence $\ldots \to \mathcal{C}^V_i \to \mathcal{C}^G_{i+1} \to \mathcal{C}^V_{i+1} \to \ldots$ of codes with increasing block length $i \cdot n_t$, $i = 1, 2, \ldots$

IV. Extremal properties of the diversity

In this section we examine the distribution of pairwise angles in $\mathcal{C}^G$ to find criteria for maximum diversity in particular for the combined code $\mathcal{C}^V = \mathcal{C}^G \cdot \mathcal{C}^U$ in $\sqrt{\frac{T}{n_t}} V^{\mathbb{C}}_{n_t,T}$. We focus on the diversity sum and diversity product, representing the most important diversity quantities (since they dominate the small and high SNR regime of diversity) while still being simple functions of the principal angles.

To get some first insight into the interplay between diversity sum and product (with respect to a fixed pair Φ, Ψ

of code symbols) we exploit the homogeneity of the elementary symmetric polynomials. For both coherent and non-coherent channel case it is quite natural to write $\hat{\sigma} := \frac{\sigma}{\|\sigma\|}$, $\hat{s}_i := \frac{s_i}{\|\sigma\|^{2i}} = \mathrm{sym}_i(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_{n_t}^2)$, The importance of this factorization arises from the identity $d = \|\sigma\|$, thus we can now write

$$\mathcal{D}iv = \sum_{i=0}^{n_t} \hat{s}_i \left(d^2 \varrho\right)^i \tag{42'}$$

which emphasizes the intuitively obvious fact, that scaling of ϱ (resp. ρ) behaves reciprocal to scaling of the distances. Moreover, we see that the diversity scales (term wise) with (an appropriate power of) the metric d, which means in particular that the task of maximizing the diversity behaves in its higher order terms (especially the diversity product) like a constraint on the packing problem determined by the diversity sum, contrasting the impression one might have gotten by considering only the Chernov bound (35), which seems dominated by its highest order term. Consequently we have to control p constrained on the unit sphere S_d. In summary the homogeneity property (42') scales all orders of diversity by the pairwise metric distances, turning the diversity orders ≥ 2 into local quantities. Thus maximizing diversity corresponds roughly to locally maximizing the diversity product while globally maximizing the diversity sum (constrained packing problem). The behavior of the diversity product on large scales becomes unimportant due to the contributions of the lower order terms. Let us therefore perform a Lagrangian analysis for the diversity product constrained on the unit sphere.

Lagrangian analysis: The non-coherent channel diversity sum and product and the corresponding lower bounds for their coherent channel analogues (by embedding), are functions of type $H(\vartheta) = \sum_{i=1}^{n_t} h(\vartheta_i)$ or $K(\vartheta) = \prod_{i=1}^{n_t} h(\vartheta_i)$ with either $h = \sin^2$ (for $\underline{d}^2$ (29), $\underline{p}^2$ (44)) or $h = 2(1-\cos)$ (for $\underline{d}\mathrm{dist}^2$ (49), $\bar{p}\mathrm{dist}^2$ (50)). Their domain of definition is the closed simplex $\bar{\Theta}$ of principal angles (see figure 1)

$$\bar{\Theta} := \left\{ \vartheta = (\vartheta_1, \ldots, \vartheta_{n_t}) \mid 0 \leq \vartheta_1 \leq \cdots \leq \vartheta_{n_t} \leq \frac{\pi}{2} \right\} \tag{63}$$

(the open simplex being $\Theta := \{\vartheta \in \mathbb{R}^{n_t} \,|\, 0 < \vartheta_1 < \cdots < \vartheta_{n_t} < \pi/2\}$) but since the principal angles vary

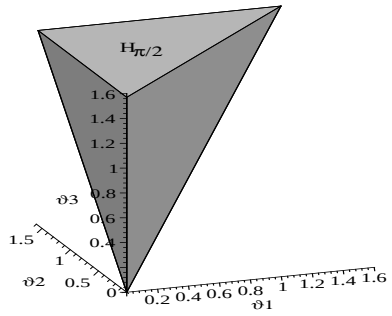

Fig. 1. $\bar{\Theta}$ for $n_t = 3$

like the identity map id for $0 \leq \vartheta_i \leq \pi/2$ but extend to $\pi - \mathrm{id}$ for $\pi/2 < \vartheta_i < \pi$ (considered as a function on the aperture angle), the function h (i.e. $1 - \cos$) fails in general to be differentiable transversal to the closed facet $H_{\pi/2}$ of $\bar{\Theta}$ containing $\vartheta_{n_t} = \pi/2$. Transversal to any other edge h is smooth, of course. In order to apply the classical Lagrangian formalism of constrained optimization problems in $\mathbb{R}^{n_t}$ to the present situation, we must convince ourselves, that the non-smooth edges and the $\pi/2$-facet of $\bar{\Theta}$ do not interfere. Our next task therefore consists of an appropriate decomposition of $\bar{\Theta}$ into smooth pieces, which decompose the optimization into a series of smaller tasks of one single type, solvable simultaneously in $\mathbb{R}^l$, $l \leq n_t$ (formula (67) shows the resulting problem formulation).

We need a little bit more notation. Let $\bar{\Theta}_< := \bar{\Theta} \setminus H_{\pi/2}$, and $\partial\Theta_< := \bar{\Theta}_< \setminus \Theta$ the C^0 boundary manifold of Θ with the problematic facet removed. For $l = 0, \ldots, k$ the faces contained in $\bar{\Theta}_<$ of dimension l are given by

$$\partial^{(l)}\Theta_< := \left\{ \begin{array}{l|c} 0 = \vartheta_1 = \cdots = \vartheta_{p_1} & \\ \quad < \vartheta_{p_1+1} = \cdots = \vartheta_{p_2} & p_1 < \cdots < p_l \\ \quad \vdots & p_i = i - 1 \ldots n_t - (l - i + 1) \\ \quad < \vartheta_{p_l+1} = \cdots = \vartheta_{n_t} & \end{array} \right\}$$

thus $\partial^{(l)}\Theta_<$ consists exactly of those faces in $\bar{\Theta}_<$, which are given by p_1 (possibly $= 0$) zero angles followed by l 'blocks' each of equal nonzero angles, in increasing order, in particular $\partial^{(0)}\Theta_< = \{0\}$, $\partial^{(n_t)}\Theta_< = \Theta$. Each face $\partial^{(l)}\Theta_<$ is a smooth submanifold of $\partial\Theta_<$, with $\partial(\partial^{(l)}\Theta_<) = \partial^{(l-1)}\Theta_<$ and $\bar{\Theta}_< = \dot{\cup}_{l=1}^{n_t} \partial^{(l)}\Theta_<$. The tangent spaces are

$$T\left(\partial^{(l)}\Theta_<\right) = \left\{ \sum_{i=1}^{l} \lambda_i (e_{p_i+1} + \cdots + e_{p_{i+1}}) \,\middle|\, \begin{array}{c} \lambda_i \in \mathbb{R} \\ p_1 < \cdots < p_l < p_{l+1} = n_t \\ p_i = i - 1 \ldots n_t - (l - i + 1) \end{array} \right\}$$

Then we have

Lemma IV.1

Given $H(\vartheta) = \sum_{i=1}^{n_t} h(\vartheta_i)$, $K(\vartheta) = \prod_{i=1}^{n_t} h(\vartheta_i)$ on $\bar{\Theta}_<$ with $h \in C^\infty([0, \pi/2[)$ (this means, that h is differentiable from the right in 0) and $h'(0) = 0$. Then for $l = 0, \ldots, n_t$

$$\mathrm{grad}\, H_{|\partial^{(l)}\Theta_<} \in T\left(\partial^{(l)}\Theta_<\right) \tag{64}$$

$$\mathrm{grad}\, K_{|\partial^{(l)}\Theta_<} \in T\left(\partial^{(l)}\Theta_<\right) \tag{65}$$

thus restricting the gradients remains intrinsic.

Proof:

We have $\mathrm{grad}\, H(\vartheta) = (h'(\vartheta_1), \ldots, h'(\vartheta_{n_t}))$. Since $h'(0) = 0$ the symmetry of H ensures $\mathrm{grad}\, H_{|\partial^{(l)}\Theta_<}(\vartheta) = \sum_{i=1}^{l} h'(\vartheta_{p_i+1})(e_{p_i+1} + \cdots + e_{p_{i+1}}) \in T\left(\partial^{(l)}\Theta_<\right)$, with $\lambda_i = h'(\vartheta_{p_i+1})$. Similarly, $\mathrm{grad}\, K(\vartheta) = (h'(\vartheta_i) \prod_{j \neq i} h(\vartheta_j))_i$, thus $\mathrm{grad}\, K_{|\partial^{(l)}\Theta_<}(\vartheta) = 0$ (for $p_1 > 0$) or $\mathrm{grad}\, K_{|\partial^{(l)}\Theta_<}(\vartheta) = \sum_{i=1}^{l} \lambda_i (e_{p_i+1} + \cdots + e_{p_{i+1}}) \in T\left(\partial^{(l)}\Theta_<\right)$ with $\lambda_i = h'(\vartheta_{p_i+1}) h(\vartheta_{p_i+1})^{p_{i+1} - p_i - 1} \prod_{j \neq i} h(\vartheta_{p_j+1})^{p_{j+1} - p_j}$. ■

The lemma ensures, that the Lagrangian functional $F = f - \lambda(g - \delta)$ on a neighborhood of $\bar{\Theta}$ for critical points of f obeying the constraint $g = \delta$ (f, g either H or K, $\delta \in \mathbb{R}$) applies to the boundary $\partial\Theta_<$. Since f and g are not necessarily

differentiable transversal to $H_{\pi/2}$, extremal points in $\bar{\Theta}$ lie in the set $\{g=\delta\}\cap H_{\pi/2}\cup\bar{\Theta}_<\cap\{DF=0\}$ (recall, that $\cap$ has higher precedence than $\cup$ and D denotes differentiation). For a more unified treatment we define $\bar{\Theta}^{(l)}:=\{0\le\vartheta_1\le\cdots\le\vartheta_l\le\pi/2\}$ (in particular $\bar{\Theta}^{(0)}=\emptyset$ and $\bar{\Theta}^{(n_t)}=\bar{\Theta}$) and recall, that on the one hand f, g are differentiable tangential to $H_{\pi/2}$ and on the other hand $\bar{\Theta}^{l-1}\cup\{\vartheta_l=\cdots=\vartheta_{n_t}=\pi/2\}$, $l=1,\dots,n_t-1$ exhausts $H_{\pi/2}$. This leads to the following recursion scheme: For $l=0,\dots,n_t$ set

$$\begin{aligned}H^{(l)}(\vartheta_1,\dots,\vartheta_l)&=\sum_{i=1}^{l}h(\vartheta_i)+h(\tfrac{\pi}{2})(n_t-l)\\K^{(l)}(\vartheta_1,\dots,\vartheta_l)&=\prod_{i=1}^{l}h(\vartheta_i)\cdot h(\tfrac{\pi}{2})^{(n_t-l)}\end{aligned}\tag{66}$$

and $f^{(l)},g^{(l)}$ given by either $H^{(l)}$ or $K^{(l)}$. Then the extremal points in $\bar{\Theta}$ lie in the set

$$\bigcup_{l=0}^{n_t}\Big(\{DF^{(l)}=0\}\cap\bar{\Theta}^{(l)}_<\cup\{\vartheta_{l+1}=\cdots=\vartheta_{n_t}=\frac{\pi}{2}\}\Big)\tag{67}$$

whereas $F^{(l)}:=f^{(l)}-\lambda(g^{(l)}-\delta)$. Furthermore, for h monotonically increasing on $[0,\pi/2]$ and zero at 0, the conditions $g^{(l)}=\delta$ forces $\delta\in[0,h(\frac{\pi}{2})n_t]$ (resp. $\delta\in[0,h(\frac{\pi}{2})^n_t]$) and in the first case $l\ge\lceil n_t-\delta/h(\frac{\pi}{2})\rceil=:\underline{l}$, thus l restricts to $\{\underline{l},\dots,n_t\}$.

Let us now start the Lagrangian analysis of the diversity (resp. with the analysis of the various diversity sums and products). The non-coherent channel diversity sum/product $(\underline{d})(\underline{p})$ as well as the lower bounds $(\bar{d}\mathrm{dist})(\bar{p}\mathrm{dist})$ for the coherent channel analogues depend on the $U(n_t)$-fibers only (moreover they depend only on the principal angles). By left invariance we can always assume $\Psi=\left(\begin{smallmatrix}\mathbf{1}\\\mathbf{0}\end{smallmatrix}\right)$ and consider the diversity terms as functions (marked with an $\underline{0}$) on the single argument Φ, for which $\langle\Phi\rangle$ is separated by principal angles $\vartheta\in\bar{\Theta}$ from $\langle\left(\begin{smallmatrix}\mathbf{1}\\\mathbf{0}\end{smallmatrix}\right)\rangle$.

$\underline{p}^2_{\underline{0}}\big|_{S_{\underline{d}^2}(\delta)}$:
In order to find the maximum of $\underline{p}$ in the unit distance sphere we constrain $f=\underline{p}^2_{\underline{0}}=\prod_i\sin^2$ on $S_{\underline{d}^2}(\delta)$, by setting $g=\underline{d}^2_{\underline{0}}=\sum_i\sin^2$. Thus we get the Lagrangian functional $F(\vartheta)=\prod_i\sin^2\vartheta_i-\lambda(\sum_i\sin^2\vartheta_i-\delta)$, $0<\delta\le n_t$. Here we have $\underline{l}=\lceil n_t-\delta\rceil$ and from (67) we get for each $l\in\{\underline{l},\dots,n_t\}$

$$\{DF^{(l)}=0\}\cap\bar{\Theta}^{(l)}_<=\Big\{g^{(l)}=\delta\Big\}\cap\left\{\begin{array}{l|l}0=\vartheta_1=\cdots=\vartheta_{p_1}&\\\quad<\vartheta_{p_1+1}\le\cdots\le\vartheta_l<\dfrac{\pi}{2}&\lambda=\displaystyle\prod_{i\ne j}^{l}\sin^2\vartheta_i,\ \forall_{p_1<j\le l}\end{array}\right\}$$

From this we get extremal points with $f\ne0$ only for $p_1=0$, $\vartheta_1=\cdots=\vartheta_l=\theta^{(l)}$ with $l\sin^2\theta^{(l)}=\delta-(n_t-l)$ and $f^{(l)}=(\sin^2\theta^{(l)})^l=(1-\frac{n_t-\delta}{l})^l$, monotonically increasing with l, therefore

$$\max\underline{p}^2_{\underline{0}}\big|_{S_{\underline{d}^2}(\delta)}=f^{(n_t)}=\left(\frac{\delta}{n_t}\right)^{n_t}\tag{68}$$

attained in $\vartheta_1=\dots\vartheta_{n_t}=\theta$, $\sin^2\theta=\delta/n_t$.

Conclusion IV.2

Locally the code points of $\mathcal{C}^G$ for the non-coherent channel have to be distributed with as many of their pairwise principal angles to be nonzero and equal in modulus as possible.

In principle, the same holds for the coherent channel, if we consider $\bar{p}$ and $\bar{d}$ as functions of $\bar{\sigma}$: The maximum diversity product is attained for $\bar{\sigma}_1=\cdots=\bar{\sigma}_{n_t}=\mathfrak{s}$, $\mathfrak{s}^2=\delta/n_t$, $\bar{p}^2=\mathfrak{s}^{2n_t}=(\delta/n_t)^{n_t}$, but it seems difficult to characterize all $\Phi\in S_{\bar{d}^2}(\delta)$ subject to $\bar{\sigma}_1=\dots\bar{\sigma}_{n_t}=\mathfrak{s}$. Let us embed $\mathcal{C}^G$ into $V^{\mathbb{C}}_{n_t,T}$ instead and investigate the question, which conditions have to be imposed on $\mathcal{C}^G$ in order to achieve some diversity gain in terms of $\bar{d}\mathrm{dist}$ and $\bar{p}\mathrm{dist}$.

$\bar{\boldsymbol{d}}\mathbf{dist}^{\mathbf{2}}_{\underline{\mathbf{0}}}\big|_{\boldsymbol{S_{\underline{d}^2}(\delta)}}$:
For the metric fiber distance $\bar{d}\mathrm{dist}$ constrained on $S_{\bar{d}^2}(\delta)$ the Lagrangian functional reads $F(\vartheta)=\bar{d}\mathrm{dist}^2_{\underline{0}}(\vartheta)-\lambda(\underline{d}^2_{\underline{0}}(\vartheta)-\delta)=2k-2\sum_i\cos\vartheta_i-\lambda(\sum_i\sin^2\vartheta_i-\delta)$. Again we have $0<\delta\le n_t$, $\underline{l}=\lceil n_t-\delta\rceil$ and from (67) we now get for each $l\in\{\underline{l},\dots,n_t\}$

$$\{DF^{(l)}=0\}\cap\bar{\Theta}^{(l)}_<=\Big\{g^{(l)}=\delta\Big\}\cap\left\{\begin{array}{l|l}0=\vartheta_1=\cdots=\vartheta_{p_1}&\\\quad<\vartheta_{p_1+1}\le\cdots\le\vartheta_l<\dfrac{\pi}{2}&\lambda=\dfrac{1}{\cos\vartheta_j},\ \forall_{p_1<j\le l}\end{array}\right\}$$

Extremal points are contained in $0=\vartheta_1=\cdots=\vartheta_{p_1}$, $\vartheta_{p_1+1}=\cdots=\vartheta_l=\theta^{(l)}$ subject to $(l-p_1)\sin^2\theta^{(l)}=\delta-(n_t-l)$ for $l=\underline{l}=\lceil n_t-\delta\rceil,\dots,n_t$ and $p_1=0,\dots,\bar{p}_1=\lfloor n_t-\delta\rfloor$, with $f^{(l)}_{p_1}=2(l-p_1)(1-\cos\theta^{(l)})+2(n_t-l)=2(l-p_1)(1-\sqrt{1-q})+2(n_t-l)$, $q:=\frac{\delta-n_t+l}{l-p_1}=\sin^2\theta^{(l)}$. For fixed l the function $p_1\longmapsto f^{(l)}_{p_1}$ is monotonically increasing (by analyzing the derivative, where defined) and we find $\underline{f}^{(l)}:=\min_{p_1}f^{(l)}_{p_1}=f^{(l)}_0$ and $\bar{f}^{(l)}:=\max_{p_1}f^{(l)}_{p_1}=f^{(l)}_{\bar{p}_1}$. As functions of l both terms turn out to be monotonically decreasing (by analyzing the derivatives with respect to l) and we find

$$\min\bar{d}\mathrm{dist}^2_{\underline{0}}\big|_{S_{\underline{d}^2}(\delta)}=\underline{f}^{(n_t)}=2n_t\left(1-\sqrt{1-\delta/n_t}\right)\tag{69a}$$

attained in $\vartheta_1=\dots\vartheta_{n_t}=\theta$, $\sin^2\theta=\delta/n_t$, and

$$\max\bar{d}\mathrm{dist}^2_{\underline{0}}\big|_{S_{\underline{d}^2}(\delta)}=\bar{f}^{(\underline{l})}=2\left(1-\sqrt{1-(\delta-\lfloor\delta\rfloor)}\right)+2\lfloor\delta\rfloor\tag{69b}$$

attained for $\vartheta_1=\dots\vartheta_{\lfloor n_t-\delta\rfloor}=0$, $\sin^2\vartheta_{\lceil n_t-\delta\rceil}=\delta-\lfloor\delta\rfloor$, $\vartheta_{\lceil n_t-\delta\rceil+1}=\dots\vartheta_{n_t}=\pi/2$.

$\bar{\boldsymbol{p}}\mathbf{dist}^{\mathbf{2}}_{\underline{\mathbf{0}}}\big|_{\boldsymbol{S_{\underline{d}^2}(\delta)}}$:
Examining $\bar{p}\mathrm{dist}$ instead of $\bar{d}\mathrm{dist}$ we have the Lagrangian $F(\vartheta)=2^k\prod_i(1-\cos\vartheta_i)-\lambda(\sum_i\sin^2\vartheta_i-\delta)$ and

$$\{DF^{(l)}=0\}\cap\bar{\Theta}^{(l)}_<=\Big\{g^{(l)}=\delta\Big\}\cap\left\{\begin{array}{l|l}0=\vartheta_1=\cdots=\vartheta_{p_1}&\\\quad<\vartheta_{p_1+1}\le\cdots\le\vartheta_l<\dfrac{\pi}{2}&\lambda=\dfrac{2^{n_t-1}}{\cos\vartheta_j}\displaystyle\prod_{i\ne j}^{l}(1-\cos\vartheta_i)\\&\qquad\forall_{p_1<j\le l}\end{array}\right\}$$

From this we get extremal points with $f \neq 0$ only for $p_1 = 0$, $\vartheta_1 = \cdots = \vartheta_l = \theta^{(l)}$ with $l \sin^2\theta^{(l)} = \delta - (n_t - l)$ and $f^{(l)} = 2^{n_t}(1-\cos\theta^{(l)})^l = 2^{n_t}(1-\sqrt{\frac{n_t-\delta}{l}})^l$, monotonically increasing with l and therfore

$$\max \bar{p}\mathrm{dist}^2_{\underline{0}|S_{\underline{d}^2}(\delta)} = f^{(n_t)} = 2^{n_t}\left(1-\sqrt{1-\delta/n_t}\right)^{n_t} \quad (70)$$

in $\vartheta_1 = \ldots \vartheta_{n_t} = \theta$, $\sin^2\theta = \delta/n_t$.

Remark IV.3

If we had constrained $\bar{p}\mathrm{dist}$ on $S_{\bar{d}\mathrm{dist}^2}(\delta)$, $0 < \delta \leq 2n_t$ we would have gotten

$$\max \bar{p}\mathrm{dist}^2_{\underline{0}|S_{\bar{d}\mathrm{dist}^2}(\delta)} = \left(\frac{\delta}{n_t}\right)^{n_t} \quad (71)$$

attained in $\vartheta_1 = \ldots \vartheta_{n_t} = \theta$, $2(1-\cos\theta) = \delta/n_t$, by analogy to the case of $\underline{p}$ constrained on $\underline{d}$. The apparent discrepancy of (70) to (71) is caused by (69b), since at $\vartheta_1 = \cdots = \vartheta_{n_t} = \theta$ $\bar{d}\mathrm{dist}^2_{\underline{0}}$ attains its minimum $\tilde{\delta} = 2n_t(1-\sqrt{1-\delta/n_t})$ on $S_{\underline{d}^2}(\delta)$ which yields (with (71)) $\bar{p}\mathrm{dist}^2_{\underline{0}|\bar{d}\mathrm{dist}^2_{\underline{0}}=\tilde{\delta}} = (\tilde{\delta}/n_t)^{n_t} = 2^{n_t}(1-\sqrt{1-\delta/n_t})^{n_t}$, which coincides with (70).

Conclusion IV.4

What remains is the general rule, that for $\bar{\varrho} = \frac{pT}{4n_t} \gg 1$ (compare (36)) in favor of $\bar{p}\mathrm{dist}$ instead of $\bar{d}\mathrm{dist}$ one should distribute the pairwise principal angles in $\mathcal{C}^G$ locally to be all nonzero and equal in modulus (by (70)*). This coincides with the preferred strategy* (68) *for the non-coherent channel code. For $\bar{\varrho} \leq 1$, when the higher order diversity terms become less important, it might be better to distribute the pairwise principal angles globally to maximize $\bar{d}\mathrm{dist}$, thus separating points in $\mathcal{C}^G$ by as many of the principal angles to be zero such that the remaining ones attain large values.*

$\bar{p}\mathrm{dist}^2_{\underline{0}|S_{\underline{p}^2}(\delta)}$:

Finally, to get an product analogue of (69) we analyze $\bar{p}\mathrm{dist}^2_{\underline{0}}$ constrained on $S_{\underline{p}^2}(\delta)$, $0 < \delta \leq 1$, thus the Lagrangian reads $F(\vartheta) = 2^k\prod_i(1-\cos\vartheta_i) - \lambda(\prod_i \sin^2\vartheta_i - \delta)$ and for $l \in \{0,\ldots,n_t\}$ we have

$$\{DF^{(l)} = 0\} \cap \bar{\Theta}^{(l)}_{<} = \left\{g^{(l)} = \delta\right\} \cap$$
$$\left\{ \begin{array}{l|l} 0 = \vartheta_1 = \cdots = \vartheta_{p_1} & \\ \quad < \vartheta_{p_1+1} \leq \cdots \leq \vartheta_l < \dfrac{\pi}{2} & \lambda = \dfrac{2^{n_t-1}}{\cos\vartheta_j}\displaystyle\prod_{i\neq j}^{l}\frac{1-\cos\vartheta_i}{\sin^2\vartheta_i} \\ & \forall_{p_1<j\leq l} \end{array} \right\}$$

$\delta > 0$ forces $p_1 = 0$ and we get $\vartheta_1 = \cdots = \vartheta_l = \theta^{(l)}$, $\sin^{2l}\theta^{(l)} = \delta$, thus $f^{(l)} = 2^{n_t}(1-\cos\theta^{(l)})^l = 2^{n_t}(1-\sqrt{1-\delta^{1/l}})^l$ which is monotonically decreasing in l, so

$$\min \bar{p}\mathrm{dist}^2_{\underline{0}|S_{\underline{p}^2}(\delta)} = f^{(n_t)} = 2^{n_t}\left(1-\sqrt{1-\delta^{1/n_t}}\right)^{n_t} \quad (72a)$$

attained in $\vartheta_1 = \ldots \vartheta_{n_t} = \theta$, $\sin^2\theta = \delta^{1/n_t}$, and

$$\max \bar{p}\mathrm{dist}^2_{\underline{0}|S_{\underline{p}^2}(\delta)} = f^{(1)} = 2^{n_t}\left(1-\sqrt{1-\delta}\right) \quad (72b)$$

attained in $\sin^2\vartheta_1 = \delta$, $\vartheta_2 = \cdots = \vartheta_{n_t} = \pi/2$ (whereas $f^{(0)}$ is contained in $f^{(1)}$ when $\delta = 1$).

Observing, that the minima in (69) and (72) are monotonically increasing in δ we end up with

Proposition IV.5

In the situation of Proposition III.4 we have

$$\bar{d}\mathrm{dist}^{2\min} \geq \min \bar{d}\mathrm{dist}^2_{\underline{0}|S_{\underline{d}^2}(\underline{d}^{2\min})} \quad (73)$$
$$= 2n_t\left(1-\sqrt{1-\underline{d}^{2\min}/n_t}\right) \geq \underline{d}^{2\min}$$
$$\bar{p}\mathrm{dist}^{2\min} \geq \min \bar{p}\mathrm{dist}^2_{\underline{0}|S_{\underline{p}^2}(\underline{p}^{2\min})} \quad (74)$$
$$= 2^{n_t}\left(1-\sqrt{1-\underline{p}^{2\min 1/n_t}}\right)^{n_t} \geq \underline{p}^{2\min}$$

The benefit of this proposition is, that it relates $\bar{d}\mathrm{dist}^{\min}$ (resp. $\bar{p}\mathrm{dist}^{\min}$) directly to $\underline{d}^{\min}$ (resp. $\underline{p}^{\min}$), regardless if the minimum distances (resp. diversity products) are realized by the same pair of points or not.

Proof:

Its only left up to show the second inequality in each formula, which is elementary (setting $x := \underline{d}^{2\min}$ or $x := \underline{p}^{2\min}$, respectively):

(73) $\iff$ $\sqrt{1-x/n_t} \leq 1-x/2n_t$ $\iff$ $1-x/n_t \leq (1-x/2n_t)^2 = 1-x/n_t+x^2/4n_t^2$.

(74) $\iff$ $\sqrt{1-x^{1/n_t}} \leq 1-x^{1/k}/2$ $\iff$ $1-x^{1/n_t} \leq (1-x^{1/k}/2)^2 = 1-x^{1/n_t}+x^{2/n_t}/4$ ∎

V. Conclusions

This work should be seen as a second step towards a geometry based analysis of general space time block codes, inspired by the results in [16], opening the door to potentially high performing space time block codes, when $T \gg n_t$. The various estimates and interrelations explored in this work assemble the following overall picture:

- **Diversity monotony:** The performance analysis revealed nice embedding properties (with respect to $G^{\mathbb{C}}_{n_t,T} \subset V^{\mathbb{C}}_{n_t,T}$) of the diversity quantities (Corollaries III.2, III.3, (57)), leading to a diversity growth (Proposition III.4) in the transition from the non-coherent channel to the coherent channel. This turned out to be due to the various invariance properties satisfied by the diversity, though tied to distinct underlying topologies of the coding spaces induced by the maximum likelihood receiver metrics. Moreover, for the diversity sum and product, more explicit estimates have been derived (Proposition IV.5).
- **Complexity reduction:** Embeddings of both $G^{\mathbb{C}}_{n_t,T}$ and $U(n_t)$ into $V^{\mathbb{C}}_{n_t,T}$ can be used to construct codes on $V^{\mathbb{C}}_{n_t,T}$ from 'smaller' pieces (Theorem III.5), both of them being already in the focus of current research. The other way round, given an non-coherent channel space time code and a 'small' coherent channel code, the performance of the resulting (larger dimensional) product code on $V^{\mathbb{C}}_{n_t,T}$ is lower bounded by the diversity expressions stated in the theorem. Thus the design complexity has been reduced to the smaller problems on $G^{\mathbb{C}}_{n_t,T}$ and $U(n_t)$. Together with Proposition I.1 this opens the door to potentially

high performing space time block codes, when $T \gg n_t$. As already indicated in the introduction this may be of some importance in the context of space frequency codes also.

- **Localization:** The local nature of the higher order diversity quantities turns space time coding into a constrained packing problem. The diversity sum still represents a major criteria, locally superposed by the diversity product as a rigidity constraint: The optimal code in the high SNR regime is packed as 'diagonal' as possible, uniformly maximizing the principal angles (Conclusion IV.2 and IV.4).

There are still many open issues. Some immediate will be listed next. The explicit bounds of Proposition IV.5 are very coarse and improvements are necessary. Moreover, it would be desirable to obtain further decompositions in Theorem III.5. Furthermore this work has to be related to the differential coding scheme [9], which benefits from high rates compared to non differential codes. Finally it remains the challenge of effective high dimensional code construction (especially for the non-coherent channel) with low complexity decoding properties.

ACKNOWLEDGMENT

I would like to thank Eduard Jorswieck, Peter Jung, and Aydin Sezgin for reading the manuscript and helpful comments.

REFERENCES

[1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, 1998.

[2] B. M. Hochwald and T. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, pp. 543–565, 2000.

[3] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Select. Areas in Communications*, vol. 16, pp. 1451–1458, 1998.

[4] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1456–1467, 1999.

[5] B. Hassibi and B. M. Hochwald, "High-rate codes that are linear in space and time," *IEEE Trans. Inform. Theory*, vol. 48, no. 7, pp. 1804–1824, 2002.

[6] R. H. Gohary and T. N. Davidson, "Design of linear dispersion codes: Asymptotic guidelines and their implementation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2892–2906, 2005.

[7] J.-C. Belfiore, G. Rekaya, and E. Viterbo, "The golden code: A 2×2 full-rate space-time code with nonvanishing determinants," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1432–1436, 2005.

[8] V. Tarokh, "Existence and construction of non-coherent unitary space-time codes," preprint, url: http://www.mit.edu/~vahid/sample.html.

[9] B. M. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Comm.*, vol. 48, pp. 2041–2052, 2000.

[10] A. Shokrollahi, B. Hassibi, B. M. Hochwald, and W. Sweldens, "Representation theory for high-rate multiple-antenna code design," *IEEE Trans. Inform. Theory*, vol. 47, pp. 335–2367, 2001.

[11] G. Han and J. Rosenthal, "Unitary space time constellation analysis: An upper bound for the diversity," 2004, preprint arXiv:math.CO/0401045.

[12] X.-B. Liang and X.-G. Xia, "Unitary signal constellations for differential space-time modulation with two transmit antennas: Parametric codes, optimal designs, and bounds," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2291–2322, 2002.

[13] H. Wang, G. Wang, and X.-G. Xia, "Some 2×2 unitary space-time codes from sphere packing theory with optimal diversity product of code size 6," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3361–3368, 2004.

[14] B. M. Hochwald and T. Marzetta, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, 1999.

[15] A. Barg and D. Y. Nogin, "Bounds on packings of spheres in the Grassmann manifold," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2450–2454, 2002.

[16] O. Henkel, "Sphere packing bounds in the Grassmann and Stiefel manifolds," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3445–3456, 2005.

[17] ——, "Space time codes from permutation codes," in *Proceedings of IEEE GlobeCom 2006*, San Francisco, California, 2006.

[18] H. Bölcskei, M. Borgmann, and A. J. Paulraj, "Space-Frequency coded MIMO-OFDM with variable multiplexing-diversity tradeoff," in *IEEE International Conference on Communications (ICC'03)*, vol. 4, 2003, pp. 2837– 2841.

[19] H. Bölcskei and M. Borgmann, "Code design for non-coherent MIMO-OFDM systems," in *Proc. 40th Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, 2002, pp. 237–246.

[20] O. Henkel, "Space frequency codes from spherical codes," in *Proceedings of the 2005 IEEE International Symposium on Information Theory (ISIT 05)*, 2005, pp. 1305–1309. [Online]. Available: http://arxiv.org/abs/cs.IT/0501085

[21] I. Kammoun and J.-C. Belfiore, "A new family of Grassmann space-time codes for non-coherent MIMO systems," *IEEE Comm. Letters*, vol. 7, no. 11, pp. 528–530, 2003.

[22] W. M. Boothby, *An Introduction to Differentiable Manifolds and Riemannian Geometry*, ser. Pure and Applied Mathematics 120. Academic Press, Inc., Orlando, FL, 1986.

[23] L. Conlon, *Differentiable Manifolds: A first course*. Birkhäuser Boston, 1993.

[24] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*, 2nd ed. Springer, 1993.

[25] W. A. Poor, *Differential Geometric Structures*. McGraw-Hill Inc., 1981.

[26] A. Edelman, T. A. Arias, and S. T. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, pp. 303–353, 1998.

[27] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, 2002.

[28] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions on Telecommunications*, vol. 10, pp. 585–595, 1999.

[29] Å. Björck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Mathematics of Computation*, vol. 27, no. 123, pp. 579–594, 1973.

[30] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: Packings in Grassmannian spaces," *Experimental Mathematics*, vol. 5, pp. 139–159, 1996. [Online]. Available: http://www.research.att.com/∼njas/grass/index.html

[31] R. Bhatia, *Matrix Analysis*. Springer, 1997.

[32] O. Henkel and G. Wunder, "Space frequency codes from sphere packings," in *Proceedings International ITG/IEEE Workshop on Smart Antennas (WSA 2005)*, Duisburg-Essen, Germany, Apr 2005.