

# Degrees of Freedom in Vector Interference Channels

David Stotz and Helmut Bölcskei, *Fellow, IEEE*

**Abstract**—This paper continues the Wu-Shamai-Verdú program [3] on characterizing the degrees of freedom (DoF) of interference channels (ICs) through Rényi information dimension. Specifically, we find a single-letter formula for the DoF of vector ICs, encompassing multiple-input multiple-output (MIMO) ICs, time- and/or frequency-selective ICs, and combinations thereof, as well as scalar ICs as considered in [3]. The DoF-formula we obtain lower-bounds the DoF of *all* channels—with respect to the choice of the channel matrix—and upper-bounds the DoF of *almost all* channels. It applies to a large class of noise distributions, and its proof is based on an extension of a result by Guionnet and Shlyakthenko [4] to the vector case in combination with the Ruzsa triangle inequality for differential entropy introduced by Kontoyiannis and Madiman [5]. As in scalar ICs, achieving full DoF requires the use of singular input distributions. Strikingly, in the vector case it suffices to enforce singularity on the joint distribution of each transmit vector. This can be realized through signaling in subspaces of the ambient signal space, which is in accordance with the idea of interference alignment, and, most importantly, allows the scalar entries of the transmit vectors to have non-singular distributions. The DoF-formula for vector ICs we obtain enables a unified treatment of “classical” interference alignment à la Cadambe and Jafar [6], and Maddah-Ali et al. [7], and the number-theoretic schemes proposed in [8], [9]. Moreover, it allows to calculate the DoF achieved by new signaling schemes for vector ICs. We furthermore recover the result by Cadambe and Jafar on the non-separability of parallel ICs [10] and we show that almost all parallel ICs are separable in terms of DoF. Finally, our results apply to complex vector ICs, thereby extending the main findings of [3] to the complex case.

## I. INTRODUCTION

Sparked by the surprising finding of Cadambe and Jafar [6] stating that  $K/2$  degrees of freedom (DoF) can be realized in  $K$ -user interference channels (ICs) through interference alignment, the study of DoF in wireless networks has seen significant activity in recent years. The essence of interference alignment is to exploit channel variations in time/frequency/space to align interference at the receiving terminals in low-dimensional subspaces. This is accomplished through a clever vector-signaling scheme.

Following the discovery in [6] it was shown that the basic idea of aligning interference to realize DoF can be applied to numerous further settings [11]. Perhaps the most surprising of these results is that  $K/2$  DoF can be realized in  $K$ -user scalar ICs with constant channel matrix [8], [9], i.e., in complete

absence of channel variations. Moreover, it is shown in [8] that  $K/2$  DoF are achievable for almost all (with respect to the channel matrix) constant scalar ICs. The schemes introduced in [8], [9] use Diophantine approximation or lattice structures to design full DoF-achieving transmit signals.

In a tour de force Wu et al. [3] discovered that Rényi information dimension<sup>1</sup> [12] is a suitable tool for systematically characterizing the DoF achievable in constant scalar ICs. Specifically, it follows from the results in [3] that the real interference alignment schemes proposed in [8], [9] correspond to the use of singular input distributions. What is more, Wu et al. [3] found that  $K/2$  DoF can be achieved in a  $K$ -user constant scalar IC only by input distributions that have a non-trivial singular component. This is a strong, negative, result as input distributions with a singular component are difficult to realize, or, more specifically, to approximate in practice. On the other hand, it is well-known that full DoF can be realized in ICs with channel variations in time/frequency or in constant multiple-input multiple-output (MIMO) ICs, even if the individual entries of the transmit vectors do not have singular components. Reconciling these two lines of results is one of the central goals of this paper.

*Contributions:* We continue the Wu-Shamai-Verdú program [3] by showing how information dimension can be used to characterize the DoF of vector ICs. This extension is relevant as “classical” interference alignment à la Cadambe and Jafar [6], and Maddah-Ali et al. [7] relies on vector-valued signaling. The vector IC we consider contains, as special cases, the MIMO IC, time- and/or frequency-selective ICs, and combinations thereof, as well as the constant scalar IC studied in [3], [8], [9]. The Wu-Shamai-Verdú theory builds on a little known but highly useful result by Guionnet and Shlyakthenko [4, Thm. 2.7], stating that the DoF in a scalar additive noise channel achieved by a given input distribution equal the input distribution’s information dimension. We present an extension of this result to the vector case and to more general noise distributions, and we give an information-theoretic proof based on (a slight generalization of) the Ruzsa triangle inequality for differential entropy as introduced by Kontoyiannis and Madiman [5]. The (generalized) Ruzsa triangle inequality is key for the extension to general noise distributions and, in addition, allows for a considerable simplification of the existing proofs (for the scalar case) in [13], [4].

Our main result is a single-letter formula for the DoF of vector ICs. This formula has the same structure as the one for the constant scalar case in [3]. It lower-bounds the DoF of *all* channels—with respect to the channel matrix—and upper-bounds the DoF of (Lebesgue) *almost all* channels,

The material in this paper was presented in part in [1] at the 50th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, Oct. 2012. We point out that the statement in [1, Thm. 1] needs to be restricted to hold true for almost all  $\mathbf{H}$  only. An Online Addendum [2] to this paper is available at [http://www.nari.ee.ethz.ch/commth/research/downloads/dof\\_addendum.pdf](http://www.nari.ee.ethz.ch/commth/research/downloads/dof_addendum.pdf).

The authors are with the Department of Information Technology and Electrical Engineering, ETH Zurich, CH-8092 Zurich, Switzerland (e-mail: {dstotz,boelcskei}@nari.ee.ethz.ch).

<sup>1</sup>In the remainder of the paper we call “Rényi information dimension” simply “information dimension”.

including channel matrices with all entries rational. While we adopt the strategy underlying the proof of the DoF-formula for constant scalar ICs [3, Sect. V-B], our extension of the result by Guionnet and Shlyakthenko [4, Thm. 2.7]—besides pertaining to the vector case—yields a DoF-formula that applies to a significantly larger class of noise distributions. Moreover, on a conceptual basis the DoF-formula we obtain leads to fundamentally new implications. Specifically, we find that while input distributions with a singular component are still needed to achieve full DoF in vector ICs, it suffices to enforce singularity on the joint distribution of each transmit vector. This form of singularity is realized by taking, e.g., the transmit vectors to live in lower-dimensional subspaces of the ambient signal space, as is, in fact, done in interference alignment à la Cadambe and Jafar [6], and Maddah-Ali et al. [7].

We demonstrate that our DoF-formula for vector ICs allows for a unified treatment of “classical” interference alignment as introduced in [6], [7] and the “number-theoretic” interference alignment schemes for constant scalar ICs as proposed in [8], [9]. In addition, the formula constitutes a tool for evaluating the DoF achieved by new signaling schemes for vector ICs.

Furthermore, we recover the result by Cadambe and Jafar on the non-separability of parallel ICs [10], and we show that almost all (again, with respect to the channel matrix) parallel ICs are separable in terms of DoF, i.e., for almost all parallel ICs independent coding across subchannels achieves full DoF. Finally, our results apply to complex signals and channel matrices, thereby extending the main result in [3] to the complex case.

*Notation:* Random vectors are represented by uppercase letters, deterministic vectors by lowercase letters, in both cases using letters from the end of the alphabet. Boldface uppercase letters are used to indicate matrices. The  $n \times n$  identity matrix is  $\mathbf{I}_n$  and the all-zeros matrix is  $\mathbf{0}$ . The  $nm \times nm$  block diagonal matrix with blocks  $\mathbf{A}_1, \dots, \mathbf{A}_n \in \mathbb{R}^{m \times m}$  on its main diagonal is denoted by  $\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_n)$ .  $\|\mathbf{A}\|_\infty := \max\{|a_{i,j}| \mid 1 \leq i \leq m, 1 \leq j \leq n\}$  stands for the  $\ell^\infty$ -norm of the matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ . For  $x \in \mathbb{R}$ , we write  $\lfloor x \rfloor$  for the largest integer not exceeding  $x$ . For  $k \in \mathbb{N} \setminus \{0\}$ , we set  $\langle x \rangle_k := \lfloor kx \rfloor / k$  and  $[x]_k := \langle x \rangle_{2^k}$ . The notation conventions  $\lfloor \cdot \rfloor$ ,  $\langle \cdot \rangle_k$ , and  $[\cdot]_k$  are extended to real vectors and matrices through application on an entry by entry basis. For  $x \in \mathbb{R}^n$ , let  $\|x\|_{\mathbb{Z}} := \min_{u \in \mathbb{Z}^n} \|x - u\|_\infty$ . We define the minimum and maximum distance of a set  $\mathcal{W} \subseteq \mathbb{R}^n$ , respectively, as  $m(\mathcal{W}) := \inf_{w_1, w_2 \in \mathcal{W}, w_1 \neq w_2} \|w_1 - w_2\|_\infty$  and  $M(\mathcal{W}) := \sup_{w_1, w_2 \in \mathcal{W}, w_1 \neq w_2} \|w_1 - w_2\|_\infty$ , and we extend these definitions to sets of matrices accordingly. For a discrete random matrix  $\mathbf{X}$ , we let  $H(\mathbf{X})$  be the entropy of the vector obtained by stacking the columns of  $\mathbf{X}$ . The differential entropy  $h(\mathbf{X})$  is defined analogously. All logarithms are to the base 2. By “ $a > 0$ ” we mean that the real constant  $a$  is positive and finite. For sets  $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^n$ ,  $m \in \mathbb{N}$ , and a scalar  $\alpha \in \mathbb{R}$ , we let  $\mathcal{A} \times \mathcal{B}$  denote the cartesian product,  $\mathcal{A}^m$  the  $m$ -th cartesian power of  $\mathcal{A}$ ,  $\mathcal{A} + \mathcal{B} := \{a + b \mid a \in \mathcal{A}, b \in \mathcal{B}\}$ , and  $\alpha\mathcal{A} := \{\alpha a \mid a \in \mathcal{A}\}$ . The dimension of a subspace  $\mathcal{V} \subseteq \mathbb{R}^n$  is denoted by  $\dim \mathcal{V}$  and  $\mathcal{V}^\perp$  represents the orthogonal complement of  $\mathcal{V}$  with respect to the Euclidean inner product.

$\mathbb{E}[\cdot]$  stands for the expectation operator and  $\stackrel{D}{=}$  means equality in distribution. For a measurable real-valued function  $f$  and a measure<sup>2</sup>  $\mu$  on its domain, the pushforward measure is given by  $(f_*\mu)(\mathcal{A}) = \mu(f^{-1}(\mathcal{A}))$  for Borel sets  $\mathcal{A}$ . Integration of  $f$  with respect to  $\mu$  is denoted by  $\int f(x)\mu\langle x \rangle$ , where for the special case of the Lebesgue measure  $\lambda$  we put  $dx := \lambda\langle x \rangle$ . We say that an integral  $\int f(x)\mu\langle x \rangle$  exists if the integral of the positive part  $\max\{f(x), 0\}$  or the integral of the negative part  $-\min\{f(x), 0\}$  is finite. Furthermore, the integral  $\int f(x)\mu\langle x \rangle$  is said to be finite if it exists and if  $|\int f(x)\mu\langle x \rangle| < \infty$ . The differential entropy  $h(X)$  of a random vector with density  $f_X$  is said to exist (to be finite), if the integral  $h(X) = \int_{\mathbb{R}^n} f_X(x) \log\left(\frac{1}{f_X(x)}\right) dx$  exists (is finite). For Borel sets  $\mathcal{A}$ , we write  $\mathbf{1}_{\mathcal{A}}(x)$  for the characteristic function on  $\mathcal{A}$ .

## II. SETUP AND DEFINITIONS

We consider a memoryless  $K$ -user (with  $K \geq 2$ ) vector IC with input-output (I/O) relation

$$Y_i = \sqrt{\text{snr}} \sum_{j=1}^K \mathbf{H}_{i,j} X_j + W_i, \quad i = 1, \dots, K, \quad (1)$$

where  $X_i = (X_i[1] \dots X_i[M])^T \in \mathbb{R}^M$  and  $Y_i = (Y_i[1] \dots Y_i[M])^T \in \mathbb{R}^M$  is the transmit and receive vector, respectively, corresponding to user  $i$ ,  $M$  is the dimension of the I/O signal space,  $\mathbf{H}_{i,j} \in \mathbb{R}^{M \times M}$  denotes the channel matrix between transmitter  $j$  and receiver  $i$ , and the  $W_i$  are i.i.d. zero-mean Gaussian random vectors with identity covariance matrix. For simplicity of exposition, we treat the real case throughout and show in Section V-E how our results can be extended to the complex case. Defining the  $KM \times KM$  matrix

$$\mathbf{H} := \begin{pmatrix} \mathbf{H}_{1,1} & \cdots & \mathbf{H}_{1,K} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{K,1} & \cdots & \mathbf{H}_{K,K} \end{pmatrix}, \quad (2)$$

we can rewrite (1) as

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_K \end{pmatrix} = \sqrt{\text{snr}} \mathbf{H} \begin{pmatrix} X_1 \\ \vdots \\ X_K \end{pmatrix} + \begin{pmatrix} W_1 \\ \vdots \\ W_K \end{pmatrix}. \quad (3)$$

The channel matrix  $\mathbf{H}$  is assumed to be known perfectly at all transmitters and receivers and remains constant across channel uses. For each user  $i = 1, \dots, K$ , we impose the average power constraint

$$\frac{1}{MN} \sum_{n=1}^N \sum_{m=1}^M \left(x_i^{(n)}[m]\right)^2 \leq 1, \quad (4)$$

on codeword matrices  $(x_i^{(1)} \dots x_i^{(N)})$  of blocklength  $N$ , where  $x_i^{(n)} = (x_i^{(n)}[1] \dots x_i^{(n)}[M])^T \in \mathbb{R}^M$ . In contrast to the setting in [3], where the action of each link is represented through scaling by a single coefficient, here each link acts as

<sup>2</sup>Throughout the paper, the terms “measurable” and “measure” are to be understood with respect to the Borel  $\sigma$ -algebra.

a linear operator (represented by a finite-dimensional matrix) on the corresponding transmit vector. To differentiate clearly, we henceforth refer to the setting in [3] as the “scalar IC”. Note that this terminology includes the channel matrix being constant, as opposed to the classical interference alignment setup in [6], which also deals with scalar ICs but has varying channel matrices.

The vector IC setting encompasses the MIMO IC (with  $M$  antennas at each user and  $\mathbf{H}_{i,j}$  denoting the MIMO channel matrix between transmitter  $j$  and receiver  $i$ ), and time-frequency-selective ICs, transformed into memoryless vector ICs through the use of guard periods/bands. For purely frequency-selective single-antenna ICs, e.g., the use of OFDM [14] results in a memoryless vector channel with diagonal  $\mathbf{H}_{i,j}$  matrices. If guard intervals (of sufficient length) filled with zeros are used instead of the cyclic prefix in OFDM, we get memoryless vector ICs with  $\mathbf{H}_{i,j}$  matrices that are not diagonal. For purely time-selective single-antenna ICs, we obtain a memoryless vector IC, again with diagonal  $\mathbf{H}_{i,j}$  matrices, without using guard regions, by simply identifying the main diagonal entries of the  $\mathbf{H}_{i,j}$ 's with the corresponding channel coefficients. When all the  $\mathbf{H}_{i,j}$  are diagonal we obtain an important special case of the vector IC in (3), namely a parallel IC (with  $M$  subchannels) as studied in [10]. Finally, the vector IC also covers time-frequency-selective MIMO ICs.

Let  $C_{\text{sum}}(\mathbf{H}; \text{snr})$  be the sum-capacity<sup>3</sup> for a given snr. The degrees of freedom<sup>4</sup> of the IC (1) are then defined as

$$\text{DoF}(\mathbf{H}) := \limsup_{\text{snr} \rightarrow \infty} \frac{C_{\text{sum}}(\mathbf{H}; \text{snr})}{\frac{1}{2} \log \text{snr}}. \quad (5)$$

We call  $\text{DoF}(\mathbf{H})/M$  the normalized DoF.

### III. RÉNYI INFORMATION DIMENSION

One of the main feats of [3] was to recognize that (Rényi) information dimension is a suitable tool for characterizing the DoF achievable in scalar ICs. The main conceptual contribution of the present paper is to show that information dimension is a natural tool for analyzing the DoF in vector ICs as well. This, in turn, leads to a unified framework for real interference alignment [8], [9]—hinging on number-theoretic properties of transmit signals and channel coefficients—and “classical” interference alignment relying on channel variations and vector signaling [6], [7], [11].

Throughout the paper we will deal with general distributions  $\mu$  on  $\mathbb{R}^n$ , the nature of which can often be understood better by invoking the following decomposition.

**Proposition 1.** *Every distribution  $\mu$  on  $\mathbb{R}^n$  can be decomposed uniquely as*

$$\mu = \alpha \mu_{ac} + \beta \mu_d + \gamma \mu_s, \quad (6)$$

<sup>3</sup>For a definition of the sum-capacity of discrete ICs see [15, Sect. 6.1]; for continuous alphabets with a cost constraint, as needed here, we refer to [15, Sections 3.3 and 3.4].

<sup>4</sup>For motivation on why to study this quantity see [11, App. A].

where  $\mu_{ac}$  is an absolutely continuous,  $\mu_d$  a discrete, and  $\mu_s$  a singular distribution,<sup>5</sup> and  $\alpha, \beta, \gamma \geq 0$  satisfy  $\alpha + \beta + \gamma = 1$ .

*Proof:* See for example [17, Thm. 2.7.19 combined with Exercise 2.9.14]. ■

We begin by defining information dimension and collecting some of its basic properties used in the remainder of the paper.

**Definition 1.** *Let  $X$  be a random vector. We define the lower and upper information dimension of  $X$  as*

$$\underline{d}(X) := \liminf_{k \rightarrow \infty} \frac{H(\langle X \rangle_k)}{\log k} \quad \text{and} \quad \bar{d}(X) := \limsup_{k \rightarrow \infty} \frac{H(\langle X \rangle_k)}{\log k}, \quad (7)$$

respectively.<sup>6</sup> If  $\underline{d}(X) = \bar{d}(X)$  (possibly  $= \infty$ ), then we say that the information dimension  $d(X)$  of  $X$  exists and we set  $d(X) := \underline{d}(X) = \bar{d}(X)$ .

At first sight (7) may suggest that information dimension depends on the specific way  $X$  is quantized, namely through application of the floor-operation on an entry-wise basis. However, the following equivalent definition shows that this is, in fact, not the case.

**Lemma 1.** *For a random vector  $X$  with distribution  $\mu$ , we have*

$$\underline{d}(X) = \liminf_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[\log \mu(B(X; \varepsilon))]}{\log \varepsilon} \quad \text{and} \quad \bar{d}(X) = \limsup_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[\log \mu(B(X; \varepsilon))]}{\log \varepsilon},$$

where  $B(x; \varepsilon) \subseteq \mathbb{R}^n$  denotes the ball with center  $x$  and radius  $\varepsilon$  with respect to an arbitrary norm on  $\mathbb{R}^n$ .

*Proof:* See [18, App. I]. A (slightly) different proof can be found in the Online Addendum [2, Sect. III-A]. ■

Further basic properties of information dimension we shall need are summarized next.

**Lemma 2** ([12],[3],[4]). *1) Let  $X$  be a random vector in  $\mathbb{R}^n$ . Then  $\underline{d}(X), \bar{d}(X) < \infty$  if and only if  $H(\lfloor X \rfloor) < \infty$ . Moreover, in this case we have*

$$0 \leq \underline{d}(X) \leq \bar{d}(X) \leq n. \quad (8)$$

*2) Let  $X_1, \dots, X_K$  be independent random vectors in  $\mathbb{R}^n$ , such that  $d(X_i)$  exists for  $i = 1, \dots, K$ . Then*

$$d \begin{pmatrix} X_1 \\ \vdots \\ X_K \end{pmatrix} = \sum_{i=1}^K d(X_i). \quad (9)$$

<sup>5</sup>“Absolutely continuous” is to be understood with respect to Lebesgue measure. A discrete distribution is supported on a countable set of points, whereas a singular distribution lives on a set of Lebesgue measure zero and, in addition, does not have any point masses. An example of a singular distribution on  $\mathbb{R}$  is the Cantor distribution [16, Ex. 1.2.4].

<sup>6</sup>Note that due to the scaling invariance of entropy, we have  $H(\langle X \rangle_k) = H(\lfloor kX \rfloor)$  for all integers  $k > 0$ . The motivation for using  $\langle X \rangle_k$  in Definition 1 is the pointwise convergence of  $\langle X \rangle_k$ , as  $k \rightarrow \infty$ , to the original random vector  $X$ .

3) Let  $X$  and  $Y$  be independent random vectors in  $\mathbb{R}^n$ . Then

$$\max\{\bar{d}(X), \bar{d}(Y)\} \leq \bar{d}(X+Y) \quad (10)$$

$$\leq \bar{d}(X) + \bar{d}(Y). \quad (11)$$

The inequality (10) also holds when  $\bar{d}(\cdot)$  is replaced by  $\underline{d}(\cdot)$  throughout.

4) Let  $X$  be a random vector in  $\mathbb{R}^n$ ,  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$  a map, and  $\|\cdot\|$  a norm on  $\mathbb{R}^n$ .

- If  $F$  is Lipschitz, i.e., there exists a constant  $C > 0$  such that  $\|F(x) - F(y)\| \leq C\|x - y\|$ , for all  $x, y \in \mathbb{R}^n$ , then

$$\underline{d}(F(X)) \leq \underline{d}(X) \quad \text{and} \quad \bar{d}(F(X)) \leq \bar{d}(X). \quad (12)$$

- If  $F$  is bi-Lipschitz, i.e., there exist constants  $c, C > 0$  such that  $c\|x - y\| \leq \|F(x) - F(y)\| \leq C\|x - y\|$ , for all  $x, y \in \mathbb{R}^n$ , then

$$\underline{d}(F(X)) = \underline{d}(X) \quad \text{and} \quad \bar{d}(F(X)) = \bar{d}(X). \quad (13)$$

*Proof:* The statement in 1) follows directly from [2, Lem. 4] with  $p = k$  and  $q = 1$ . Eq. (9) follows from  $H(X_1, \dots, X_K) = \sum_{i=1}^K H(X_i)$  for independent  $X_i$ . For the inequalities (10) and (11), we first note that  $|H(\langle X \rangle_k + \langle Y \rangle_k) - H(\langle X+Y \rangle_k)| \leq n \log 3$ , which follows from Lemma 9 (in Appendix B) taking into account that  $\| [k(X+Y)] - [kX] - [kY] \|_\infty \leq 1$  and that the minimum distance of the value set of  $\langle X \rangle_k + \langle Y \rangle_k$  is lower-bounded by  $1/k$  by definition of  $\langle \cdot \rangle_k$ . The inequality (10) then follows from

$$\begin{aligned} H(\langle X \rangle_k) &= H(\langle X \rangle_k + \langle Y \rangle_k | \langle Y \rangle_k) \\ &\leq H(\langle X \rangle_k + \langle Y \rangle_k) \\ &\leq H(\langle X+Y \rangle_k) + n \log 3, \end{aligned}$$

$$\begin{aligned} H(\langle Y \rangle_k) &= H(\langle X \rangle_k + \langle Y \rangle_k | \langle X \rangle_k) \\ &\leq H(\langle X \rangle_k + \langle Y \rangle_k) \\ &\leq H(\langle X+Y \rangle_k) + n \log 3. \end{aligned}$$

Similarly for (11), we note that the value set of  $\langle X+Y \rangle_k$  is lower-bounded by  $1/k$ , and we apply Lemma 9 to obtain

$$\begin{aligned} H(\langle X+Y \rangle_k) &\leq H(\langle X \rangle_k + \langle Y \rangle_k) + n \log 3 \\ &\leq H(\langle X \rangle_k) + H(\langle Y \rangle_k) + n \log 3, \end{aligned}$$

where the last inequality follows from  $H(V) \geq H(V) - H(V|U+V) = I(V; U+V) = H(U+V) - H(U+V|V) = H(U+V) - H(U)$  for discrete random vectors  $U, V$ . To prove (12) we use the equivalent characterization of information dimension from Lemma 1. Since  $F$  is continuous, it is measurable and we find that

$$B\left(x; \frac{\varepsilon}{C}\right) \subseteq F^{-1}(B(F(x); \varepsilon)), \quad (14)$$

which implies

$$\mu\left(B\left(x; \frac{\varepsilon}{C}\right)\right) \leq F_*\mu(B(F(x); \varepsilon)), \quad (15)$$

where  $\mu$  is the distribution of  $X$ . Dividing (15) by  $\log \varepsilon$  and taking  $\liminf_{\varepsilon \rightarrow 0}$  yields

$$\liminf_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[\log F_*\mu(B(F(X); \varepsilon))]}{\log \varepsilon} \leq \liminf_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[\log \mu(B(X; \varepsilon))]}{\log \varepsilon}. \quad (16)$$

As  $F_*\mu$  is the distribution of  $F(X)$ , this implies  $\underline{d}(F(X)) \leq \underline{d}(X)$ . If  $F$  is bi-Lipschitz, we additionally have

$$F^{-1}(B(F(x); \varepsilon)) \subseteq B\left(x; \frac{\varepsilon}{c}\right),$$

which together with (16) implies that

$$\liminf_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[\log F_*\mu(B(F(X); \varepsilon))]}{\log \varepsilon} = \liminf_{\varepsilon \rightarrow 0} \frac{\mathbb{E}[\log \mu(B(X; \varepsilon))]}{\log \varepsilon}, \quad (17)$$

and hence  $\underline{d}(F(X)) = \underline{d}(X)$ . The statements in (16) and (17) also hold for “lim sup” in place of “lim inf”. ■

**Remark 1.** 1) Since linear isomorphisms  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$  are bi-Lipschitz with constants  $c = \inf_{\|x\|=1} \|F(x)\|$  and  $C = \sup_{\|x\|=1} \|F(x)\|$ , an immediate consequence of (13) and (9) is the following: Suppose  $(v_1, \dots, v_n)$  are linearly independent vectors in some vector space and  $X_1, \dots, X_n$  are independent random variables with  $d(X_i) = 1$ , for all  $i$ . Then, we have

$$d(X_1 v_1 + \dots + X_n v_n) \stackrel{(13)}{=} d\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \stackrel{(9)}{=} \sum_{i=1}^n d(X_i) = n. \quad (18)$$

2) For independent random variables  $X, Y$  with  $d(X) = d(Y) = 1$ , it follows by (8) and (10) that  $d(X+Y)$  exists and

$$d(X+Y) = 1. \quad (19)$$

This fact will often be used in the remainder of the paper.

Information dimension is, in general, difficult to compute analytically. However, for distributions that are discrete-continuous mixtures, there is an explicit formula, which is a simple extension of [12, Thm. 3] to the vector case.

**Proposition 2.** Let  $X$  be a random vector in  $\mathbb{R}^n$  with  $H(\lfloor X \rfloor) < \infty$  and distribution  $\mu$  that admits a decomposition into an absolutely continuous part  $\mu_{ac}$  and a discrete part  $\mu_d$  according to  $\mu = \alpha\mu_{ac} + (1-\alpha)\mu_d$ , where  $\alpha \in [0, 1]$ . Then, we have

$$d(X) = n\alpha. \quad (20)$$

*Proof:* The proof follows closely that of the corresponding result for the scalar case reported in [12, Thm. 3], and is provided in the Online Addendum [2, Sect. III-B] for completeness. ■

Another class of distributions that is amenable to analytical expressions for information dimension and will turn out crucial in the proof of our main result, Theorem 3, is that of self-similar homogeneous distributions.

A self-similar distribution  $\mu$  is characterized as follows. Let  $\{F_1, \dots, F_m\}$  be a finite set of contractions on  $\mathbb{R}^n$  endowed with norm  $\|\cdot\|$ , i.e.,  $F_i: \mathbb{R}^n \rightarrow \mathbb{R}^n$  with  $\|F_i(x) - F_i(y)\| < \|x - y\|$  for all  $x, y \in \mathbb{R}^n$  with  $x \neq y$ . The set  $\{F_1, \dots, F_m\}$  is called an iterated function system. Given a probability vector  $(\rho_1, \dots, \rho_m)$  we let  $F$  be the random contraction drawn from  $\{F_1, \dots, F_m\}$  according to  $(\rho_1, \dots, \rho_m)$ , i.e.,  $F = F_i$  with

probability  $\rho_i$ . It turns out (see [19, Sect. 1: (2)]) that there is a unique distribution  $\mu$  on  $\mathbb{R}^n$ , which is invariant under the random contraction  $F$ , i.e.,  $F(X) \stackrel{D}{=} X$  if  $X$  has distribution  $\mu$  and is independent of  $F$ . This distribution satisfies

$$\mu = \sum_{i=1}^m \rho_i F_{i*} \mu \quad (21)$$

and is called the self-similar distribution associated with the pair  $(\{F_1, \dots, F_m\}, (\rho_1, \dots, \rho_m))$ . We refer the interested reader to [19] for general facts on self-similar distributions. Analytical expressions for the information dimension of self-similar distributions are available only under suitable ‘‘regularity’’ conditions, such as the open set condition [20].

**Definition 2.** An iterated function system  $\{F_1, \dots, F_m\}$  is said to satisfy the open set condition if there exists a nonempty bounded open set  $\mathcal{U} \subseteq \mathbb{R}^n$  such that  $\bigcup_{i=1}^m F_i(\mathcal{U}) \subseteq \mathcal{U}$  and  $F_i(\mathcal{U}) \cap F_j(\mathcal{U}) = \emptyset$  for all  $i \neq j$ .

In the special case where all  $F_i$  are of the form  $F_i(x) := rx + w_i$  for a positive scalar  $r < 1$  and some set of vectors  $\mathcal{W} := \{w_1, \dots, w_m\} \subseteq \mathbb{R}^n$ , the associated measure  $\mu$  in (21) is said to be a self-similar *homogeneous* distribution with similarity ratio  $r$ . In this case, we can give the following explicit expression for a random vector  $X$  with the distribution  $\mu$  in (21)

$$X := \sum_{i \geq 0} r^i W_i, \quad (22)$$

where  $\{W_i\}_{i \geq 0}$  is a set of i.i.d. copies of a random vector  $W$  drawn from the set  $\mathcal{W}$  according to the probability vector  $(\rho_1, \dots, \rho_m)$ . To see this simply note that with  $\widetilde{W}$  another copy of  $W$ , independent of  $\{W_i\}_{i \geq 0}$ , we have

$$F(X) = rX + \widetilde{W} = \sum_{i \geq 0} r^{i+1} W_i + \widetilde{W} \stackrel{D}{=} \sum_{i \geq 0} r^i W_i = X.$$

By uniqueness of  $\mu$  in (21) it follows that  $X$  must have distribution  $\mu$ . Random vectors of the form (22) will be used later in the paper to construct full DoF-achieving transmit vectors. This ingenious idea was first described in [3] for scalar input distributions. The information dimension of random vectors with self-similar homogeneous distribution satisfying the open set condition takes on a particularly simple form.

**Theorem 1.** Let  $X$  be defined by (22) with the underlying iterated function system satisfying the open set condition. Then, we have

$$d(X) = \frac{H(W)}{\log \frac{1}{r}}. \quad (23)$$

*Proof:* See [21, Thm. 2] and [22, Thm. 4.4]. ■

We will see in the proof of Theorem 3 that the open set condition can be satisfied by suitably restricting the contraction parameter  $r$  as a function of  $m(\mathcal{W})$  and  $M(\mathcal{W})$ .

#### IV. MAIN RESULTS

Our first result is an achievability statement and shows that information dimension plays a fundamental role in the characterization of the DoF of vector ICs.

**Theorem 2.** Let  $X_1, \dots, X_K$  be independent random vectors in  $\mathbb{R}^M$  such that  $H(\lfloor X_i \rfloor) < \infty$ , for all  $i$ , and such that all information dimension terms appearing in

$$\text{dof}(X_1, \dots, X_K; \mathbf{H}) := \sum_{i=1}^K \left[ d \left( \sum_{j=1}^K \mathbf{H}_{i,j} X_j \right) - d \left( \sum_{j \neq i}^K \mathbf{H}_{i,j} X_j \right) \right], \quad (24)$$

exist. Then, we have

$$\text{DoF}(\mathbf{H}) \geq \text{dof}(X_1, \dots, X_K; \mathbf{H}). \quad (25)$$

*Proof:* See Appendix B. ■

The existence of the information dimension terms in (24) is not guaranteed for general input distributions. In particular, existence of  $d(X_i)$ ,  $i = 1, \dots, K$ , does not necessarily imply existence of the information dimension terms in (24). However, when all input distributions are either discrete-continuous mixtures or self-similar homogeneous, then the distributions of  $\sum_{j=1}^K \mathbf{H}_{i,j} X_j$  and  $\sum_{j \neq i}^K \mathbf{H}_{i,j} X_j$  are of the same nature as the input distributions, which by Proposition 2 and Theorem 1 guarantees existence of the information dimension terms in (24).

Even though a single-letter characterization of the capacity region of the (vector) IC is not available in the literature, the next result shows that we can get a single-letter formula for the DoF, which holds for almost all channel matrices  $\mathbf{H}$ .

**Theorem 3 (DoF-formula).** For (Lebesgue) almost all channel matrices  $\mathbf{H}$  we have

$$\text{DoF}(\mathbf{H}) = \sup_{X_1, \dots, X_K} \text{dof}(X_1, \dots, X_K; \mathbf{H}), \quad (26)$$

where the supremum is taken over all independent  $X_1, \dots, X_K$  such that  $H(\lfloor X_i \rfloor) < \infty$ , for all  $i$ , and such that all information dimension terms appearing in (24) exist.<sup>7,8</sup>

*Proof:* See Appendix B. ■

This formula extends the DoF-formula for scalar ICs established by Wu et al. [3] to the case of vector ICs. The quantity  $\text{dof}(X_1, \dots, X_K; \mathbf{H})$  can be interpreted as the DoF achieved by the particular choice of transmit vectors  $X_1, \dots, X_K$ . The terms  $d \left( \sum_{j=1}^K \mathbf{H}_{i,j} X_j \right) - d \left( \sum_{j \neq i}^K \mathbf{H}_{i,j} X_j \right)$  in (24) can be understood as the difference of the information dimensions of the noise-free receive signal and the noise-free interference at user  $i$ . The DoF of a given channel are obtained by maximizing  $\text{dof}(\cdot)$  over all independent transmit vectors satisfying  $H(\lfloor X_i \rfloor) < \infty$ ,  $i = 1, \dots, K$ .<sup>9</sup>

We hasten to add that Theorem 3 applies to *almost all* channel matrices  $\mathbf{H}$  only, and that an explicit characterization of the set of these matrices is not available. For a given channel matrix, it is therefore not clear whether Theorem 3 applies. The

<sup>7</sup>It turns out (in the proof of Theorem 3) that it is not necessary to restrict the supremization to transmit vectors satisfying the power constraint (4).

<sup>8</sup>We point out that the statement in [1, Thm. 1] needs to be restricted to hold true only for almost all  $\mathbf{H}$ .

<sup>9</sup>The expression  $\text{dof}(X_1, \dots, X_K; \mathbf{H})$  is defined only if all information dimension terms in (24) exist. For brevity, we will not always mention this existence condition.

set of exceptions is, however, of Lebesgue measure zero. What we do know is that Theorem 3 applies to channel matrices  $\mathbf{H}$  that have all entries rational.

**Theorem 4.** *If all entries of  $\mathbf{H}$  are rational, then the DoF-formula (26) holds.*

*Proof:* See Appendix B. ■

For scalar ICs [23] provides an explicit condition on the channel matrix  $\mathbf{H}$  to admit  $K/2$  DoF. This condition is satisfied by almost all  $\mathbf{H}$  and the DoF of the corresponding class of channels are shown in [23] to be given by (26). In addition, for the scalar case, channel matrices that have all entries non-zero and rational admit strictly less than  $K/2$  DoF as shown in [9], and their exact DoF are characterized by Theorem 4 through (26). Thus, for the scalar case we have explicit characterizations of channels where (26) applies and where  $\text{DoF}(\mathbf{H}) = K/2$  or  $\text{DoF}(\mathbf{H}) < K/2$ . In [3] it is furthermore shown that Theorem 3 also applies in the scalar case if all entries of  $\mathbf{H}$  are algebraic numbers. These arguments can be extended to the vector case. For conciseness, we do, however, not pursue this extension here.

The channel model in (1) assumes Gaussian noise, as was done in [3]. The results in Theorems 2–4 extend, however, to a fairly general class of noise distributions, as shown next.

**Theorem 5.** *The results in Theorems 2–4 continue to hold if the Gaussian noise distribution in the channel model (1) is replaced by an absolutely continuous noise distribution satisfying  $h(W_1), \dots, h(W_K) > -\infty$  and  $H(\lfloor W_1 \rfloor), \dots, H(\lfloor W_K \rfloor) < \infty$ .*

*Proof:* See Appendix B. ■

Our Theorems 2–5 above extend [3, Thm. 4] from the scalar case to the vector case and to a broader class of noise distributions. Like the proof of [3, Thm. 4], the proof of our Theorem 3 builds on two results, whose extensions from the scalar to the vector case are provided next. We begin with the following multi-letter characterization of the vector IC sum-capacity.

**Proposition 3.** *The sum-capacity of the channel (1) is given by*

$$C_{\text{sum}}(\mathbf{H}; \text{snr}) = \lim_{N \rightarrow \infty} \frac{1}{N} \sup_{\mathbf{X}_1^N, \dots, \mathbf{X}_K^N} \sum_{i=1}^K I(\mathbf{X}_i^N; \mathbf{Y}_i^N), \quad (27)$$

where  $\mathbf{X}_i^N = (X_i^{(1)} \dots X_i^{(N)})$  and the supremum is taken over all independent  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  satisfying

$$\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \mathbb{E} \left[ \left( X_i^{(n)}[m] \right)^2 \right] \leq 1, \quad \text{for all } i. \quad (28)$$

*Proof:* The proof is a straightforward extension of Ahlswede's limiting characterization of the sum-capacity of the discrete memoryless IC [24, Sect. 2: Lem. 1] to continuous alphabet channels under an average power constraint. ■

For the scalar IC, Wu et al. [3] single-letterized (27) through an ingenious construction, which we extend to the vector case

in the proof of Theorem 3. We hasten to add that this extension is relatively straightforward.

The second result we shall need in the proof of Theorem 3 relates information dimension to the high-snr asymptotics of mutual information in additive noise vector channels.

**Theorem 6.** *Let  $X$  and  $W$  be independent random vectors in  $\mathbb{R}^n$  such that  $W$  has an absolutely continuous distribution with  $h(W) > -\infty$  and  $H(\lfloor W \rfloor) < \infty$ . Then*

$$\limsup_{\text{snr} \rightarrow \infty} \frac{I(X; \sqrt{\text{snr}}X + W)}{\frac{1}{2} \log \text{snr}} = \bar{d}(X). \quad (29)$$

*Proof:* See Appendix A. ■

**Remark 2.** *By Lemma 11 in Appendix B it follows that  $\mathbb{E}[W^T W] < \infty$  implies  $H(\lfloor W \rfloor) < \infty$ . Finite second moment of  $W$  and  $h(W) > -\infty$  are therefore sufficient for the conditions in Theorem 6 to be satisfied.*

For  $X$  and  $W$  scalar, (29) was proved in [4, Thm. 2.7] under the following assumptions:  $h(W) > -\infty$ ,  $\mathbb{E}[\log(1 + |W|)] < \infty$ , and  $\mathbb{E}[\log(1 + |X|)] < \infty$ . This set of assumptions is more restrictive than that in Theorem 6 as it includes a condition on  $X$  as well and, in addition,  $H(\lfloor W \rfloor) < \infty$  is slightly weaker than  $\mathbb{E}[\log(1 + |W|)] < \infty$ . In [13, Thm. 9], (29) was shown in the scalar case for arbitrary  $X$  and Gaussian  $W$ . Theorem 6 above applies to the vector case and has weaker assumptions than both [13, Thm. 9] and [4, Thm. 2.7] as it does not impose restrictions on  $X$  and also covers more general noise distributions  $W$ . The proof of Theorem 6, provided in Appendix A, proceeds by first showing the result for the case of uniform noise, similar to [4], and then uses the Ruzsa triangle inequality for differential entropy [5, Thm. 3.1] to generalize to noise  $W$  satisfying  $h(W) > -\infty$  and  $H(\lfloor W \rfloor) < \infty$ . Recognizing that the Ruzsa triangle inequality for differential entropy can be applied considerably simplifies the arguments in [4, Thm. 2.7] (for the scalar case), which are based on explicit manipulations of the densities involved.

**Remark 3.** *An alternative proof of Theorem 6 can be obtained as follows. As in the proof provided in Appendix A, one starts by employing the Ruzsa triangle inequality for differential entropy to establish that*

$$\begin{aligned} & |I(X; \sqrt{\text{snr}}X + W) - I(X; \sqrt{\text{snr}}X + Z)| \\ & \leq \max\{I(Z; Z - W), I(W; Z - W)\}, \end{aligned} \quad (30)$$

for independent random vectors  $X$ ,  $W$ , and  $Z$ , with  $H(\lfloor X \rfloor) < \infty$ ,  $Z$  standard Gaussian, and  $W$  of absolutely continuous distribution with  $h(W) > -\infty$  and  $H(\lfloor W \rfloor) < \infty$ . By [25, Thm. 6] it follows that  $h(Z - W) < \infty$ , which implies that the right-hand side (RHS) of (30) is finite, and thus dividing (30) by  $\frac{1}{2} \log \text{snr}$  and letting  $\text{snr} \rightarrow \infty$  yields zero. In particular, this means that the asymptote on the left-hand side (LHS) of (29) remains unchanged if general noise  $W$  (with  $h(W) > -\infty$  and  $H(\lfloor W \rfloor) < \infty$ ) is replaced by Gaussian noise  $Z$ . The proof is then completed through an extension of [13, Thm. 9] to the vector case. While this route would lead to a shorter proof, the proof presented in Appendix A is more direct, completely self-contained, and illuminates the

link between  $I(X; \sqrt{\text{snr}}X + W)$  for uniformly distributed  $W$  and information dimension through its alternative definition in Lemma 1.

We note that the conditions  $h(W) > -\infty$ ,  $H(\lfloor W \rfloor) < \infty$  in Theorem 6 do not imply each other. Informally speaking,  $h(W) > -\infty$  is a condition on the microscopic structure of the density of  $W$  whereas  $H(\lfloor W \rfloor) < \infty$  restricts the density in a global sense. The condition  $h(W) > -\infty$  cannot be relaxed. To see this, suppose that  $h(W) = -\infty$  and consider a random vector  $X$  in  $\mathbb{R}^n$  with absolutely continuous distribution, and such that  $h(X) > -\infty$  and  $H(\lfloor X \rfloor) < \infty$  (these assumptions are satisfied, e.g., for Gaussian  $X$ ). By Proposition 2, we have  $d(X) = n$ . On the other hand,  $I(\sqrt{\text{snr}}X + W; W) = h(\sqrt{\text{snr}}X + W) - h(\sqrt{\text{snr}}X) \geq 0$  and therefore  $h(\sqrt{\text{snr}}X + W) \geq \frac{1}{2} \log \text{snr} + h(X) > -\infty$  for all  $\text{snr} > 0$ , which implies

$$I(X; \sqrt{\text{snr}}X + W) = h(\sqrt{\text{snr}}X + W) - h(W) = \infty,$$

and thus  $\limsup_{\text{snr} \rightarrow \infty} \frac{I(X; \sqrt{\text{snr}}X + W)}{\frac{1}{2} \log \text{snr}} = \infty$ . This obviously violates (29).

## V. IMPLICATIONS OF THE DOF-FORMULA

We now show how the results in Theorems 2–4 can be used to develop insights into the fundamental limits of interference alignment. Specifically, the DoF-formula (26) allows us to treat scalar [8], [9] and vector interference alignment schemes [6], [7], [11] in a unified fashion, and constitutes a tool for evaluating new interference alignment schemes both for the scalar and the vector case.

We begin the discussion with an extension of [3, Thm. 5] to vector ICs, which states that for almost all channel matrices no more than  $K/2$  normalized DoF can be achieved.

**Proposition 4.** *Suppose that there is a permutation  $\sigma$  of  $\{1, \dots, K\}$  such that  $\sigma(i) \neq i$  and  $\det \mathbf{H}_{i, \sigma(i)} \neq 0$ , for all  $i$ . Then, we have*

$$\frac{\text{DoF}(\mathbf{H})}{M} \leq \frac{K}{2}. \quad (31)$$

*Proof:* Let  $R_1(\text{snr}_k), \dots, R_K(\text{snr}_k)$  be achievable rates for users  $1, \dots, K$ , respectively, and consider a sequence  $\text{snr}_k \rightarrow \infty$  such that

$$\lim_{k \rightarrow \infty} \frac{R_1(\text{snr}_k) + \dots + R_K(\text{snr}_k)}{\frac{1}{2} \log \text{snr}_k} = \text{DoF}(\mathbf{H}). \quad (32)$$

We set

$$d_i := \limsup_{k \rightarrow \infty} \frac{R_i(\text{snr}_k)}{\frac{1}{2} \log \text{snr}_k}, \quad i = 1, \dots, K. \quad (33)$$

A straightforward extension of the arguments in [6, Sect. IV-A] to the vector case yields that  $d_i + d_j \leq M$  if  $\det \mathbf{H}_{i,j} \neq 0$ .

We thus have

$$2\text{DoF}(\mathbf{H}) \leq 2 \sum_{i=1}^K d_i \quad (34)$$

$$= \sum_{i=1}^K d_i + \sum_{i=1}^K d_{\sigma(i)} \quad (35)$$

$$= \sum_{i=1}^K \underbrace{(d_i + d_{\sigma(i)})}_{\leq M, \text{ by assumption}} \quad (36)$$

$$\leq KM, \quad (37)$$

where (34) follows since the sum of the lim sups (33) is larger than or equal to the limit of the sum in (32). ■

**Remark 4.** *The conditions for Proposition 4 to hold are satisfied for almost all channel matrices  $\mathbf{H}$ . To see this first note that the determinant of a matrix is a polynomial in the entries of the matrix and the set of zeros of every non-trivial polynomial has Lebesgue measure zero. This allows us to conclude that for almost all  $\mathbf{H}$  we have  $\det \mathbf{H}_{i,j} \neq 0$ , for all  $i, j$ . Second, we can always find a permutation  $\sigma$  with  $\sigma(i) \neq i$ , for  $i = 1, \dots, K$ , as we assume (throughout the paper) that  $K \geq 2$ .*

For channel matrices  $\mathbf{H}$  that satisfy the conditions in Proposition 4, it follows from Theorem 2 that (26) holds if we can construct input distributions such that  $\text{dof}(X_1, \dots, X_K; \mathbf{H})/M = K/2$ .

### A. Singular input distributions

For scalar ICs it was observed in [3] that input distributions with a singular component play a central role in achieving the supremum in (26). The natural extension of this result to the vector case holds, albeit with vastly different consequences as we shall see below.

**Proposition 5.** *Suppose  $X_1, \dots, X_K$  are independent random vectors in  $\mathbb{R}^M$ , whose distributions are discrete-continuous mixtures<sup>10</sup> and are such that  $H(\lfloor X_i \rfloor) < \infty$ , for  $i = 1, \dots, K$ . If  $\det \mathbf{H}_{i,j} \neq 0$ , for all  $i, j$ , then*

$$\frac{\text{dof}(X_1, \dots, X_K; \mathbf{H})}{M} \leq 1. \quad (38)$$

*Proof:* The proof is inspired by the proof of the corresponding result for the scalar case in [3, Sect. III-C], and is detailed in the Online Addendum [2, Sect. III-C]. ■

The condition  $\det \mathbf{H}_{i,j} \neq 0$ , for all  $i, j$ , means that all transmit signals fully contribute to the signal at each receiver. As explained in Remark 4, this condition is satisfied for almost all  $\mathbf{H}$ . Proposition 5 has far-reaching consequences as it says that restricting the transmit vectors to have distributions that do not contain a singular component, we can achieve no more than one normalized DoF with single-letter transmit vectors,<sup>11</sup> irrespectively of the number of users and for almost

<sup>10</sup>This means that  $\gamma = 0$  in the decomposition (6).

<sup>11</sup>By Theorem 3 single-letter inputs are DoF-optimal for almost all channels. This, however, leaves open the possibility that on a set of channel matrices of (Lebesgue) measure zero multi-letter inputs are needed to achieve  $\text{DoF}(\mathbf{H})$ .

all channels. This shows that the quantity  $\text{dof}(X_1, \dots, X_K; \mathbf{H})$  is highly sensitive to the nature of the input distributions, a surprising fact first discovered for scalar ICs in [3]. The implications of this result in the scalar case and in the vector case are, however, vastly different. In the scalar case singular input distributions are difficult to realize, or, more specifically, to approximate, as uncountable subsets of  $\mathbb{R}$  with zero Lebesgue measure have a complicated structure. While Gaussian input distributions can be efficiently approximated by discrete constellations through shaping [26], corresponding techniques for the approximation of singular scalar input distributions do not seem to be available in the literature. In the case of vector ICs the input symbols are vector-valued and a striking new feature appears. Singularity of the transmit symbols (vectors) can be realized by simply choosing input distributions that are (continuously) supported on lower-dimensional subspaces of the ambient signal space. These subspaces are chosen to maximize  $\text{dof}(X_1, \dots, X_K; \mathbf{H})$ , which is precisely the idea underlying interference alignment à la Cadambe and Jafar [6], and Maddah-Ali et al. [7]. We illustrate this point by way of an example.

**Example 1.** Let  $K = 3$ ,  $M = 2$ , and

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 2 & 2 & 0 & 1 & 1 & 1 \\ 1 & 0 & 2 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Choose the transmit vectors as

$$X_1 := \tilde{X}_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad X_2 := \tilde{X}_2 \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad X_3 := \tilde{X}_3 \begin{pmatrix} 1 \\ 3 \end{pmatrix}, \quad (39)$$

where  $\tilde{X}_1, \tilde{X}_2$ , and  $\tilde{X}_3$  are independent random variables with absolutely continuous distributions and  $H(\lfloor \tilde{X}_i \rfloor) < \infty$ , for  $i = 1, 2, 3$ . Note that the vectors  $X_1, X_2, X_3$  are continuously distributed on lines in 2-dimensional space and hence have distributions that are supported on sets of Lebesgue measure zero; this renders the input distributions singular. Therefore  $\text{dof}(X_1, X_2, X_3; \mathbf{H})/2$  is not bound by (38), even though  $\det \mathbf{H}_{i,j} \neq 0$ , for  $i, j = 1, 2, 3$ . Indeed, a simple calculation reveals that

$$\begin{aligned} \text{dof}(X_1, X_2, X_3; \mathbf{H}) &= d\left(\tilde{X}_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + (\tilde{X}_2 + \tilde{X}_3) \begin{pmatrix} 1 \\ 3 \end{pmatrix}\right) - d\left((\tilde{X}_2 + \tilde{X}_3) \begin{pmatrix} 1 \\ 3 \end{pmatrix}\right) \\ &+ d\left(\tilde{X}_2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + (\tilde{X}_1 + \tilde{X}_3) \begin{pmatrix} 1 \\ 4 \end{pmatrix}\right) - d\left((\tilde{X}_1 + \tilde{X}_3) \begin{pmatrix} 1 \\ 4 \end{pmatrix}\right) \\ &+ d\left(\tilde{X}_3 \begin{pmatrix} 1 \\ 3 \end{pmatrix} + (\tilde{X}_1 + 2\tilde{X}_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) - d\left((\tilde{X}_1 + 2\tilde{X}_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) \\ &= 2 - 1 + 2 - 1 + 2 - 1 = 3, \end{aligned} \quad (40)$$

and hence  $\text{DoF}(\mathbf{H})/2 \geq 3/2$ . Here, we used (18) and (20) together with  $d(\tilde{X}_i + \alpha\tilde{X}_j) = 1$ , for  $i \neq j$  and  $\alpha = 1, 2$ , which follows from (19) as  $d(\alpha\tilde{X}_j) = d(\tilde{X}_j) = 1$  by (13). Since  $\det \mathbf{H}_{i,j} \neq 0$  for  $i, j = 1, 2, 3$ , Proposition 4 implies that  $\text{DoF}(\mathbf{H})/2 \leq 3/2$ , and thus the transmit vectors in

(39) achieve the supremum in (26). We emphasize that while the distributions of the transmit vectors in this example are singular, their marginals are not. What is more, the marginals have absolutely continuous distributions, as the  $\tilde{X}_i$  have absolutely continuous distributions. We could, e.g., take the  $\tilde{X}_i$  to be Gaussian which would render the marginals Gaussian.

The reader can readily verify that the example above is in the spirit of interference alignment as put forward in [6], [7]. The underlying idea is to choose the inputs continuously distributed in  $M/2$ -dimensional subspaces and such that at each receiver the interference aligns within an  $M/2$ -dimensional subspace, while the desired signal plus interference is continuously distributed in  $\mathbb{R}^M$ . We finally note that the transmission scheme discussed in this example only works for  $M$  even.

### B. Classical interference alignment and parallel ICs

Interference alignment as introduced by Cadambe and Jafar [6] was first applied to single-antenna ICs with channel coefficients that vary across channel uses. The signal model in [6] differs from the one considered in the present paper as we take the channel coefficients to be constant across (vector-)channel uses. We can, however, cast the signal model in [6] into the vector IC model used in this paper as follows. We take a finite block generated by the model in [6] and stack the corresponding input and output symbols into vectors whose dimension equals the blocklength. The submatrices  $\mathbf{H}_{i,j}$  of the resulting channel matrix  $\mathbf{H}$  are diagonal and  $\mathbf{H}$  is then assumed to be constant across (vector-)channel uses. We illustrate the procedure by way of an example.

**Example 2.** Consider a time-varying 3-user single-antenna IC whose channel matrix alternates between

$$\mathbf{H}[1] = \begin{pmatrix} 1 & 1 & -1 \\ -1 & 1 & 1 \\ 1 & -1 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{H}[2] = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{pmatrix}. \quad (42)$$

As  $\mathbf{H}[1] + \mathbf{H}[2] = 2\mathbf{I}_2$ , a simple transmit scheme where each data symbol is sent over two consecutive channel uses allows the receivers interference-free access to the data symbols by simply adding the corresponding two consecutive receive symbols. Since each data symbol is repeated once, this scheme achieves  $3/2$  normalized DoF. Stacking the I/O relation induced by (42) over blocks of length 2 results in a vector IC of dimension  $M = 2$  with channel matrix

$$\mathbf{H} = \begin{pmatrix} 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 & 0 & 1 \end{pmatrix}. \quad (43)$$

Reflecting that each data symbol is sent over two consecutive channel uses, the corresponding transmit vectors are given by

$$X_1 := \tilde{X}_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad X_2 := \tilde{X}_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad X_3 := \tilde{X}_3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad (44)$$



where  $\tilde{X}_1, \tilde{X}_2$ , and  $\tilde{X}_3$  are independent random variables with absolutely continuous distributions and  $H(\lfloor \tilde{X}_i \rfloor) < \infty$ , for  $i = 1, 2, 3$ . A calculation similar to the one in (40)–(41) shows that  $\text{dof}(X_1, X_2, X_3; \mathbf{H})/2 = 3/2$ , in accordance with what was found above.

The submatrices  $\mathbf{H}_{i,j}$  in (43) are, indeed, all diagonal. This represents the structure of a parallel IC. For general parallel ICs with  $\mathbf{H}_{i,j} = \text{diag}(h_{i,j}[1], \dots, h_{i,j}[M])$ ,  $i, j = 1, \dots, K$ , we introduce the notation

$$\mathbf{H}[m] := \begin{pmatrix} h_{1,1}[m] & \cdots & h_{1,K}[m] \\ \vdots & \ddots & \vdots \\ h_{K,1}[m] & \cdots & h_{K,K}[m] \end{pmatrix}, \quad m = 1, \dots, M. \quad (45)$$

Here,  $\mathbf{H}[m]$  is the interference matrix of the  $m$ -th (scalar) subchannel (as already encountered in (42)), which acts on  $(X_1[m] \dots X_K[m])^T$ , i.e., the vector consisting of the  $m$ -th entry of each transmit vector. We note that for parallel ICs the conditions on  $\mathbf{H}$  in Propositions 4 and 5 are satisfied if all subchannel matrices  $\mathbf{H}[m]$  are fully connected, i.e.,  $h_{i,j}[m] \neq 0$ , for  $i, j = 1, \dots, K$ ,  $m = 1, \dots, M$ , as this guarantees that

$$\det \mathbf{H}_{i,j} = \prod_{m=1}^M h_{i,j}[m] \neq 0, \quad i, j = 1, \dots, K. \quad (46)$$

Since parallel ICs describe the classical interference alignment setup à la [6], [7], it behooves us, in the next subsection, to particularize our main results to that case and to discuss the resulting implications. Besides recovering results known in the literature [6], [10], we obtain a number of new insights. We emphasize, however, that the general vector IC setting is still highly relevant as MIMO ICs, and general time-frequency-selective channels [27] usually lead to  $\mathbf{H}_{i,j}$ -matrices that are not diagonal.

### C. Parallel ICs

An interesting question in the context of parallel ICs is that of separability. Concretely, a parallel IC is said to be separable if its capacity region equals the (Minkowski) sum of the capacity regions of the individual subchannels. This means that input distributions that are independent across subchannels are optimal for separable parallel ICs. Specifically, if a separable parallel IC is obtained from a purely time-selective channel as in Example 2, then coding across time is not needed to achieve capacity. It was shown in [10] that non-separable parallel ICs do exist. This phenomenon stands in stark contrast to the (Gaussian) multiple access and the (Gaussian) broadcast channel, both of which are always separable [10].

Here, we shall deal with the weaker notion of separability in terms of DoF, i.e., the question of whether full DoF can be achieved by transmit vectors that are independent across subchannels. The non-separability result in [10] was proved by showing that the channel with  $K = 3$ ,  $M = 2$ , and the subchannel matrices in (42) is not separable. We discussed this very example in the previous subsection showing that, indeed, transmit vectors with identical (and hence dependent) entries (cf. (44)) achieve full DoF.

We turn to analyzing the question of separability in more generality. First, we show how the non-separability result in [10] (in terms of DoF) can be recovered through our DoF-formula.

**Proposition 6.** *For a parallel IC with channel matrix  $\mathbf{H}$  such that (26) is satisfied for  $\mathbf{H}[m]$ ,  $m = 1, \dots, M$ , we have*

$$\text{DoF}(\mathbf{H}) \geq \sum_{m=1}^M \text{DoF}(\mathbf{H}[m]). \quad (47)$$

There exist cases where the inequality in (47) is strict.

*Proof:* By Theorem 2 we have  $\text{DoF}(\mathbf{H}) \geq \sup_{X_1, \dots, X_K} \text{dof}(X_1, \dots, X_K; \mathbf{H})$  for all  $\mathbf{H}$ . The inequality in (47) follows by restricting the supremization in (26) to vectors  $X_i = (X_i[1] \dots X_i[M])^T$  which, in addition to satisfying the conditions in the supremization, have independent entries (across  $m = 1, \dots, M$ ) and for which all information dimension terms appearing in the expression

$$\begin{aligned} & \text{dof}(X_1[m], \dots, X_K[m]; \mathbf{H}[m]) \\ &= \sum_{i=1}^K \left[ d \left( \sum_{j=1}^K h_{i,j}[m] X_j[m] \right) - d \left( \sum_{j \neq i}^K h_{i,j}[m] X_j[m] \right) \right] \end{aligned} \quad (48)$$

exist, for  $m = 1, \dots, M$ . Evaluating  $\text{dof}(X_1, \dots, X_K; \mathbf{H})$  for this class of inputs yields

$$\text{dof}(X_1, \dots, X_K; \mathbf{H}) \quad (49)$$

$$= \sum_{i=1}^K \left[ d \begin{pmatrix} \sum_{j=1}^K h_{i,j}[1] X_j[1] \\ \vdots \\ \sum_{j=1}^K h_{i,j}[M] X_j[M] \end{pmatrix} \right] \quad (50)$$

$$- d \begin{pmatrix} \sum_{j \neq i}^K h_{i,j}[1] X_j[1] \\ \vdots \\ \sum_{j \neq i}^K h_{i,j}[M] X_j[M] \end{pmatrix} \right] \quad (51)$$

$$\stackrel{(9)}{=} \sum_{m=1}^M \sum_{i=1}^K \left[ d \begin{pmatrix} \sum_{j=1}^K h_{i,j}[m] X_j[m] \end{pmatrix} \right] \quad (52)$$

$$- d \begin{pmatrix} \sum_{j \neq i}^K h_{i,j}[m] X_j[m] \end{pmatrix} \right] \quad (53)$$

$$= \sum_{m=1}^M \text{dof}(X_1[m], \dots, X_K[m]; \mathbf{H}[m]). \quad (54)$$

Taking the supremum in (49)–(54) over the class of input distributions specified above proves (47).

A class of examples that renders the inequality in (47) strict, and is different from the example reported in [10], is obtained as follows. Take  $K = 3$  and  $M = 2$  with

$$\mathbf{H}[m] = \begin{pmatrix} 1 & 0 & 0 \\ 1 & \lambda[m] & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad m = 1, 2, \quad (55)$$

where  $\lambda[1]$  and  $\lambda[2]$  are nonzero, rational, and satisfy  $\lambda[1] \neq \lambda[2]$ . From [3, Thm. 11], we get  $\text{DoF}(\mathbf{H}[m]) < 3/2$ , for  $m = 1, 2$ . Now, consider the transmit vectors

$$X_1 := \tilde{X}_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad X_2 := \tilde{X}_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad X_3 := \tilde{X}_3 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad (56)$$

where  $\tilde{X}_1, \tilde{X}_2, \tilde{X}_3$  are independent random variables with  $d(\tilde{X}_i) = 1$ , for  $i = 1, 2, 3$ . Note that the transmit vectors  $X_1$  and  $X_2$  have dependent, in fact identical, entries, which implies that we perform joint coding across subchannels. Inserting into (24), we find

$$\begin{aligned} & \text{dof}(X_1, X_2, X_3; \mathbf{H}) \\ &= d\left(\tilde{X}_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) - d\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) \\ &+ d\left(\tilde{X}_2 \begin{pmatrix} \lambda[1] \\ \lambda[2] \end{pmatrix} + \tilde{X}_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) - d\left(\tilde{X}_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) \\ &+ d\left(\tilde{X}_3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (\tilde{X}_1 + \tilde{X}_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) - d\left((\tilde{X}_1 + \tilde{X}_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}\right) \\ &\stackrel{(18),(20)}{=} 1 - 0 + 2 - 1 + 2 - 1 = 3, \end{aligned} \quad (57)$$

where we used  $d(\tilde{X}_i + \tilde{X}_j) = 1$ , for  $i \neq j$ , cf. (19). Applying Theorem 2, we therefore find that

$$\frac{\text{DoF}(\mathbf{H})}{2} \stackrel{(57)}{\geq} \frac{3}{2} > \frac{\text{DoF}(\mathbf{H}[1]) + \text{DoF}(\mathbf{H}[2])}{2}. \quad (58)$$

The channel matrices in (55) constitute an entire family of parallel ICs that are non-separable. We next show that, however, almost all parallel ICs are separable (in terms of DoF). This means that for almost all parallel ICs, we can achieve full DoF without coding across subchannels or, in the setting of [6], across time or frequency.

**Proposition 7.** *For almost all parallel ICs, we have*

$$\text{DoF}(\mathbf{H}) = \sum_{m=1}^M \text{DoF}(\mathbf{H}[m]). \quad (59)$$

*Proof:* We will make repeated use of the fact that if two different properties hold individually for almost all channel matrices, then they also hold simultaneously for almost all channel matrices. We may assume that all subchannel matrices  $\mathbf{H}[m]$  are fully connected, since this holds for almost all parallel ICs. Then, based on (46) we can apply Proposition 4 and find together with Proposition 6 that

$$\frac{\text{DoF}(\mathbf{H}[1]) + \dots + \text{DoF}(\mathbf{H}[M])}{M} \leq \frac{\text{DoF}(\mathbf{H})}{M} \leq \frac{K}{2}, \quad (60)$$

for almost all parallel ICs. Since we know that  $\text{DoF}(\mathbf{H}[m]) = K/2$  for almost all  $\mathbf{H}[m]$  [8, Thm. 1], for  $m = 1, \dots, M$ , it follows that for almost all parallel ICs we have  $\text{DoF}(\mathbf{H}[m]) = K/2$ , for  $m = 1, \dots, M$ , as the cartesian product of sets of full (Lebesgue) measure (i.e., their complement has (Lebesgue) measure 0) again has full (Lebesgue) measure. From (60) we thus obtain that  $\text{DoF}(\mathbf{H})/M = K/2$  for almost all parallel ICs. ■

The RHS of (59) is achieved by transmit vectors with independent entries (across subchannels), where the individual

entries are taken to achieve  $\text{DoF}(\mathbf{H}[m])$ , for  $m = 1, \dots, M$ . This means that for almost all parallel ICs full DoF can be achieved without coding across subchannels. The downside is, however, that the distributions of the scalar entries of the corresponding full DoF-achieving transmit vectors must have a singular component as each subchannel constitutes a scalar IC. If we allow joint coding across subchannels, on the other hand, there are ICs where full DoF can be achieved even with transmit vectors whose entries have distributions that do not contain a singular component as the example class in the proof of Proposition 6 shows. Specifically, the input vectors in (56) have absolutely continuous scalar entries. The same interpretation applies to the example in (42) with the transmit vectors in (44).

We investigate the separability of parallel ICs and the nature of full DoF-achieving input distributions further by studying a parallel IC with  $K = 3$ ,  $M = 2$ , and fully connected subchannel matrices  $\mathbf{H}[1]$  and  $\mathbf{H}[2]$ . The fully connected scalar IC (i.e.,  $M = 1$ ) was studied in [8], where it was shown that  $\text{DoF}(\mathbf{H}) = K/2 (= 3/2$  in this example) for almost all channel matrices. In the following, we identify explicit conditions on  $\mathbf{H}[1]$  and  $\mathbf{H}[2]$  for  $\text{DoF}(\mathbf{H})/M = K/2 = 3/2$  as well as conditions for  $\text{DoF}(\mathbf{H})/M < 3/2$ . Note that Proposition 7 and the result in [8] imply that  $\text{DoF}(\mathbf{H})/M = 3/2$  for almost all parallel 3-user ICs with 2 subchannels. We begin by stating a result on the scaling invariance of  $\text{DoF}(\mathbf{H})$ , frequently used below.

**Lemma 3.** *Let  $\mathbf{D}_1 = \text{diag}(\mathbf{D}_{1,1}, \dots, \mathbf{D}_{1,K})$ ,  $\mathbf{D}_2 = \text{diag}(\mathbf{D}_{2,1}, \dots, \mathbf{D}_{2,K})$ , with  $\mathbf{D}_{i,j} \in \mathbb{R}^{M \times M}$  and  $\det \mathbf{D}_{i,j} \neq 0$ , for  $i = 1, 2$  and  $j = 1, \dots, K$ . Then, we have*

$$\text{DoF}(\mathbf{H}) = \text{DoF}(\mathbf{D}_1 \mathbf{H} \mathbf{D}_2).$$

*In particular,  $\text{DoF}(\mathbf{H})$  is invariant with respect to scaling of a row or a column by a nonzero constant.*

*Proof:* The proof follows the same line of arguments as the proof of the corresponding result for the scalar case [9, Lem. 1]. ■

Back to our example, we first note that, thanks to the scaling invariance in Lemma 3, it suffices to consider the “standard 3-user IC matrices” from [8, Sec. VI-B]

$$\mathbf{H}[m] = \begin{pmatrix} a[m] & 1 & 1 \\ 1 & b[m] & 1 \\ 1 & d[m] & c[m] \end{pmatrix}, \quad m = 1, 2, \quad (61)$$

where  $a[m]$ ,  $b[m]$ ,  $c[m]$ , and  $d[m]$  are nonzero for  $m = 1, 2$ . Since  $\mathbf{H}[1]$  and  $\mathbf{H}[2]$  are fully connected, we have  $\det \mathbf{H}_{i,j} = \prod_{m=1}^2 h_{i,j}[m] \neq 0$ , for all  $i, j$ , which, by Proposition 4, implies that

$$\text{DoF}(\mathbf{H})/2 \leq 3/2. \quad (62)$$

- 1) If, for  $m = 1, 2$ ,  $d[m]$  is rational and  $a[m], b[m], c[m]$  are all irrational, then by [3, Thm. 6] (26) is satisfied for  $\mathbf{H}[m]$  in place of  $\mathbf{H}$ , for  $m = 1, 2$ , and  $\text{DoF}(\mathbf{H}[1]) = \text{DoF}(\mathbf{H}[2]) = 3/2$ , which combined with Proposition 6 and (62) yields

$$\frac{\text{DoF}(\mathbf{H})}{2} = \frac{3}{2}. \quad (63)$$

Therefore,  $\text{DoF}(\mathbf{H}) = \text{DoF}(\mathbf{H}[1]) + \text{DoF}(\mathbf{H}[2])$ , which shows that full DoF are achieved without coding across subchannels.

- 2) If  $d[1] = d[2]$  and  $a[1] \neq a[2]$ ,  $b[1] \neq b[2]$ ,  $c[1] \neq c[2]$ , a calculation similar to the one in Example 1 (see [2, Sect. III-D]) reveals that  $\text{dof}(X_1, X_2, X_3; \mathbf{H}) = 3$  for independent transmit vectors each of which is continuously distributed along the direction  $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ . Together with (62), we hence get

$$\frac{\text{DoF}(\mathbf{H})}{2} = \frac{3}{2}. \quad (64)$$

This is achieved by joint coding across subchannels. Specifically, each transmit vector has scalar entries that are dependent, in fact identical, across subchannels, which corresponds to repetition coding. Note that, here, to achieve full DoF we do not need the distributions of the scalar entries of the transmit vectors to contain singular components.

- 3) If all entries of the subchannel matrices are rational and one of the entries on the main diagonal in (61) is identical across the two subchannels, then

$$\frac{\text{DoF}(\mathbf{H})}{2} < \frac{3}{2}, \quad (65)$$

i.e., we have strictly less than 3/2 normalized DoF. This follows from an extension of [3, Thm. 8] to the vector case, detailed in the Online Addendum [2, Sect. III-E].

To further demonstrate what we can get out of our DoF-formula, we next analyze an example that was studied in [6, Sect. III-C]. Specifically, we recover the statement that 4/3 normalized DoF can be achieved for almost all 3-user parallel ICs with 3 subchannels using joint coding across subchannels.

**Example 3.** We take  $K = 3$ ,  $M = 3$ , and again assume that the subchannel matrices  $\mathbf{H}[m]$  are in standard form as in (61). We further assume that the channel coefficients  $a[m], b[m], c[m], d[m]$ ,  $m = 1, 2, 3$ , are chosen randomly with respect to a joint (across  $a, b, c, d$ , and  $m$ ) absolutely continuous distribution. We take the transmit vectors as

$$X_1 := \tilde{X}_{1,1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \tilde{X}_{1,2} \begin{pmatrix} d[1] \\ d[2] \\ d[3] \end{pmatrix}, \quad X_2 := \tilde{X}_2 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

$$\text{and } X_3 := \tilde{X}_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad (66)$$

where  $\tilde{X}_{1,1}, \tilde{X}_{1,2}, \tilde{X}_2$ , and  $\tilde{X}_3$  are independent random variables with absolutely continuous distributions and  $H(\lfloor \tilde{X}_{1,1} \rfloor), H(\lfloor \tilde{X}_{1,2} \rfloor), H(\lfloor \tilde{X}_2 \rfloor), H(\lfloor \tilde{X}_3 \rfloor) < \infty$ . We then find (67)–(69) displayed at the top of the next page, where (68) holds almost surely with respect to the channel coefficients

$a[m], b[m], c[m], d[m]$ , as a consequence of each of the sets

$$\left\{ \begin{pmatrix} a[1] \\ a[2] \\ a[3] \end{pmatrix}, \begin{pmatrix} a[1]d[1] \\ a[2]d[2] \\ a[3]d[3] \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\},$$

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} d[1] \\ d[2] \\ d[3] \end{pmatrix}, \begin{pmatrix} b[1] \\ b[2] \\ b[3] \end{pmatrix} \right\}, \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} d[1] \\ d[2] \\ d[3] \end{pmatrix}, \begin{pmatrix} c[1] \\ c[2] \\ c[3] \end{pmatrix} \right\}$$

being linearly independent almost surely. In (69) we used (19). The transmit vectors in (66) therefore achieve 4/3 normalized DoF almost surely (with respect to the channel coefficients).

#### D. Verifying the MIMO interference alignment conditions in [28]

We now show how our general DoF-formula can be used to verify the interference alignment conditions for MIMO ICs with constant channel matrices analyzed in [28, Eqs. (3)–(4)]. For simplicity of exposition, we assume  $M$  antennas at each transmitter and receiver. The interference alignment conditions in [28, Eqs. (3)–(4)] are summarized as follows. To achieve  $\ell$  DoF, find pairs of subspaces  $(\mathcal{U}_i, \mathcal{V}_i)$ ,  $i = 1, \dots, K$  of  $\mathbb{R}^M$  with  $d_i := \dim \mathcal{U}_i = \dim \mathcal{V}_i$  and  $\sum_{i=1}^K d_i = \ell$ , such that

$$\mathbf{H}_{i,j} \mathcal{U}_j \subseteq \mathcal{V}_i^\perp, \quad \text{for all } i \neq j, \quad (71)$$

$$\dim(\pi_{\mathcal{V}_i}(\mathbf{H}_{i,i} \mathcal{U}_i)) = \dim(\mathcal{U}_i), \quad \text{for all } i = 1, \dots, K, \quad (72)$$

where  $\pi_{\mathcal{V}_i}$  denotes the orthogonal projection operator onto  $\mathcal{V}_i$ . These pairs of subspaces are associated with a transmit-receive scheme as follows. We choose  $X_i$  to be continuously distributed in the subspace  $\mathcal{U}_i$  (i.e., the distribution of  $X_i$  has a density supported on  $\mathcal{U}_i$ ) and such that  $H(\lfloor X_i \rfloor) < \infty$ . We furthermore take  $X_1, \dots, X_K$  to be independent. The  $i$ -th receiver computes the orthogonal projection onto  $\mathcal{V}_i$ . By (71) this results in interference-free signals at all receivers and (72) guarantees that in the process the desired signal at the  $i$ -th receiver does not experience dimensionality reduction.

Next, we show how our achievability result (25) can be used to verify that the transmit-receive scheme summarized above, indeed, achieves  $\ell$  DoF. To this end, we first organize the elements of a basis for  $\mathcal{V}_i$  into the first  $d_i$  columns of an  $M \times M$  matrix  $\mathbf{V}_i = (v_1, \dots, v_{d_i}, \tilde{v}_1, \dots, \tilde{v}_{M-d_i})$  with the remaining columns consisting of a basis for the  $(M - d_i)$ -dimensional orthogonal complement of  $\mathbf{H}_{i,i} \mathcal{U}_i$ . Note that by (72) multiplication of  $\mathcal{U}_i$  by  $\mathbf{H}_{i,i}$  must necessarily be dimensionality-preserving, i.e.,  $\mathbf{H}_{i,i} \mathcal{U}_i$  must be  $d_i$ -dimensional. We now show that  $\det \mathbf{V}_i \neq 0$ . To this end, suppose that  $v \in \mathbb{R}^M$  lies in the orthogonal complement of the column space of  $\mathbf{V}_i$ . Then,  $v$  has to be orthogonal to both  $\mathcal{V}_i$  and the orthogonal complement of  $\mathbf{H}_{i,i} \mathcal{U}_i$ , and it follows that  $\pi_{\mathcal{V}_i}(v) = 0$  and  $v \in \mathbf{H}_{i,i} \mathcal{U}_i$ . But by (72) this can only hold for  $v = 0$  as the application of  $\pi_{\mathcal{V}_i}$  to the subspace  $\mathbf{H}_{i,i} \mathcal{U}_i$  can not result in dimensionality reduction. The orthogonal complement of the column space of

$$\begin{aligned}
\text{dof}(X_1, X_2, X_3; \mathbf{H}) &= d\left(\tilde{X}_{1,1} \begin{pmatrix} a[1] \\ a[2] \\ a[3] \end{pmatrix} + \tilde{X}_{1,2} \begin{pmatrix} a[1]d[1] \\ a[2]d[2] \\ a[3]d[3] \end{pmatrix} + (\tilde{X}_2 + \tilde{X}_3) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right) - d\left((\tilde{X}_2 + \tilde{X}_3) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right) \\
&+ d\left((\tilde{X}_{1,1} + \tilde{X}_3) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \tilde{X}_{1,2} \begin{pmatrix} d[1] \\ d[2] \\ d[3] \end{pmatrix} + \tilde{X}_2 \begin{pmatrix} b[1] \\ b[2] \\ b[3] \end{pmatrix}\right) - d\left((\tilde{X}_{1,1} + \tilde{X}_3) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \tilde{X}_{1,2} \begin{pmatrix} d[1] \\ d[2] \\ d[3] \end{pmatrix}\right) \\
&+ d\left(\tilde{X}_{1,1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (\tilde{X}_{1,2} + \tilde{X}_2) \begin{pmatrix} d[1] \\ d[2] \\ d[3] \end{pmatrix} + \tilde{X}_3 \begin{pmatrix} c[1] \\ c[2] \\ c[3] \end{pmatrix}\right) - d\left(\tilde{X}_{1,1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + (\tilde{X}_{1,2} + \tilde{X}_2) \begin{pmatrix} d[1] \\ d[2] \\ d[3] \end{pmatrix}\right)
\end{aligned} \tag{67}$$

$$\begin{aligned}
&\stackrel{(13)}{=} d\left(\begin{pmatrix} \tilde{X}_{1,1} \\ \tilde{X}_{1,2} \\ \tilde{X}_2 + \tilde{X}_3 \end{pmatrix}\right) - d(\tilde{X}_2 + \tilde{X}_3) + d\left(\begin{pmatrix} \tilde{X}_{1,1} + \tilde{X}_3 \\ \tilde{X}_{1,2} \\ \tilde{X}_2 \end{pmatrix}\right) - d\left(\begin{pmatrix} \tilde{X}_{1,1} + \tilde{X}_3 \\ \tilde{X}_{1,2} \end{pmatrix}\right) \\
&+ d\left(\begin{pmatrix} \tilde{X}_{1,1} \\ \tilde{X}_{1,2} + \tilde{X}_2 \\ \tilde{X}_3 \end{pmatrix}\right) - d\left(\begin{pmatrix} \tilde{X}_{1,1} \\ \tilde{X}_{1,2} + \tilde{X}_2 \end{pmatrix}\right)
\end{aligned} \tag{68}$$

$$\stackrel{(9)}{=} 3 - 1 + 3 - 2 + 3 - 2 = 4, \tag{69}$$

$\mathbf{V}_i$  is therefore trivial, and we can conclude that  $\det \mathbf{V}_i \neq 0$  for all  $i = 1, \dots, K$ . This yields

$$\begin{aligned}
&\text{dof}(X_1, \dots, X_K; \mathbf{H}) \\
&= \sum_{i=1}^K \left[ d\left(\sum_{j=1}^K \mathbf{H}_{i,j} X_j\right) - d\left(\sum_{j \neq i}^K \mathbf{H}_{i,j} X_j\right) \right]
\end{aligned} \tag{73}$$

$$\stackrel{(13)}{=} \sum_{i=1}^K \left[ d\left(\sum_{j=1}^K \mathbf{V}_i^T \mathbf{H}_{i,j} X_j\right) - d\left(\sum_{j \neq i}^K \mathbf{V}_i^T \mathbf{H}_{i,j} X_j\right) \right] \tag{74}$$

$$\begin{aligned}
&\stackrel{(71)}{=} \sum_{i=1}^K \left[ d\left(\begin{pmatrix} v_1^T \mathbf{H}_{i,i} X_i \\ \vdots \\ v_{d_i}^T \mathbf{H}_{i,i} X_i \\ \tilde{v}_1^T \sum_{j \neq i}^K \mathbf{H}_{i,j} X_j \\ \vdots \\ \tilde{v}_{M-d_i}^T \sum_{j \neq i}^K \mathbf{H}_{i,j} X_j \end{pmatrix}\right) \right. \\
&\quad \left. - d\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \\ \tilde{v}_1^T \sum_{j \neq i}^K \mathbf{H}_{i,j} X_j \\ \vdots \\ \tilde{v}_{M-d_i}^T \sum_{j \neq i}^K \mathbf{H}_{i,j} X_j \end{pmatrix}\right) \right]
\end{aligned} \tag{75}$$

$$\stackrel{(9)}{=} \sum_{i=1}^K d\left(\begin{pmatrix} v_1^T \mathbf{H}_{i,i} X_i \\ \vdots \\ v_{d_i}^T \mathbf{H}_{i,i} X_i \end{pmatrix}\right) \tag{76}$$

$$= \sum_{i=1}^K d(\mathbf{V}_i^T \mathbf{H}_{i,i} X_i) \tag{77}$$

$$\stackrel{(13)}{=} \sum_{i=1}^K d_i \tag{78}$$

$$= \ell, \tag{79}$$

where in the application of (13) in (74) we used the fact that the map  $\mathbf{V}_i$  is bi-Lipschitz as a consequence of  $\det \mathbf{V}_i \neq 0$ . Furthermore, in the application of (13) in (78) we used that  $d(X_i) = d_i$  by Proposition 2 together with the fact that  $\mathbf{V}_i^T \mathbf{H}_{i,i}$  induces a bi-Lipschitz map on  $\mathcal{U}_i$  since  $\det \mathbf{V}_i \neq 0$  and  $\mathbf{H}_{i,i}$  acts isomorphically on  $\mathcal{U}_i$  by (72).

Conditions on the existence of pairs of subspaces that satisfy (71) and (72) can be expressed in terms of algebraic equations, which can be studied using methods from algebraic geometry, see [28], [11]. In particular, it is shown in [28, Corollary 8] that if  $d_1 = \dots = d_K$ , then for almost all MIMO ICs no more than  $\ell/M = 2$  normalized DoF can be achieved with the interference alignment scheme described in this subsection.

### E. The complex case

The results stated so far apply to real signals and channel matrices. We next describe how Theorems 2 and 3 can be extended to the complex case, thereby also providing an extension of the main result in [3] to the complex case.

For a vector IC with transmit and receive signals in  $\mathbb{C}^M$  and  $\mathbf{H}_{i,j} \in \mathbb{C}^{M \times M}$ , for  $i, j = 1, \dots, K$ , simply stack the real and imaginary parts of the transmit, receive, and noise vectors in (1), and stack the real and imaginary parts of the matrices  $\mathbf{H}_{i,j}$  correspondingly according to

$$\begin{pmatrix} \text{Re}(\mathbf{H}_{i,j}) & -\text{Im}(\mathbf{H}_{i,j}) \\ \text{Im}(\mathbf{H}_{i,j}) & \text{Re}(\mathbf{H}_{i,j}) \end{pmatrix}. \tag{80}$$

Then apply Theorem 2 to the resulting vector IC in  $\mathbb{R}^{2M}$  and divide the so obtained number of DoF by 2 to account for the fact that we would like to think in terms of complex dimensions (degrees of freedom) for the complex IC. As the scalar complex IC is turned into a 2-dimensional real vector IC, we see that having a characterization of the DoF for vector ICs constitutes the basis for extending the main result in [3] to the complex case. The DoF-formula in Theorem 3 can analogously

be shown to hold for the complex case. The achievability part of the corresponding proof is straightforward, the converse requires certain modifications taking into account the specific structure of the effective channel matrix in (80). For brevity, we do not present the proof of this extension.

The extensions of Theorems 2 and 3 to the complex case allow us to treat interference alignment schemes for complex channel matrices  $\mathbf{H}$ . This is illustrated next by analyzing the following well-known example using Theorem 2.

**Example 4** ([6]). *We consider a cyclic variant of the example in [6, App. 1], whose main idea is to exploit channel-induced differences in propagation delays to perform interference alignment. Let  $M$  be even, take  $K \geq 3$ , and consider the complex channel with I/O relation*

$$Y_k[m] = \sqrt{\text{snr}} \left( X_k[m] + \sum_{\ell \neq k}^K X_\ell[m-1] \right) + W_k[m], \quad (81)$$

for  $k = 1, \dots, K$ ,  $m = 1, \dots, M$ , where the argument in square brackets, e.g., in  $X_k[m]$ , is to be understood modulo  $M$ . The corresponding channel matrices are given by  $\mathbf{H}_{k,k} = \mathbf{I}_M$ , and

$$\mathbf{H}_{k,\ell} = \begin{pmatrix} 0 & \cdots & \cdots & 0 & 1 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \quad k \neq \ell, \quad k, \ell = 1, \dots, K.$$

Denoting the  $m$ -th vector of the standard basis in  $\mathbb{R}^M$  by  $e_m$ , we choose the transmit vectors  $X_k$  to be continuously distributed in the subspace spanned by  $\{e_m \mid 1 \leq m \leq M, m \text{ even}\}$  such that  $H(\lfloor X_k \rfloor) < \infty$ . This accounts for the fact that the transmitters remain silent in odd time slots. The vectors  $\mathbf{H}_{k,\ell} X_\ell$ ,  $k \neq \ell$ , are then continuously distributed in the subspace spanned by  $\{e_m \mid 1 \leq m \leq M, m \text{ odd}\}$ , which reflects the fact that interference at each receiver is confined to odd time slots. Inserting into (24), we find that

$$\begin{aligned} \text{dof}(X_1, \dots, X_K; \mathbf{H}) &= \sum_{k=1}^K \left[ d\left(\sum_{\ell=1}^K \mathbf{H}_{k,\ell} X_\ell\right) - d\left(\sum_{\ell \neq k}^K \mathbf{H}_{k,\ell} X_\ell\right) \right] \\ &= \sum_{k=1}^K \left[ M - \frac{M}{2} \right] \\ &= KM/2, \end{aligned}$$

where we used  $d\left(\sum_{\ell=1}^K \mathbf{H}_{k,\ell} X_\ell\right) = d\left(\mathbf{H}_{k,k} X_k + \sum_{\ell \neq k}^K \mathbf{H}_{k,\ell} X_\ell\right) = M$  and  $d\left(\sum_{\ell \neq k}^K \mathbf{H}_{k,\ell} X_\ell\right) = M/2$ , both thanks to (18). Moreover, since  $\det \mathbf{H}_{k,\ell} \neq 0$ , for all  $k, \ell$ , we can apply<sup>12</sup> Proposition 4 to find that  $\text{DoF}(\mathbf{H})/M \leq K/2$  and therefore

$$\frac{\text{DoF}(\mathbf{H})}{M} = \frac{K}{2}. \quad (82)$$

<sup>12</sup>In fact, we apply the extension of Proposition 4—obtained mutatis mutandis—to complex ICs.

This recovers the result in [6, App. 1], showing that the scheme above achieves full DoF.

The example can equivalently be analyzed in the frequency domain. We pursue this in the following as it reveals interesting properties of the full DoF-achieving transmit vectors in the frequency domain. First, we note that the cyclic channel matrices  $\mathbf{H}_{k,\ell}$  are simultaneously diagonalized by the discrete Fourier transform (DFT). With  $\hat{X}[m]$ ,  $\hat{Y}[m]$ , and  $\hat{W}[m]$  denoting the  $M$ -point DFT of the transmit, receive, and noise vectors, respectively, we obtain from (81) the following transformed I/O relation

$$\hat{Y}_k[m] = \sqrt{\text{snr}} \left( \hat{X}_k[m] + e^{-2\pi i \frac{m}{M}} \sum_{\ell \neq k}^K \hat{X}_\ell[m] \right) + \hat{W}_k[m], \quad (83)$$

for  $m = 1, \dots, M$  and  $k = 1, \dots, K$ . This is a parallel IC with subchannel matrices

$$\hat{\mathbf{H}}[m] = \begin{pmatrix} 1 & e^{-2\pi i \frac{m}{M}} & \cdots & e^{-2\pi i \frac{m}{M}} \\ e^{-2\pi i \frac{m}{M}} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & e^{-2\pi i \frac{m}{M}} \\ e^{-2\pi i \frac{m}{M}} & \cdots & e^{-2\pi i \frac{m}{M}} & 1 \end{pmatrix}. \quad (84)$$

As the DFT applied at each transmitter and receiver is unitary, (83) constitutes an equivalent channel, i.e., every code for the original channel corresponds to a code for the modified channel (83) that achieves the same rate. Moreover, since the DFT is invertible the transformed noise distributions meet the conditions in Theorem 5 if  $W_1, \dots, W_K$  in (81) do. Indeed, since the DFT is unitary we have  $h(\hat{W}_k) = h(W_k) > -\infty$ , for  $k = 1, \dots, K$ . Moreover, by the first property in Lemma 2  $H(\lfloor W_k \rfloor) < \infty$  is equivalent to  $\underline{d}(W_k), \bar{d}(W_k) < \infty$ , and we have  $\underline{d}(W_k) = \underline{d}(\hat{W}_k)$ ,  $\bar{d}(W_k) = \bar{d}(\hat{W}_k)$  by (13), which, since  $\underline{d}(W_k), \bar{d}(W_k) < \infty$  by assumption, again using the first property in Lemma 2, establishes that  $H(\lfloor \hat{W}_k \rfloor) < \infty$ , for  $k = 1, \dots, K$ . We next analyze the parallel IC in the frequency domain with respect to its separability. The matrices  $\hat{\mathbf{H}}[m]$  are fully connected; we can therefore (cf. (46)) apply Proposition 4 to the individual subchannels and find that

$$\text{DoF}(\hat{\mathbf{H}}[m]) \leq \frac{K}{2}, \quad m = 1, \dots, M. \quad (85)$$

For  $m = \frac{M}{2}$  and  $m = M$  the channel matrix  $\hat{\mathbf{H}}[m]$  has real, in fact rational, entries. We can therefore apply [9, Thm. 2]<sup>13</sup> to conclude that the inequality (85) is strict for  $m = \frac{M}{2}$  and  $m = M$ . This implies that the inequality (47) is strict, which means that the channel (83) is a non-separable parallel IC

<sup>13</sup>Although [9, Thm. 2] is stated for real scalar ICs, it also applies to scalar ICs with real channel matrix and complex inputs and outputs, as is the case here. To see this, simply stack the real and imaginary parts of input and output to obtain a parallel IC with 2 subchannels and identical subchannel matrices. Since the overall channel matrix has rational entries only, we know by Theorem 4 that single-letter inputs are DoF-optimal. Therefore transmit vectors with independent entries achieving full DoF of the real scalar IC are DoF-optimal for the induced parallel IC and we have equality in (47). This implies that the DoF of the channel with complex inputs and outputs equal twice the DoF of the real scalar channel. By [9, Thm. 2] the DoF of each of the real scalar ICs are strictly less than  $K/2$ , and so the same holds true for the normalized DoF of the IC with complex inputs and outputs.

and joint coding across subchannels is mandatory to achieve full DoF. We next show how the transmit vectors we chose in the time domain amount to coding across subchannels in the frequency domain. Noting that

$$\widehat{X}_k \left[ m + \frac{M}{2} \right] = \frac{1}{\sqrt{M}} \sum_{m' \text{ even}}^M X_k[m'] e^{-2\pi i \frac{m m'}{M}} \underbrace{e^{-\pi i m'}}_{=(-1)^{m'}} \quad (86)$$

$$= \widehat{X}_k[m], \quad (87)$$

for all  $m$  and  $k$ , it follows that each user sends the same symbol over two different frequency slots, implementing repetition coding across subchannels. The receiver computes  $\widehat{Y}_k[m] + \widehat{Y}_k[m + \frac{M}{2}] = 2\sqrt{\text{snr}}\widehat{X}_k[m] + \widehat{W}_k[m] + \widehat{W}_k[m + \frac{M}{2}]$  to recover the transmitted data symbols  $\widehat{X}_k[m]$ . We see that interference is canceled by the phase difference induced by the channel across the frequency slots  $m$  and  $m + M/2$ , i.e., for  $\ell \neq k$ , we have

$$\begin{aligned} & e^{-2\pi i \frac{m}{M}} \widehat{X}_\ell[m] + e^{-2\pi i \frac{m + \frac{M}{2}}{M}} \widehat{X}_\ell \left[ m + \frac{M}{2} \right] \\ &= e^{-2\pi i \frac{m}{M}} \left( \widehat{X}_\ell[m] - \widehat{X}_\ell \left[ m + \frac{M}{2} \right] \right) \\ &\stackrel{(87)}{=} 0. \end{aligned}$$

## VI. OPEN PROBLEMS

It would be desirable to prove that the DoF-formula (26) holds for *all* rather than only *almost all* channel matrices. This appears to be a very hard problem, whose solution may require novel results in number theory. We believe that the main obstacle in extending (26) to all channel matrices resides in the approximation of the entropy of the noiseless output signals by the entropy of their quantized versions, see (213)–(227). These approximations depend crucially on results from Diophantine approximation theory, which guarantee an approximation behavior as in [3, (112)], and are known to hold for almost all  $\mathbf{H}$  including special cases such as  $\mathbf{H}$  with all entries rational or algebraic numbers.

Despite the fact that our DoF-formula holds for almost all  $\mathbf{H}$  only, it provides a comprehensive framework for studying the DoF in vector ICs. Knowing that the DoF-formula holds for *all* channel matrices could be useful in finding sharper upper bounds on the DoF of ICs that have strictly less than  $K/2$  DoF.

## ACKNOWLEDGMENTS

The authors would like to thank Y. Wu, S. Shamai (Shitz), and S. Verdú for interesting discussions and A. Guionnet for advice on the proof of [4, Sect. 2]. We gratefully acknowledge an anonymous reviewer for pointing out the alternative proof of Theorem 6 explained in Remark 3.

## APPENDIX A PROOF OF THEOREM 6

Since by Lemma 2,  $\bar{d}(X) < \infty$  if and only if  $H(\lfloor X \rfloor) < \infty$ , we distinguish the two cases  $H(\lfloor X \rfloor) = \infty$  and  $H(\lfloor X \rfloor) < \infty$ . First, we show that  $H(\lfloor X \rfloor) = \infty$  implies

$$\limsup_{\text{snr} \rightarrow \infty} \frac{I(X; \sqrt{\text{snr}}X + W)}{\frac{1}{2} \log \text{snr}} = \infty.$$

For  $\text{snr} > 0$ , we have

$$I(X; \sqrt{\text{snr}}X + W) \geq I(\lfloor X \rfloor; \sqrt{\text{snr}}X + W) \quad (88)$$

$$\geq I(\lfloor X \rfloor; \lfloor \sqrt{\text{snr}}X + W \rfloor) \quad (89)$$

$$= H(\lfloor X \rfloor) - H(\lfloor X \rfloor | \lfloor \sqrt{\text{snr}}X + W \rfloor), \quad (90)$$

where (88) and (89) are by the data processing inequality applied to  $\lfloor X \rfloor - X - \sqrt{\text{snr}}X + W$  and  $\lfloor X \rfloor - \sqrt{\text{snr}}X + W - \lfloor \sqrt{\text{snr}}X + W \rfloor$ , respectively. Now

$$H(\lfloor X \rfloor | \lfloor \sqrt{\text{snr}}X + W \rfloor) \quad (91)$$

$$\leq H(\lfloor X \rfloor, \lfloor W \rfloor | \lfloor \sqrt{\text{snr}}X + W \rfloor) \quad (92)$$

$$= H(\lfloor W \rfloor | \lfloor \sqrt{\text{snr}}X + W \rfloor) + H(\lfloor X \rfloor | \lfloor \sqrt{\text{snr}}X + W \rfloor, \lfloor W \rfloor) \quad (93)$$

$$\leq H(\lfloor W \rfloor) + H(\lfloor X \rfloor | \lfloor \sqrt{\text{snr}}X + W \rfloor, \lfloor W \rfloor) \quad (94)$$

$$< \infty, \quad (95)$$

where (92) is by the chain rule and the non-negativity of entropy, (93) is again by the chain rule, and (95) holds since by assumption  $H(\lfloor W \rfloor) < \infty$  and since the second term in (94) is finite, which is seen as follows. We show that given  $\lfloor \sqrt{\text{snr}}X + W \rfloor$  and  $\lfloor W \rfloor$  the number of possible values of  $\lfloor X \rfloor$  is bounded independently of  $\text{snr}$  when  $\text{snr}$  is sufficiently large. First, we note that  $\lfloor \sqrt{\text{snr}}X \rfloor$  has integer-valued entries which are sandwiched according to

$$\lfloor \sqrt{\text{snr}}X_i + W_i \rfloor - \lfloor W_i \rfloor - 2 < \sqrt{\text{snr}}X_i + W_i - W_i - 1 \quad (96)$$

$$< \lfloor \sqrt{\text{snr}}X_i \rfloor \quad (97)$$

$$\leq \sqrt{\text{snr}}X_i + W_i - W_i \quad (98)$$

$$< \lfloor \sqrt{\text{snr}}X_i + W_i \rfloor + 1 - \lfloor W_i \rfloor. \quad (99)$$

Here  $X_i$  and  $W_i$  denote the  $i$ -th entry of  $X$  and  $W$ , respectively. Second,  $\lfloor X \rfloor$  can be sandwiched according to

$$\frac{1}{\sqrt{\text{snr}}} \lfloor \sqrt{\text{snr}}X_i \rfloor - 1 \leq X_i - 1 \quad (100)$$

$$< \lfloor X_i \rfloor \quad (101)$$

$$\leq \frac{1}{\sqrt{\text{snr}}} (\lfloor \sqrt{\text{snr}}X_i \rfloor + 1). \quad (102)$$

Combining (96)–(99) with (102) we obtain

$$\frac{1}{\sqrt{\text{snr}}} (\lfloor \sqrt{\text{snr}}X_i + W_i \rfloor - \lfloor W_i \rfloor - 2) - 1 \quad (103)$$

$$< \lfloor X_i \rfloor \quad (104)$$

$$< \frac{1}{\sqrt{\text{snr}}} (\lfloor \sqrt{\text{snr}}X_i + W_i \rfloor + 2 - \lfloor W_i \rfloor). \quad (105)$$

From (103)–(105) we conclude that the number of values  $\lfloor X \rfloor$  can take on given  $\lfloor \sqrt{\text{snr}}X + W \rfloor$  and  $\lfloor W \rfloor$  is bounded by  $(4/\sqrt{\text{snr}} + 1)^n$ . If we take  $\text{snr}$  sufficiently large, e.g.,  $\text{snr} > 1$ , we obtain a bound that is independent of  $\text{snr}$ , which concludes the argument. Finally, using (90) and (95) yields  $I(X; \sqrt{\text{snr}}X + W) = \infty$  for  $H(\lfloor X \rfloor) = \infty$  and hence

$$\limsup_{\text{snr} \rightarrow \infty} \frac{I(X; \sqrt{\text{snr}}X + W)}{\frac{1}{2} \log \text{snr}} = \infty, \quad (106)$$

as was to be shown.

It remains to establish the result for  $H(\lfloor X \rfloor) < \infty$ . We begin by decomposing

$$I(X; \sqrt{\text{snr}}X + W) = h(\sqrt{\text{snr}}X + W) - h(W),$$

which holds provided that  $h(\sqrt{\text{snr}}X + W)$  exists. Note that  $h(W)$  exists and satisfies  $h(W) > -\infty$  by assumption. The existence of  $h(\sqrt{\text{snr}}X + W)$  is established as part of Proposition 8 below. Setting  $t := \frac{1}{\sqrt{\text{snr}}}$  and using the scaling property for differential entropy, we have

$$h(\sqrt{\text{snr}}X + W) = h(X + tW) - n \log t. \quad (107)$$

To finalize the proof of Theorem 6 it therefore suffices to establish the following:

**Theorem 7.** *Let  $X$  and  $W$  be independent random vectors in  $\mathbb{R}^n$  such that  $H(\lfloor X \rfloor) < \infty$  and  $W$  has an absolutely continuous distribution with  $h(W) > -\infty$  and  $H(\lfloor W \rfloor) < \infty$ . Then, we have*

$$\limsup_{t \rightarrow 0} \frac{h(X + tW)}{|\log t|} = \bar{d}(X) - n.$$

The proof of Theorem 7 proceeds as follows. We first state a few auxiliary results, then prove the statement for the special case of uniformly distributed  $W$ , and finally extend the proof to the case where  $W$  has general absolutely continuous distribution with  $h(W) > -\infty$  and  $H(\lfloor W \rfloor) < \infty$ .

#### A. Auxiliary results

We begin with the following result on the existence of differential entropy.

**Lemma 4.** *Let  $f, g$  be probability densities on  $\mathbb{R}^n$ .*

1) *If  $\int_{\mathbb{R}^n} f(x) \log\left(\frac{1}{g(x)}\right) dx$  exists and does not equal  $+\infty$  (but possibly  $-\infty$ ), then  $\int_{\mathbb{R}^n} f(x) \log\left(\frac{1}{f(x)}\right) dx$  exists and*

$$\int_{\mathbb{R}^n} f(x) \log\left(\frac{1}{f(x)}\right) dx \leq \int_{\mathbb{R}^n} f(x) \log\left(\frac{1}{g(x)}\right) dx. \quad (108)$$

2) *If  $\int_{\mathbb{R}^n} f(x) \log\left(\frac{1}{f(x)}\right) dx$  is finite, then  $\int_{\mathbb{R}^n} f(x) \log\left(\frac{1}{g(x)}\right) dx$  exists and (108) holds.*

*Proof:* See [29, Lem. 8.3.1].  $\blacksquare$

**Corollary 1.** *Suppose that  $W$  is a random vector in  $\mathbb{R}^n$  with absolutely continuous distribution. If  $H(\lfloor W \rfloor) < \infty$ , then  $h(W)$  exists and*

$$h(W) \leq H(\lfloor W \rfloor).$$

*Proof:* Let  $f$  be the density of  $W$ . For  $x \in \mathbb{R}^n$ , we define a function  $g$  according to

$$g(x) := \int_{\mathcal{Q}(\ell)} f(y) dy,$$

where  $\mathcal{Q}(\ell) := [\ell_1, \ell_1 + 1) \times \dots \times [\ell_n, \ell_n + 1)$ , with  $\ell = (\ell_1, \dots, \ell_n) \in \mathbb{Z}^n$ , is the  $n$ -dimensional cube of side-length 1 containing  $x$ . Since  $g$  is constant on each cube  $\mathcal{Q}(\ell)$ , and each  $\mathcal{Q}(\ell)$  is a Borel set,  $g$  is measurable, and  $\int_{\mathbb{R}^n} g(x) dx = \int_{\mathbb{R}^n} f(x) dx = 1$  implies that  $g$  is a probability density. Decomposing  $\mathbb{R}^n$  into the union of all cubes  $\mathcal{Q}(\ell) = [\ell_1, \ell_1 + 1) \times \dots \times [\ell_n, \ell_n + 1)$ , for  $\ell = (\ell_1, \dots, \ell_n) \in \mathbb{Z}^n$ , we get

$$\begin{aligned} & \int_{\mathbb{R}^n} f(x) \log\left(\frac{1}{g(x)}\right) dx \\ &= \sum_{\ell \in \mathbb{Z}^n} \int_{\mathcal{Q}(\ell)} f(x) \log\left(\frac{1}{\int_{\mathcal{Q}(\ell)} f(y) dy}\right) dx \\ &= H(\lfloor W \rfloor) < \infty, \end{aligned}$$

and by the first part of Lemma 4, the proof is complete.  $\blacksquare$

We will make repeated use of the fact that the sum  $X + W$  of independent random vectors  $X$  and  $W$  in  $\mathbb{R}^n$  has absolutely continuous distribution if  $X$  or  $W$  is of absolutely continuous distribution [30, p. 175]. Assuming that  $W$  has density  $f_W$ , the density  $f_{X+W}$  of  $X + W$  is given by

$$f_{X+W}(y) = \mathbb{E}[f_W(y - X)] = \int_{\mathbb{R}^n} f_W(y - x) \mu(x), \quad (109)$$

where  $\mu$  is the distribution of  $X$ . Next we establish an existence result for the differential entropy term appearing on the RHS in (107).

**Proposition 8.** *Suppose that  $X$  and  $W$  are independent random vectors in  $\mathbb{R}^n$  such that  $W$  has an absolutely continuous distribution and  $H(\lfloor X \rfloor), H(\lfloor W \rfloor) < \infty$ . Then  $H(\lfloor X + tW \rfloor) < \infty$ , for all  $t > 0$ , and  $h(X + tW)$  exists and satisfies  $h(X + tW) \leq H(\lfloor X + tW \rfloor)$  for all  $t > 0$ . Moreover, we have*

$$\limsup_{t \rightarrow 0} \frac{h(X + tW)}{|\log t|} \leq 0. \quad (110)$$

*Proof:* Applying the chain rule, we get

$$\begin{aligned} & H(\lfloor X + tW \rfloor) \\ & \leq H(\lfloor X + tW \rfloor, \lfloor X \rfloor) \\ & = H(\lfloor X \rfloor) + H(\lfloor X + tW \rfloor | \lfloor X \rfloor) \\ & \leq H(\lfloor X \rfloor) + H(\lfloor tW \rfloor) + H(\lfloor X + tW \rfloor | \lfloor X \rfloor, \lfloor tW \rfloor). \end{aligned} \quad (111)$$

For  $i = 1, \dots, n$ , let  $X_i$  and  $W_i$  be the  $i$ -th entry of  $X$  and  $W$ , respectively. We next show that the three terms on the RHS of (111) are finite for all  $t > 0$  and, moreover, bounded uniformly for  $t \leq 1$ , which establishes the existence of  $h(X + tW)$  by Corollary 1 and also proves (110). The term  $H(\lfloor X \rfloor)$  in (111) is bounded by assumption. For the last term in (111), we have  $H(\lfloor X + tW \rfloor | \lfloor X \rfloor, \lfloor tW \rfloor) \leq n \log 2$ , since

$$\lfloor X_i \rfloor + \lfloor tW_i \rfloor \leq \lfloor X_i + tW_i \rfloor \leq \lfloor X_i \rfloor + \lfloor tW_i \rfloor + 1$$

for  $i = 1, \dots, n$ . To upper-bound the remaining term  $H(\lfloor tW \rfloor)$ , we apply the chain rule and find

$$H(\lfloor tW \rfloor) \leq H(\lfloor tW \rfloor | \lfloor tW \rfloor) + H(\lfloor tW \rfloor | \lfloor tW \rfloor) \quad (112)$$

$$\leq H(\lfloor W \rfloor) + H(\lfloor tW \rfloor | \lfloor tW \rfloor), \quad (113)$$

where (113) follows from the fact that  $\lfloor tW \rfloor$  is a function of  $\lfloor W \rfloor$  and the application of a function to a discrete random variable cannot increase entropy. Since

$$\begin{aligned} \lfloor tW_i \rfloor &\leq \lfloor tW_i \rfloor \leq \underbrace{t(W_i - \lfloor W_i \rfloor)}_{< 1} + t\lfloor W_i \rfloor \\ &< t + \lfloor t\lfloor W_i \rfloor \rfloor + 1, \end{aligned}$$

we have  $H(\lfloor tW \rfloor | \lfloor tW \rfloor) < \infty$  for all  $t > 0$ . Moreover,  $H(\lfloor tW \rfloor | \lfloor tW \rfloor)$  is uniformly bounded by  $n \log 2$  if  $t \leq 1$ . The sum  $X + tW$  has absolutely continuous distribution as a consequence of  $tW$  being of absolutely continuous distribution. It thus follows from Corollary 1 and  $H(\lfloor X + tW \rfloor) < \infty$  that  $h(X + tW)$  exists and  $h(X + tW) \leq H(\lfloor X + tW \rfloor)$ . Moreover, since  $H(\lfloor X + tW \rfloor)$  is uniformly bounded for all  $t \leq 1$ , we have

$$\limsup_{t \rightarrow 0} \frac{h(X + tW)}{|\log t|} \leq 0. \quad (114)$$

## B. Proof of Theorem 7

The proof of Theorem 7 constitutes one of the main technical difficulties in this paper. We proceed in two steps. First, we prove the statement for uniformly distributed noise  $W$ . Then, by employing a slight generalization of the Ruzsa triangle inequality for differential entropy [5], presented in Proposition 10 below, we extend it to general noise distributions satisfying  $h(W) > -\infty$  and  $H(\lfloor W \rfloor) < \infty$ . This considerably simplifies the proof of [4, Thm. 2.7] (for the scalar case), which is based on explicit manipulations of the densities involved. We begin by showing the statement for uniformly distributed noise with the corresponding proof inspired by [4, Thm. 3.1].

**Proposition 9.** *Let  $X$  and  $U$  be independent random vectors in  $\mathbb{R}^n$  such that  $H(\lfloor X \rfloor) < \infty$  and  $U$  is uniformly distributed on  $B(0; 1)$ , where the ball is taken with respect to the  $\ell^\infty$ -norm. Then, we have*

$$\limsup_{t \rightarrow 0} \frac{h(X + tU)}{|\log t|} = \bar{d}(X) - n.$$

*Proof:* Note first that the distribution of  $U$  is absolutely continuous and  $H(\lfloor U \rfloor) = 0 < \infty$ . Thus, by Proposition 8,  $H(\lfloor X + tU \rfloor) < \infty$  and  $h(X + tU)$  exists and satisfies  $h(X + tU) \leq H(\lfloor X + tU \rfloor)$ , for all  $t > 0$ . The density of  $X + tU$  is given by

$$f_t(x) := \frac{1}{t^n} \mu(B(x; t)), \quad (115)$$

where  $\mu$  is the distribution of  $X$ . As  $\log$  is an increasing function we have  $\mathbb{E}[\log \mu(B(X + tU; t))] \leq \mathbb{E}[\log 1] = 0$  and thus

$$h(X + tU) = \mathbb{E} \left[ \log \frac{t^n}{\mu(B(X + tU; t))} \right] \geq n \log t, \quad (116)$$

for all  $t > 0$ . Moreover, since  $U$  is uniformly distributed on  $B(0; 1)$ , we have  $\mu(B(X + tU; t)) \leq \mu(B(X; 2t))$  almost surely, and hence

$$\begin{aligned} \limsup_{t \rightarrow 0} \frac{h(X + tU)}{|\log t|} &\geq \limsup_{t \rightarrow 0} \frac{n \log t - \mathbb{E}[\log \mu(B(X; 2t))]}{|\log t|} \\ &= \limsup_{t \rightarrow 0} \frac{\mathbb{E}[\log \mu(B(X; 2t))]}{\log t} - n \\ &= \limsup_{t \rightarrow 0} \frac{\mathbb{E}[\log \mu(B(X; 2t))]}{\log(2t)} - n \\ &= \bar{d}(X) - n, \end{aligned} \quad (117)$$

where we used Lemma 1 in the last step.

To conclude the proof we need to show that  $\limsup_{t \rightarrow 0} \frac{h(X + tU)}{|\log t|} \leq \bar{d}(X) - n$ . To this end, we take  $\tilde{U}$  uniformly distributed on  $B(0, 2)$ , independent of  $X$ , and note that  $X + t\tilde{U}$  has density

$$g_t(x) := \frac{1}{(2t)^n} \mu(B(x; 2t)).$$

By (116) it follows that  $h(X + tU) > -\infty$ . Combined with  $h(X + tU) < \infty$  from above this means that  $h(X + tU)$  is finite for all  $t > 0$ . Therefore, we can apply the second part of Lemma 4 to find that

$$h(X + tU) \leq \int_{\mathbb{R}^n} f_t(x) \log \left( \frac{1}{g_t(x)} \right) dx. \quad (118)$$

Writing the density of  $X + tU$  as  $f_t(x) = \int_{\mathbb{R}^n} \frac{1}{t^n} \mathbb{1}_{B(x; t)}(y) \mu\langle y \rangle$ , we find that

$$\int_{\mathbb{R}^n} f_t(x) \log \left( \frac{1}{g_t(x)} \right) dx \quad (119)$$

$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{1}{t^n} \mathbb{1}_{B(x; t)}(y) \mu\langle y \rangle \log \left( \frac{(2t)^n}{\mu(B(x; 2t))} \right) dx \quad (120)$$

$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{1}{t^n} \mathbb{1}_{B(x; t)}(y) \log \left( \frac{(2t)^n}{\mu(B(x; 2t))} \right) dx \mu\langle y \rangle \quad (121)$$

$$= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{1}{t^n} \mathbb{1}_{B(0; t)}(x') \log \left( \frac{(2t)^n}{\mu(B(x' + y; 2t))} \right) dx' \mu\langle y \rangle \quad (122)$$

$$\leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{1}{t^n} \mathbb{1}_{B(0; t)}(x') \log \left( \frac{(2t)^n}{\mu(B(y; t))} \right) dx' \mu\langle y \rangle \quad (123)$$

$$= \int_{\mathbb{R}^n} \log \left( \frac{(2t)^n}{\mu(B(y; t))} \right) \mu\langle y \rangle \quad (124)$$

$$= \mathbb{E} \left[ \log \frac{1}{\mu(B(X; t))} \right] + n \log(2t), \quad (125)$$

where (121) follows from Fubini's Theorem, whose conditions are satisfied since the function in (121) is integrable which in turn follows from (125), in (121) we substituted  $x' := x - y$ , (123) is thanks to  $B(y; t) \subseteq B(x' + y; 2t)$  for all  $x' \in B(0; t)$ , and to get (124) we used  $\int_{\mathbb{R}^n} \frac{1}{t^n} \mathbb{1}_{B(0; t)}(x') dx' = 1$ . Together with (118) this yields

$$\limsup_{t \rightarrow 0} \frac{h(X + tU)}{|\log t|} \leq \limsup_{t \rightarrow 0} \frac{n \log(2t) - \mathbb{E}[\log \mu(B(X; t))]}{|\log t|}$$

$$= \limsup_{t \rightarrow 0} \frac{\mathbb{E}[\log \mu(B(X; t))]}{\log t} - n \quad (126)$$

$$= \bar{d}(X) - n, \quad (127)$$



where we used Lemma 1 in the last step. This completes the proof. ■

We turn to the second step in the proof of Theorem 7, namely the extension of Proposition 9 to noise  $W$  with general (absolutely continuous) distribution satisfying  $h(W) > -\infty$  and  $H(\lfloor W \rfloor) < \infty$ . To this end, we need a slight generalization of the Ruzsa triangle inequality for differential entropy stated in [5, Thm. 3.1] for scalar random variables with absolutely continuous distributions. The generalization that we state below i) applies to random vectors, and ii) allows for a general distribution for  $X$  as long as  $H(\lfloor X \rfloor) < \infty$ .

**Proposition 10.** *Let  $X$ ,  $Y$ , and  $Z$  be independent random vectors in  $\mathbb{R}^n$  such that  $H(\lfloor X \rfloor), H(\lfloor Y \rfloor), H(\lfloor Z \rfloor) < \infty$  and  $Y$  and  $Z$  have absolutely continuous distribution with  $h(Y), h(Z) > -\infty$ . Then, we have*

$$h(X + Z) - h(X + Y) \leq h(Y - Z) - h(Y). \quad (128)$$

*Proof:* First, note that, by Proposition 8, the differential entropies  $h(X + Z)$ ,  $h(X + Y)$  and  $h(Y - Z)$  exist.<sup>14</sup> We have

$$h(X + Z) - h(Z) \quad (129)$$

$$= I(X; X + Z) \quad (130)$$

$$\leq I(X; (X + Y, Y - Z)) \quad (131)$$

$$= h(X + Y, Y - Z) - h(Y, Y - Z) \quad (132)$$

$$= h(X + Y, Y - Z) - h(Y, Z) \quad (133)$$

$$= h(X + Y, Y - Z) - h(Y) - h(Z) \quad (134)$$

$$\leq h(X + Y) + h(Y - Z) - h(Y) - h(Z), \quad (135)$$

where (131) is by the data processing inequality noting that  $X - (X + Y, Y - Z) - X + Z$  forms a Markov chain, for (133) we applied the coordinate transformation  $(x, y) \mapsto (x, x - y)$  to the vector in the second term in (132) and used the fact that the determinant of the corresponding Jacobian has absolute value 1 together with [32, (8.71), p. 254], (134) follows from the independence of  $Y$  and  $Z$ , and (135) is by the chain rule. Adding  $h(Z) - h(X + Y)$  to both sides yields (128). ■

We are now ready to complete the proof of Theorem 7. Let  $U$  be uniformly distributed on the  $\ell^\infty$ -ball  $B(0; 1)$  and independent of  $(X, W)$  for  $X$  and  $W$  as in the statement of Theorem 7 and let  $t > 0$ . Noting that  $U$  and  $W$  have absolutely continuous distributions,  $h(W) > -\infty$  and  $H(\lfloor W \rfloor) < \infty$  by assumption, and  $h(U) = H(\lfloor U \rfloor) = 0$ , we apply Proposition 10 twice to get

$$-h(tU - tW) + h(tU) \leq h(X + tU) - h(X + tW) \quad (136)$$

$$\leq h(tW - tU) - h(tW). \quad (137)$$

Corollary 1 and Proposition 8 imply that all differential entropies appearing in (136) and (137) exist and  $h(tU - tW), h(tW - tU) < \infty$  for all  $t > 0$ . Furthermore, by the scaling property of differential entropy we have  $h(tU - tW) =$

<sup>14</sup>Note that  $h(Y), h(Z) < \infty$  is not sufficient for  $h(Y - Z) < \infty$  to hold. In fact, [31, Prop. V.8] shows that there exist  $Y$  with  $h(Y)$  finite such that for every  $Z$  with  $h(Z)$  finite, we have  $h(Y - Z) = \infty$ .

$h(tW - tU)$ . Moreover, since the density  $f_U$  of  $U$  is bounded, i.e.,  $f_U(x) \leq C$  for all  $x \in \mathbb{R}^n$ , it follows that the density  $f_{tU - tW}$  of  $tU - tW$  is bounded as well for all  $t > 0$ , since

$$f_{tU - tW}(x) = \frac{1}{t^n} \int_{\mathbb{R}^n} f_U(u) f_W\left(u - \frac{x}{t}\right) du \leq \frac{C}{t^n},$$

and thus

$$\begin{aligned} h(tU - tW) &= h(tW - tU) \\ &\geq \int_{\mathbb{R}^n} f_{tU - tW}(x) \log\left(\frac{t^n}{C}\right) dx \\ &= \log\left(\frac{t^n}{C}\right) > -\infty, \end{aligned}$$

for all  $t > 0$ . Since  $h(U) = 0$  and  $h(W)$  is finite by Corollary 1 and the assumption  $h(W) > -\infty$ , we find that  $h(U - W)$  and  $h(W - U)$  are finite, which yields

$$\begin{aligned} &\lim_{t \rightarrow 0} \frac{-h(tU - tW) + h(tU)}{|\log t|} \\ &= \lim_{t \rightarrow 0} \frac{-h(U - W) - n \log t + h(U) + n \log t}{|\log t|} = 0 \\ &\stackrel{(136)}{\leq} \lim_{t \rightarrow 0} \frac{h(X + tU) - h(X + tW)}{|\log t|} \\ &\stackrel{(137)}{\leq} \lim_{t \rightarrow 0} \frac{h(tW - tU) - h(tW)}{|\log t|} \\ &= \lim_{t \rightarrow 0} \frac{h(W - U) + n \log t - h(W) - n \log t}{|\log t|} = 0. \end{aligned}$$

We can therefore conclude that

$$\limsup_{t \rightarrow 0} \frac{h(X + tW)}{|\log t|} = \limsup_{t \rightarrow 0} \frac{h(X + tU)}{|\log t|} \stackrel{\text{Prop. 9}}{=} \bar{d}(X) - n,$$

which completes the proof. ■

## APPENDIX B PROOFS OF THEOREMS 2–5

The proofs build on the arguments in [3, Sect. V-B], which need to be extended to the vector case. To keep this paper as self-contained as possible, we detail the corresponding extensions and also include the steps of [3] that remain unchanged. We hasten to add that the main conceptual components in the proofs in this Appendix should be attributed to Wu et al. [3].

### A. Auxiliary results

We begin with a number of auxiliary results.

**Lemma 5.** *Let  $\mathbf{X}$  be a random matrix in  $\mathbb{R}^{M \times N}$  and let  $k \in \mathbb{N}$ . Then, we have*

$$0 \leq H([\mathbf{X}]_k) - \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-k}))} \right] \leq MN \log 3,$$

where  $\mu$  is the distribution of  $\mathbf{X}$  and  $B(\mathbf{X}; \varepsilon)$  is the  $\ell^\infty$ -ball<sup>15</sup> with center  $\mathbf{X}$  and radius  $\varepsilon$ .

<sup>15</sup> $B(\mathbf{A}; \varepsilon) := \{\mathbf{B} \in \mathbb{R}^{M \times N} \mid \|\mathbf{A} - \mathbf{B}\|_\infty < \varepsilon\}$  where  $\|\cdot\|_\infty$  is defined in the Notation paragraph.

*Proof:* For  $\mathbf{A} = (a_{i,j})_{\substack{1 \leq i \leq M \\ 1 \leq j \leq N}}$  we define the ‘‘cube’’ in  $\mathbb{R}^{M \times N}$  of sidelength  $2^{-k}$  containing  $\mathbf{A}$  as

$$\mathcal{Q}_k(\mathbf{A}) := \left[ [a_{1,1}]_k, [a_{1,1}]_k + 2^{-k} \right) \times \dots \times \left[ [a_{M,N}]_k, [a_{M,N}]_k + 2^{-k} \right). \quad (138)$$

Consider the decomposition of  $\mathbb{R}^{M \times N}$  into cubes with sidelength  $2^{-k}$  according to

$$\mathbb{R}^{M \times N} = \bigcup_{\mathcal{Q} \in \{\mathcal{Q}_k(\mathbf{A}) | \mathbf{A} \in \mathbb{R}^{M \times N}\}} \mathcal{Q}. \quad (139)$$

We will also need cubes of sidelength  $3 \cdot 2^{-k}$ , defined according to

$$\widehat{\mathcal{Q}}_k(\mathbf{A}) := \left[ [a_{1,1}]_k - 2^{-k}, [a_{1,1}]_k + 2^{-k+1} \right) \times \dots \times \left[ [a_{M,N}]_k - 2^{-k}, [a_{M,N}]_k + 2^{-k+1} \right).$$

Then, we have  $\mathcal{Q}_k(\mathbf{A}) \subseteq B(\mathbf{A}; 2^{-k}) \subseteq \widehat{\mathcal{Q}}_k(\mathbf{A})$  and since  $H([\mathbf{X}]_k) = \mathbb{E} \left[ \log \frac{1}{\mu(\mathcal{Q}_k(\mathbf{X}))} \right]$ , we get

$$0 \leq H([\mathbf{X}]_k) - \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-k}))} \right] \quad (140)$$

$$= \mathbb{E} \left[ \log \frac{\mu(B(\mathbf{X}; 2^{-k}))}{\mu(\mathcal{Q}_k(\mathbf{X}))} \right] \quad (141)$$

$$\leq \mathbb{E} \left[ \log \frac{\mu(\widehat{\mathcal{Q}}_k(\mathbf{X}))}{\mu(\mathcal{Q}_k(\mathbf{X}))} \right] \quad (142)$$

$$\stackrel{\text{Jensen}}{\leq} \log \mathbb{E} \left[ \frac{\mu(\widehat{\mathcal{Q}}_k(\mathbf{X}))}{\mu(\mathcal{Q}_k(\mathbf{X}))} \right] \quad (143)$$

$$= MN \log 3, \quad (144)$$

where in (144) we used the fact that for each  $\mathbf{A} \in \mathbb{R}^{M \times N}$  the function  $\mu(\mathcal{Q}_k(\mathbf{X}))$  is constant on the event  $\mathbf{X} \in \mathcal{Q}_k(\mathbf{A})$  and that each  $\widehat{\mathcal{Q}}_k(\mathbf{A})$  is the union of  $3^{MN}$  cubes of the form (138). ■

**Lemma 6.** *Let  $X$  be a random vector in  $\mathbb{R}^n$ . For  $a > 1$  we have*

$$\underline{d}(X) = \liminf_{\ell \rightarrow \infty} \frac{H(\langle X \rangle_{a^\ell})}{\log(a^\ell)} \quad \text{and} \quad \bar{d}(X) = \limsup_{\ell \rightarrow \infty} \frac{H(\langle X \rangle_{a^\ell})}{\log(a^\ell)}.$$

*Proof:* The proof follows closely that of the corresponding result for the scalar case reported in [18, Prop. 2], and is provided in the Online Addendum [2, Lem. 5] for completeness. ■

**Lemma 7.** *Let  $\mathbf{X}$  and  $\mathbf{W}$  be independent random matrices in  $\mathbb{R}^{M \times N}$ , where  $H([\mathbf{X}]) < \infty$  and  $\mathbf{W}$  has an absolutely continuous distribution with  $h(\mathbf{W}) > -\infty$  and  $H([\mathbf{W}]) < \infty$ . Furthermore, let  $\mathbf{U}$  be uniformly distributed on  $B(\mathbf{0}; 1)$  and independent of  $\mathbf{X}$  and  $\mathbf{W}$ . For  $k \in \mathbb{N}$ , we have*

$$\begin{aligned} & \frac{1}{MN} \left| h(2^k \mathbf{X} + \mathbf{W}) - \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-k}))} \right] \right| \\ & \leq \frac{\max\{h(\mathbf{W} - \mathbf{U}) - h(\mathbf{W}), h(\mathbf{U} - \mathbf{W}) - h(\mathbf{U})\}}{MN} + \log 6, \end{aligned} \quad (145)$$

where  $\mu$  is the distribution of  $\mathbf{X}$  and  $B(\mathbf{A}; \varepsilon)$  denotes the  $\ell^\infty$ -ball with center  $\mathbf{A} \in \mathbb{R}^{M \times N}$  and radius  $\varepsilon$ .

*Proof:* Since  $h(\mathbf{U}) > -\infty$  and  $H([\mathbf{U}]) < \infty$ , we can apply the Ruzsa triangle inequality (128) and find

$$\begin{aligned} & |h(2^k \mathbf{X} + \mathbf{W}) - h(2^k \mathbf{X} + \mathbf{U})| \\ & \leq \max\{h(\mathbf{W} - \mathbf{U}) - h(\mathbf{W}), h(\mathbf{U} - \mathbf{W}) - h(\mathbf{U})\}. \end{aligned} \quad (146)$$

The RHS of (146) is finite by Proposition 8 and Corollary 1. By the scaling property of differential entropy, we have  $h(2^k \mathbf{X} + \mathbf{U}) = h(\mathbf{X} + 2^{-k} \mathbf{U}) + MNk \log 2$ , and following the steps (118)–(125), we find that

$$h(2^k \mathbf{X} + \mathbf{U}) \quad (147)$$

$$= h(\mathbf{X} + 2^{-k} \mathbf{U}) + MNk \log 2 \quad (148)$$

$$\leq \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-k}))} \right] + MNk \log 2 + MN \log(2^{-k+1}) \quad (149)$$

$$= \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-k}))} \right] + MN \log 2. \quad (150)$$

For a bound in the other direction, we note that  $B(\mathbf{X} + 2^{-k} \mathbf{U}; 2^{-k}) \subseteq B(\mathbf{X}; 2^{-(k-1)})$  and thus

$$\begin{aligned} h(2^k \mathbf{X} + \mathbf{U}) & \stackrel{(115)}{=} \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X} + 2^{-k} \mathbf{U}; 2^{-k}))} \right] \\ & \geq \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-(k-1)})} \right]. \end{aligned}$$

It follows that

$$\mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-k}))} \right] - h(2^k \mathbf{X} + \mathbf{U}) \quad (151)$$

$$\leq \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-k}))} \right] - \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-(k-1)})} \right] \quad (152)$$

$$\stackrel{\text{Lemma 5}}{\leq} H([\mathbf{X}]_k) - H([\mathbf{X}]_{k-1}) + MN \log 3 \quad (153)$$

$$\leq H([\mathbf{X}]_k | [\mathbf{X}]_{k-1}) + MN \log 3 \quad (154)$$

$$\leq MN(\log 2 + \log 3), \quad (155)$$

where in (154) we applied the chain rule, and (155) follows since given  $[\mathbf{X}]_{k-1}$  there are at most  $2^{MN}$  possible values for  $[\mathbf{X}]_k$ . Putting together (146), (150), and (155) completes the proof. ■

The next lemma, which is a straightforward extension of [3, Lem. 4] to the vector case, provides a sufficient condition for the iterated function system used in the construction of the random vector in (22) to satisfy the open set condition.

**Lemma 8.** *Consider the iterated function system  $\{F_1, \dots, F_m\}$  with  $F_i(x) = rx + w_i$ , for  $x \in \mathbb{R}^n$ ,  $r \in (0, 1)$ , and pairwise different vectors  $w_1, \dots, w_m \in \mathbb{R}^n$ . Let furthermore  $\mathcal{W} := \{w_1, \dots, w_m\}$ . Then, the open set condition (see Definition 2) is satisfied if*

$$r \leq \frac{m(\mathcal{W})}{m(\mathcal{W}) + M(\mathcal{W})}. \quad (156)$$

*Proof:* The proof follows closely that of the corresponding result for the scalar case reported in [3, Lem. 4], and is provided in the Online Addendum [2, Lem. 1] for completeness. ■

The following lemma provides an upper bound on the difference in entropy of two discrete random matrices.

**Lemma 9** ([3, Lem. 7]). *Let  $\mathbf{U}$  and  $\mathbf{V}$  be discrete random matrices in  $\mathbb{R}^{M \times N}$  such that  $\|\mathbf{U} - \mathbf{V}\|_\infty \leq \varepsilon$  and  $m(\mathcal{U}) \geq \delta > 0$  where  $\mathcal{U}$  is the set of possible realizations of  $\mathbf{U}$ . Then,*

$$H(\mathbf{U}) - H(\mathbf{V}) \leq MN \log \left( 1 + \left\lfloor \frac{2\varepsilon}{\delta} \right\rfloor \right).$$

*Proof:* By the chain rule we have

$$H(\mathbf{U}) - H(\mathbf{V}) \leq H(\mathbf{U}|\mathbf{V}).$$

The result now follows from the fact that given  $\mathbf{V}$  each of the  $MN$  entries of  $\mathbf{U}$  can take on at most  $1 + \lfloor \frac{2\varepsilon}{\delta} \rfloor$  different values. ■

Next, we bound the error in the entropy of the quantized output signals which results from replacing the input distributions by their fractional parts, a crucial step in the proof of Theorem 2.

**Lemma 10.** *Consider the deterministic matrices  $\mathbf{H}_1, \dots, \mathbf{H}_K \in \mathbb{R}^{M \times M}$  and let  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  be random matrices in  $\mathbb{R}^{M \times N}$ . For every  $k \in \mathbb{N}$ , we have*

$$\left| H \left( \left[ \sum_{j=1}^K \mathbf{H}_j \mathbf{X}_j^N \right]_k \right) - H \left( \left[ \sum_{j=1}^K \mathbf{H}_j (\mathbf{X}_j^N) \right]_k \right) \right| \leq \sum_{j=1}^K H(\lfloor \mathbf{X}_j^N \rfloor) + MN \log 2,$$

where  $(\mathbf{A}) := \mathbf{A} - \lfloor \mathbf{A} \rfloor$  denotes the fractional part of the real matrix  $\mathbf{A}$ .

*Proof:* The proof follows closely that of the corresponding result for the scalar case reported in [3, Lem. 10], and is provided in the Online Addendum [2, Lem. 2] for completeness. ■

The following result shows that  $H(\lfloor \mathbf{X} \rfloor)$  is finite if all entries of  $\mathbf{X}$  have finite second moment.

**Lemma 11.** *Let  $\mathbf{X}^N$  be a random matrix in  $\mathbb{R}^{M \times N}$  with columns  $X^{(1)}, \dots, X^{(N)}$  where  $X^{(n)} = (X^{(n)}[1], \dots, X^{(n)}[M]) \in \mathbb{R}^M$ , for  $n = 1, \dots, N$ . If*

$$\sum_{m=1}^M \sum_{n=1}^N \mathbb{E}[(X^{(n)}[m])^2] \leq MN, \quad (157)$$

then

$$H(\lfloor \mathbf{X} \rfloor) \leq \frac{MN}{2} \log \left( \frac{26}{3} \pi e \right). \quad (158)$$

*Proof:*

$$H(\lfloor \mathbf{X} \rfloor) \quad (159)$$

$$\leq \sum_{m=1}^M \sum_{n=1}^N H(\lfloor X^{(n)}[m] \rfloor) \quad (160)$$

$$\leq MN \log 2 + \sum_{m=1}^M \sum_{n=1}^N H(\lfloor |X^{(n)}[m]| \rfloor) \quad (161)$$

$$\leq MN \log 2 + \sum_{m=1}^M \sum_{n=1}^N \frac{1}{2} \log \left( 2\pi e \left( \mathbb{E}[(X^{(n)}[m])^2] + \frac{1}{12} \right) \right) \quad (162)$$

$$\leq \frac{MN}{2} \left( \log 4 + \log \left( 2\pi e \left( \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \mathbb{E}[(X^{(n)}[m])^2] + \frac{1}{12} \right) \right) \right) \quad (163)$$

$$\leq \frac{MN}{2} \log \left( \frac{26}{3} \pi e \right), \quad (164)$$

where (160) is by the chain rule, (161) follows from

$$\begin{aligned} H(\lfloor X \rfloor) &\leq H(\lfloor X \rfloor, \lfloor |X| \rfloor) \\ &= H(\lfloor |X| \rfloor) + H(\lfloor X \rfloor | \lfloor |X| \rfloor) \\ &\leq H(\lfloor |X| \rfloor) + \log 2 \end{aligned}$$

for arbitrary random variables  $X$ , in (162) we used [32, Eq. (8.94)], (163) is due to Jensen's inequality, and (164) follows from (157). ■

The following lemma is a straightforward extension of [3, Lem. 14] to the vector case.

**Lemma 12.** *Let  $\mathcal{V} \subseteq \mathbb{R}^n$  be a set such that  $0 < m(\mathcal{V}), M(\mathcal{V}) < \infty$  and let  $r > 0$  be such that*

$$r \leq \frac{m(\mathcal{V})}{m(\mathcal{V}) + M(\mathcal{V})}. \quad (165)$$

Then, for every  $\ell \in \mathbb{N}$  with  $\ell \geq 1$ , we have

$$m(\mathcal{V} + r\mathcal{V} + \dots + r^{\ell-1}\mathcal{V}) \geq r^{\ell-1}m(\mathcal{V}). \quad (166)$$

Moreover, the mapping  $\mathcal{V}^\ell \rightarrow \mathcal{V} + r\mathcal{V} + \dots + r^{\ell-1}\mathcal{V}$ ,  $(v_1, \dots, v_\ell) \mapsto v_1 + rv_2 + \dots + r^{\ell-1}v_\ell$  is a one-to-one correspondence.

*Proof:* The proof follows closely that of the corresponding result for the scalar case reported in [3, Lem. 13], and is provided in the Online Addendum [2, Lem. 3] for completeness. ■

We next derive an alternative limiting expression for DoF( $\mathbf{H}$ ) in terms of the entropy of quantized versions of the noiseless channel output vectors. We restrict the corresponding statement in Proposition 11 below to Gaussian noise so that it can be applied directly in the proof of Theorem 3. However, in the proof of Theorem 5, we show that the statement continues to hold for more general noise distributions, specifically those satisfying the conditions of Theorem 5.

**Proposition 11.** *The DoF of the channel (1) (with Gaussian noise) are given by*

$$\text{DoF}(\mathbf{H}) = \limsup_{k \rightarrow \infty} \lim_{N \rightarrow \infty} \sup_{\mathbf{X}_1^N, \dots, \mathbf{X}_K^N} \frac{1}{Nk} \sum_{i=1}^K \left[ H \left( \left[ \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k \right) - H \left( \left[ \sum_{j \neq i}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k \right) \right],$$

where  $\mathbf{X}_i^N = (X_i^{(1)} \dots X_i^{(N)})$  and the supremum is taken over all independent  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  satisfying

$$\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \mathbb{E} \left[ (X_i^{(n)}[m])^2 \right] \leq 1, \quad \text{for all } i. \quad (167)$$

*Proof:* We begin by arguing that in computing  $\limsup_{\text{snr} \rightarrow \infty} \frac{C_{\text{sum}}(\mathbf{H}; \text{snr})}{\frac{1}{2} \log \text{snr}}$  we may restrict to an exponential sequence  $\text{snr}_k$ . Specifically, let  $\text{snr}_k$  be an increasing sequence of positive real numbers such that  $\text{snr}_{k+1} \leq c \cdot \text{snr}_k$ , for all  $k \in \mathbb{N}$  and some real constant  $c > 1$ . For a given  $\text{snr} > \text{snr}_0$ , we can then find a  $k$  such that  $\text{snr}_k < \text{snr} \leq \text{snr}_{k+1}$ . Since  $C_{\text{sum}}(\mathbf{H}; \text{snr})$  is monotonic in  $\text{snr}$ , we get

$$\frac{C_{\text{sum}}(\mathbf{H}; \text{snr})}{\frac{1}{2} \log \text{snr}} \leq \frac{C_{\text{sum}}(\mathbf{H}; \text{snr}_{k+1})}{\frac{1}{2} \log \text{snr}_k} \quad (168)$$

$$= \frac{C_{\text{sum}}(\mathbf{H}; \text{snr}_{k+1})}{\frac{1}{2} \log \text{snr}_{k+1} + \frac{1}{2} \log \frac{\text{snr}_k}{\text{snr}_{k+1}}} \quad (169)$$

$$\leq \frac{C_{\text{sum}}(\mathbf{H}; \text{snr}_{k+1})}{\frac{1}{2} \log \text{snr}_{k+1} - \frac{1}{2} \log c}. \quad (170)$$

This implies

$$\limsup_{\text{snr} \rightarrow \infty} \frac{C_{\text{sum}}(\mathbf{H}; \text{snr})}{\frac{1}{2} \log \text{snr}} \leq \limsup_{k \rightarrow \infty} \frac{C_{\text{sum}}(\mathbf{H}; \text{snr}_k)}{\frac{1}{2} \log \text{snr}_k}. \quad (171)$$

By virtue of  $\text{snr}_k$  being a sequence that tends to  $\infty$ , we also have

$$\limsup_{\text{snr} \rightarrow \infty} \frac{C_{\text{sum}}(\mathbf{H}; \text{snr})}{\frac{1}{2} \log \text{snr}} \geq \limsup_{k \rightarrow \infty} \frac{C_{\text{sum}}(\mathbf{H}; \text{snr}_k)}{\frac{1}{2} \log \text{snr}_k}, \quad (172)$$

which implies equality in (171). From now on we restrict ourselves to the particular subsequence  $\text{snr}_k = 4^k$ .

As a second preliminary remark, we note that for independent  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  with  $\mathbf{X}_i^N = (X_i^{(1)} \dots X_i^{(N)})$  and  $X_i^{(n)} \in \mathbb{R}^M$ , we have

$$\begin{aligned} I(\mathbf{X}_i^N; \mathbf{Y}_i^N) &= I \left( \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N, \text{snr} \right) - I \left( \sum_{j \neq i}^K \mathbf{H}_{i,j} \mathbf{X}_j^N, \text{snr} \right), \end{aligned}$$

where, for a random matrix  $\mathbf{X}$ , we set  $I(\mathbf{X}, \text{snr}) := I(\mathbf{X}; \sqrt{\text{snr}}\mathbf{X} + \mathbf{W})$  with  $\mathbf{W}$  consisting of i.i.d. zero-mean Gaussian random vectors, independent of  $\mathbf{X}$  and with identity

covariance matrix. The limiting characterization of sum-rate capacity in Proposition 3 thus yields

$$\text{DoF}(\mathbf{H}) = \limsup_{\text{snr} \rightarrow \infty} \lim_{N \rightarrow \infty} \sup_{\mathbf{X}_1^N, \dots, \mathbf{X}_K^N} \frac{2}{N \log \text{snr}} \sum_{i=1}^K \left[ I \left( \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N, \text{snr} \right) - \left( \sum_{j \neq i}^K \mathbf{H}_{i,j} \mathbf{X}_j^N, \text{snr} \right) \right], \quad (173)$$

where the supremum is taken over all independent  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  satisfying the power constraint (167).

Since  $h(\mathbf{W}) = \frac{MN}{2} \log(2\pi e) > -\infty$ ,  $H([\mathbf{W}]) < \infty$  by Lemma 11, and the RHS of (145) is bounded by a numerical constant as a consequence of  $h(\mathbf{W} - \mathbf{U}), h(\mathbf{W}), h(\mathbf{U} - \mathbf{W}), h(\mathbf{U})$  all being finite, it follows from Lemma 7 that for a random matrix  $\mathbf{X}$  with  $H([\mathbf{X}]) < \infty$ , we have

$$\left| I(\mathbf{X}, \text{snr}_k) - \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-k}))} \right] \right| \leq MNc, \quad (174)$$

where  $\mu$  is the distribution of  $\mathbf{X}$ ,  $c$  is a numerical constant, we used  $I(\mathbf{X}, \text{snr}_k) = h(2^k \mathbf{X} + \mathbf{W}) - h(\mathbf{W})$ , and moved  $h(\mathbf{W})$  to the other side in inequality (145). By Lemma 5, we also have

$$0 \leq H([\mathbf{X}]_k) - \mathbb{E} \left[ \log \frac{1}{\mu(B(\mathbf{X}; 2^{-k}))} \right] \leq MN \log 3. \quad (175)$$

Therefore, we get for the mutual information terms in (173) that

$$- \frac{M(c + \log 3)}{\frac{1}{2} \log \text{snr}_k} \quad (176)$$

$$\leq \frac{I \left( \sum_j \mathbf{H}_{i,j} \mathbf{X}_j^N, \text{snr}_k \right) - H \left( \left[ \sum_j \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k \right)}{N \frac{1}{2} \log \text{snr}_k} \quad (177)$$

$$\leq \frac{M(c + \log 3)}{\frac{1}{2} \log \text{snr}_k}, \quad (178)$$

for all independent  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  satisfying (167). The lower bound (176) and the upper bound (178) tend to zero for  $k \rightarrow \infty$ , and do so uniformly with respect to  $N$  and the inputs  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$ . Therefore, we may replace the mutual information terms in (173) by the corresponding entropy terms to get

$$\text{DoF}(\mathbf{H}) = \limsup_{k \rightarrow \infty} \lim_{N \rightarrow \infty} \sup_{\mathbf{X}_1^N, \dots, \mathbf{X}_K^N} \frac{1}{Nk} \sum_{i=1}^K \left[ H \left( \left[ \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k \right) - H \left( \left[ \sum_{j \neq i}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k \right) \right], \quad (179)$$

where we used  $\text{snr}_k = 4^k$ . ■

### B. Proof of Theorem 2

First, we argue that we may assume

$$\sum_{m=1}^M \mathbb{E}[(X_i[m])^2] \leq M, \quad i = 1, \dots, K, \quad (180)$$

in addition to  $H(\lfloor X_i \rfloor) < \infty$  and the existence of all information dimension terms appearing in (24). This follows, since first by Lemma 6 we can compute the information dimension by

$$d\left(\left[\sum_j \mathbf{H}_{i,j} X_j\right]_k\right) = \lim_{k \rightarrow \infty} \frac{H\left(\left[\sum_j \mathbf{H}_{i,j} X_j\right]_k\right)}{k},$$

and thus by Lemma 10 and the assumption  $H(\lfloor X_i \rfloor) < \infty$  the RHS of (24) does not change when we replace the inputs by their respective fractional parts. The argument is concluded by noting that the fractional parts always satisfy (180) as a consequence of  $0 \leq (x) < 1$ , for  $x \in \mathbb{R}$ .

Now let  $X_1, \dots, X_K$  be independent and such that the information dimension terms in (24) exist and the power constraint (180) is satisfied. For each  $i = 1, \dots, K$  and for given  $N$ , we take the  $N$  columns of  $\mathbf{X}_i^N$  to be i.i.d. copies of  $X_i$ . Since the  $X_i$  satisfy (180), the  $\mathbf{X}_i^N$  satisfy (167), and by Proposition 3, we find that

$$C_{\text{sum}}(\mathbf{H}; \text{snr}) \quad (181)$$

$$\geq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^K I(\mathbf{X}_i^N; \mathbf{Y}_i^N) \quad (182)$$

$$\geq \sum_{i=1}^K I(X_i; Y_i) \quad (183)$$

$$= \sum_{i=1}^K \left[ I\left(\sum_{j=1}^K \mathbf{H}_{i,j} X_j, \text{snr}\right) - I\left(\sum_{j \neq i}^K \mathbf{H}_{i,j} X_j, \text{snr}\right) \right]. \quad (184)$$

It therefore follows that

$$\text{DoF}(\mathbf{H}) \quad (185)$$

$$= \limsup_{\text{snr} \rightarrow \infty} \frac{C_{\text{sum}}(\mathbf{H}; \text{snr})}{\frac{1}{2} \log \text{snr}} \quad (186)$$

$$\geq \limsup_{\text{snr} \rightarrow \infty} \frac{1}{\frac{1}{2} \log \text{snr}} \sum_{i=1}^K \left[ I\left(\sum_{j=1}^K \mathbf{H}_{i,j} X_j, \text{snr}\right) - I\left(\sum_{j \neq i}^K \mathbf{H}_{i,j} X_j, \text{snr}\right) \right] \quad (187)$$

$$\stackrel{\text{Thm. 6}}{=} \sum_{i=1}^K \left[ d\left(\sum_{j=1}^K \mathbf{H}_{i,j} X_j\right) - d\left(\sum_{j \neq i}^K \mathbf{H}_{i,j} X_j\right) \right], \quad (188)$$

$$= \text{dof}(X_1, \dots, X_K; \mathbf{H}) \quad (189)$$

where we used that, by assumption, the information dimension terms in (188) exist.  $\blacksquare$

### C. Proof of Theorem 3

As achievability, i.e.,

$$\sup_{X_1, \dots, X_K} \text{dof}(X_1, \dots, X_K; \mathbf{H}) \leq \text{DoF}(\mathbf{H}) \quad (190)$$

was established in Theorem 2, it remains to prove the converse.

The proof architecture is as follows. Starting from Proposition 11 we have a limiting characterization of  $\text{DoF}(\mathbf{H})$  based on terms that are related to information dimension through Lemma 6. The main task is to show that for arbitrary multi-letter inputs we can construct corresponding single-letter inputs that achieve the same value in the supremization in Proposition 11 as the underlying multi-letter inputs. These single-letter inputs are obtained by discretizing the underlying multi-letter inputs and ‘‘encoding’’ the resulting discrete random vectors using self-similar distributions. The corresponding (noiseless) output distributions are then shown to also be self-similar. This means that we can compute the information dimension of the output distributions using Theorem 1. We will, however, have to ensure that the open set condition is satisfied, which will be accomplished by using a result from Diophantine approximation theory. This result applies to almost all channel matrices  $\mathbf{H}$  only, with the consequence of the statement of Theorem 3 being restricted to almost all channels. We emphasize that the idea for this construction was first proposed in [3] for the scalar case.

We begin by noting that Proposition 11 implies that  $\text{DoF}(\mathbf{H})$  can be expressed as

$$\text{DoF}(\mathbf{H}) = \limsup_{k \rightarrow \infty} \lim_{N \rightarrow \infty} \sup_{\mathbf{X}_1^N, \dots, \mathbf{X}_K^N} \frac{1}{Nk} \sum_{i=1}^K \left[ H\left(\left[\sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N\right]_k\right) - H\left(\left[\sum_{j \neq i}^K \mathbf{H}_{i,j} \mathbf{X}_j^N\right]_k\right) \right], \quad (191)$$

where the supremum is taken over all independent  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  satisfying the power constraint (167). We now want to apply Lemma 10 to find that we can replace the inputs in (191) by their fractional parts, which, in turn, means that it suffices to consider  $\mathbf{X}_i^N$ ,  $i = 1, \dots, K$ , that take on values in  $[0, 1]^{M \times N}$  only. For this to work out it suffices to establish an upper bound on  $\sum_{j=1}^K H(\lfloor \mathbf{X}_j^N \rfloor)$  that is uniform with respect to the inputs  $\mathbf{X}_i^N$  and grows at most linearly in  $N$ . Such an upper bound follows immediately from Lemma 11, taking into account the power constraint (167). We can therefore conclude that for every  $\varepsilon > 0$  and every  $N_0 > 0$ , there exist  $N, k \geq N_0$  and independent  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  taking on values in  $[0, 1]^{M \times N}$  such that

$$\text{DoF}(\mathbf{H}) \leq \varepsilon + \frac{1}{Nk} \quad (192)$$

$$\sum_{i=1}^K \left[ H\left(\left[\sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N\right]_k\right) - H\left(\left[\sum_{j \neq i}^K \mathbf{H}_{i,j} \mathbf{X}_j^N\right]_k\right) \right]. \quad (193)$$

It remains to single-letterize (191) by constructing  $\tilde{X}_1, \dots, \tilde{X}_K$  out of the  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  such that

dof $(\tilde{X}_1, \dots, \tilde{X}_K; \mathbf{H})$  upper-bounds the RHS of (193). This then completes the proof.

We begin the construction of the single-letter transmit vectors with preparatory steps. Consider monomials in the  $(KM)^2$  channel coefficients, i.e., let  $f_1, f_2, \dots$  be the monomials<sup>16</sup> of all degrees<sup>17</sup> in  $(KM)^2$  variables enumerated as follows:  $f_1, \dots, f_{\varphi(d)}$  are the monomials of degree not larger than  $d$ , where

$$\varphi(d) := \binom{(KM)^2 + d}{d}.$$

As  $\varphi(d)$  and  $\varphi(d+1)$  will occur frequently in the remainder of the proof, we simplify the notation by setting  $L := \varphi(d)$  and  $L' := \varphi(d+1)$ . Since  $L' = \frac{(KM)^2 + d + 1}{d+1}L$ , we have  $(L' - L)/L' \xrightarrow{d \rightarrow \infty} 0$ . Suppose we are given  $\varepsilon > 0$  in (193). We choose  $d$  large enough so that

$$(L' - L)/L' < \varepsilon. \quad (194)$$

We will construct single-letter transmit vectors with self-similar homogeneous distributions, with the corresponding finite set  $\mathcal{W}$  (cf. p. 5) given by a subset of the lattice of  $\mathbb{Z}$ -linear combinations of the monomials  $f_1(\mathbf{H}), f_2(\mathbf{H}), \dots$  in the channel coefficients. In order to guarantee the open set condition for the resulting distributions of the noiseless output and noiseless interference, we will need the following two results from Diophantine approximation theory.

**Lemma 13.** For  $\mathbf{A} \in \mathbb{R}^{m \times n}$  define  $w(\mathbf{A})$  to be the supremum of all  $w > 0$  such that for arbitrarily large  $Q > 0$  the set

$$\{u \in \mathbb{Z}^n \setminus \{0\} \mid |\mathbf{A}u|_{\mathbb{Z}} < Q^{-w}, \|u\|_{\infty} \leq Q\}$$

is non-empty. If  $g_1, \dots, g_k$  are  $k$  pairwise distinct monomials of positive degrees, then

$$w\left((g_1(a) \dots g_k(a))^T\right) = \frac{1}{k} \quad \text{and} \quad w((g_1(a) \dots g_k(a))) = k \quad (195)$$

for almost all  $a \in \mathbb{R}^{\ell}$ .

*Proof:* See [33, Thm. KM, p. 823].  $\blacksquare$

**Lemma 14.** Let  $\mathbf{A} \in \mathbb{R}^{m \times n}$ ,  $\kappa := 2^{1-m-n}((m+n)!)^2$ , and  $\alpha, \beta > 0$ . If the set

$$\left\{u \in \mathbb{Z}^m \setminus \{0\} \mid |\mathbf{A}^T u|_{\mathbb{Z}} < \frac{\kappa}{\alpha}, \|u\|_{\infty} \leq \beta\right\}$$

is empty, then for all  $z \in \mathbb{R}^m$ , the set

$$\left\{v \in \mathbb{Z}^n \setminus \{0\} \mid |\mathbf{A}v + z|_{\mathbb{Z}} \leq \frac{\kappa}{\beta}, \|v\|_{\infty} \leq \alpha\right\}$$

is non-empty.

*Proof:* See [34, Lem. 3, p. 756].  $\blacksquare$

<sup>16</sup>A “monomial”  $f$  in  $\ell$  variables is a function of the form  $f(x_1, \dots, x_{\ell}) = x_1^{d_1} x_2^{d_2} \dots x_{\ell}^{d_{\ell}}$  for non-negative integers  $d_1, \dots, d_{\ell}$ .

<sup>17</sup>The “degree” of a monomial is to be understood as the sum of all exponents of the variables involved (sometimes called the total degree).

We apply Lemma 13 to find that for almost all matrices  $\mathbf{H} \in \mathbb{R}^{KM \times KM}$

$$w\left((f_2(\mathbf{H}) \dots f_L(\mathbf{H}))^T\right) = \frac{1}{L-1} \quad \text{and} \\ w\left((f_2(\mathbf{H}) \dots f_{L'}(\mathbf{H}))\right) = L' - 1,$$

where  $w(\cdot)$  was defined in Lemma 13; note that we intentionally left out the constant monomial  $f_1$ . Therefore, for every  $s > 0$  and almost all  $\mathbf{H}$ , there is a  $Q_0$  such that for all  $Q \geq Q_0$ , we have

$$\{u \in \mathbb{Z} \setminus \{0\} \mid (f_2(\mathbf{H}) \dots f_L(\mathbf{H}))^T u|_{\mathbb{Z}} < Q^{-\frac{1}{L-1-s}}, \|u\|_{\infty} \leq Q\} = \emptyset \quad (196)$$

$$\{u \in \mathbb{Z}^{L'-1} \setminus \{0\} \mid (f_2(\mathbf{H}) \dots f_{L'}(\mathbf{H})) u|_{\mathbb{Z}} < Q^{-L'+1-s}, \|u\|_{\infty} \leq Q\} = \emptyset. \quad (197)$$

We choose  $s > 0$  to be small enough for

$$\left| \frac{L' - L + 2s}{L'} - \frac{L' - L}{L'} \right| < \varepsilon \quad (198)$$

to hold. Since (196) and (197) hold, individually, for almost all channel matrices  $\mathbf{H}$  it follows that they also hold simultaneously for almost all  $\mathbf{H}$ . In the remainder of the proof, we assume that  $\mathbf{H}$  satisfies both (196) and (197). By the scaling invariance in Lemma 3, we may furthermore assume that  $\|\mathbf{H}\|_{\infty} \leq 1$ . Next, for a given finite set  $\mathcal{A} \subseteq \mathbb{R}$ , we define the quantizer to  $\mathcal{A}$  as

$$\mathbf{Q}_{\mathcal{A}}(x) := \arg \min_{a \in \mathcal{A}} |x - a|,$$

with ties broken arbitrarily. For an arbitrary real constant  $T > 0$ , define the following sets

$$\mathcal{V} := \frac{1}{T} \left\{ \sum_{i=1}^L u_i f_i(\mathbf{H}) \mid u_1, \dots, u_L \in \mathbb{Z}, |u_1| \leq LT, |u_2|, \dots, |u_L| \leq T \right\} \subseteq [-2L, 2L], \quad (199)$$

$$\mathcal{V}' := \frac{1}{T'} \left\{ \sum_{i=1}^{L'} u_i f_i(\mathbf{H}) \mid u_1, \dots, u_{L'} \in \mathbb{Z}, |u_1|, \dots, |u_{L'}| \leq L'T' \right\} \subseteq [-L'L, L'L], \quad (200)$$

which will be used in our construction of the self-similar single-letter input distributions. The choice of  $\mathcal{V}$  and  $\mathcal{V}'$  is adapted to the channel in the sense that if  $x_1, \dots, x_K \in \mathcal{V}^M$  then  $\sum_{j=1}^K \mathbf{H}_{i,j} x_j \in \mathcal{V}'^M$ , since multiplication by a channel coefficient only increases the degrees of the involved monomials by 1.

We next discretize the inputs  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  by quantizing their entries to the set  $\mathcal{V}$  and characterize the resulting quantization error through (196) and Lemma 14. To this end, we set  $Q = \left(\frac{T}{\kappa}\right)^{\frac{1}{L-1-s}}$  and apply Lemma 14 with  $\mathbf{A} = (f_2(\mathbf{H}), \dots, f_L(\mathbf{H}))$ ,  $\alpha = T, \beta = Q, m = 1, n = L - 1$ ,

and  $\kappa = 2^{1-L}(L!)^2$ ; the condition of Lemma 14 is satisfied by (196) if we choose  $T$  large enough for  $Q \geq Q_0$  to hold. Lemma 14 then says that for every  $\theta \in [0, 1]$ , there exists a  $v = (v_2 \dots v_L)^T \in \mathbb{Z}^{L-1} \setminus \{0\}$  with  $\|v\|_\infty \leq T$  such that

$$|((f_2(\mathbf{H}), \dots, f_L(\mathbf{H}))v - T\theta)|_{\mathbb{Z}} \leq \frac{\kappa^{L-s}}{T^{L-1-s}}. \quad (201)$$

Since  $\|\mathbf{H}\|_\infty \leq 1$  we have  $f_i(\mathbf{H}) \leq 1$ , for all  $i$ , and thus  $|((f_2(\mathbf{H}), \dots, f_L(\mathbf{H}))v - T\theta)| \leq LT$ . Therefore, the integer that best approximates  $((f_2(\mathbf{H}), \dots, f_L(\mathbf{H}))v - T\theta)$  also has absolute value  $\leq LT$ . Denote this best approximating integer by  $v_1$ . Since  $f_1$  is the constant monomial, dividing both sides of (201) by  $T$  yields

$$\left| \frac{1}{T} \sum_{i=1}^L v_i f_i(\mathbf{H}) - \theta \right| \leq \left( \frac{\kappa}{T} \right)^{L-s}.$$

We thus have

$$\sup_{\theta \in [0,1]} |\mathcal{Q}_{\mathcal{V}}(\theta) - \theta| \leq \left( \frac{\kappa}{T} \right)^{L-s}. \quad (202)$$

Next, we lower-bound the minimum distance between elements of  $\mathcal{V}'$  using (197). Note that

$$\mathcal{V}' - \mathcal{V}' = \frac{1}{T} \left\{ \sum_{i=1}^{L'} u_i f_i(\mathbf{H}) \mid u_1, \dots, u_{L'} \in \mathbb{Z}, |u_1|, \dots, |u_{L'}| \leq 2LT \right\}.$$

We choose  $T$  large enough for  $2LT \geq Q_0$  to hold. By (197) we then get that for all  $u = (u_2 \dots u_{L'})^T \in \mathbb{Z}^{L'-1} \setminus \{0\}$  with  $\|u\|_\infty \leq 2LT$

$$|(f_2(\mathbf{H}), \dots, f_{L'}(\mathbf{H}))u|_{\mathbb{Z}} \geq (2LT)^{-L'+1-s}$$

must hold. Therefore, no matter how we choose the integer  $u_1$ , we have

$$\left| \sum_{i=1}^{L'} u_i f_i(\mathbf{H}) \right| \geq (2LT)^{-L'+1-s},$$

and thus

$$m(\mathcal{V}') \geq \frac{(2LT)^{-L'+1-s}}{T}. \quad (203)$$

Before getting into the juice of the argument, we need some preparatory material, which explains how to choose the parameter  $N_0$  (from (193)) sufficiently large to obtain bounds guaranteeing that the open set condition for the noiseless output distributions is satisfied. The reader can safely skip this part initially and consult it when needed later in the proof. First, note that the expression

$$\frac{1}{k} \log(c_1 + c_2 2^{c_3 k}) \quad (204)$$

for positive constants  $c_1, c_2, c_3$  converges to  $c_3$  as  $k \rightarrow \infty$ . We can therefore take  $N_0$  large enough such that for all  $k \geq N_0$  we have

$$\left| \frac{1}{k} \log\left(c_1 + c_2 2^{\frac{L'-L+2s}{L'} k}\right) - \frac{L'-L+2s}{L'} \right| < \varepsilon, \quad (205)$$

where  $c_1, c_2$  will be specified below and  $\varepsilon$  was chosen above. Finally, we set

$$T := 2^{\frac{k}{L'}} \quad (206)$$

and pick  $N_0$  large enough such that for all  $k \geq N_0$ , in addition to (205), we also have

$$2^{-k} \leq \frac{(2L)^{-L'+1-s} T^{-L'-s}}{16KML}. \quad (207)$$

We are now ready to proceed to the finale of the proof and single-letterize (193). First, we set  $\tilde{\mathbf{X}}_i^N := \mathcal{Q}_{\mathcal{V}}(\mathbf{X}_i^N)$ , for  $i = 1, \dots, K$ , where  $\mathcal{Q}_{\mathcal{V}}(\cdot)$  is applied entry-wise. Note that

$$\left\| \left[ \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k - \sum_{j=1}^K \mathbf{H}_{i,j} \tilde{\mathbf{X}}_j^N \right\|_\infty \quad (208)$$

$$\leq 2^{-k} + \left\| \sum_{j=1}^K \mathbf{H}_{i,j} (\mathbf{X}_j^N - \tilde{\mathbf{X}}_j^N) \right\|_\infty \quad (209)$$

$$\leq 2^{-k} + \sum_{j=1}^K \left\| \mathbf{H}_{i,j} (\mathbf{X}_j^N - \tilde{\mathbf{X}}_j^N) \right\|_\infty \quad (210)$$

$$\stackrel{(202)}{\leq} 2^{-k} + KM \left( \frac{\kappa}{T} \right)^{L-s}. \quad (211)$$

Applying Lemma 9 with  $\varepsilon = 2^{-k} + KM \left( \frac{\kappa}{T} \right)^{L-s}$  and  $\delta = 2^{-k}$  which is the minimum distance of the value set for  $\left[ \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k$ , we find

$$\frac{1}{Nk} \left( H \left( \left[ \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k \right) - H \left( \sum_{j=1}^K \mathbf{H}_{i,j} \tilde{\mathbf{X}}_j^N \right) \right) \quad (212)$$

$$\leq \frac{M}{k} \log \left( 1 + \frac{2 \left( 2^{-k} + KM \left( \frac{\kappa}{T} \right)^{L-s} \right)}{2^{-k}} \right) \quad (213)$$

$$= \frac{M}{k} \log(3 + 2KM\kappa^{L-s} 2^k T^{-L+s}) \quad (214)$$

$$\stackrel{(206)}{=} \frac{M}{k} \log(3 + 2KM\kappa^{L-s} 2^{(1-\frac{L-s}{L'})k}) \quad (215)$$

$$\leq \frac{M}{k} \log(3 + 2KM\kappa^{L-s} 2^{\frac{L'-L+2s}{L'} k}) \quad (216)$$

$$\stackrel{(205)}{\leq} M \left( \frac{L'-L+2s}{L'} + \varepsilon \right) \quad (217)$$

$$\stackrel{(198)}{\leq} M \left( \frac{L'-L}{L'} + 2\varepsilon \right) \quad (218)$$

$$\stackrel{(194)}{\leq} 3M\varepsilon. \quad (219)$$

Similarly, again using Lemma 9 with  $\varepsilon$  as before and  $\delta =$

$(2L)^{-L'+1-s}T^{-L'-s}$ , we obtain by (203) that

$$\frac{1}{Nk} \left( H \left( \sum_{j=1}^K \mathbf{H}_{i,j} \tilde{\mathbf{X}}_j^N \right) - H \left( \left[ \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k \right) \right) \quad (220)$$

$$\leq \frac{M}{k} \log \left( 1 + \frac{2 \left( 2^{-k} + KM \left( \frac{\kappa}{T} \right)^{L-s} \right)}{(2L)^{-L'+1-s}T^{-L'-s}} \right) \quad (221)$$

$$= \frac{M}{k} \log \left( 1 + 2(2L)^{L'-1+s}T^{L'+s} \times \left( 2^{-k} + KM \left( \frac{\kappa}{T} \right)^{L-s} \right) \right) \quad (222)$$

$$\stackrel{(206)}{=} \frac{M}{k} \log \left( 1 + 2(2L)^{L'-1+s}KM\kappa^{L-s} \times \left( 2^{\left( \frac{L'+s}{L'} - 1 \right)k} + 2^{\frac{L'-L+2s}{L'}k} \right) \right) \quad (223)$$

$$\leq \frac{M}{k} \log \left( 1 + 2(2L)^{L'-1+s}KM\kappa^{L-s} \times \left( 2^{\frac{L'-L+2s}{L'}k} + 2^{\frac{L'-L+2s}{L'}k} \right) \right) \quad (224)$$

$$\stackrel{(205)}{\leq} M \left( \frac{L' - L + 2s}{L'} + \varepsilon \right) \quad (225)$$

$$\stackrel{(198)}{\leq} M \left( \frac{L' - L}{L'} + 2\varepsilon \right) \quad (226)$$

$$\stackrel{(194)}{\leq} 3M\varepsilon, \quad (227)$$

where in (224) we used  $s/L' \leq (L' - L + 2s)/L'$ . In summary, we have shown that

$$\frac{1}{Nk} \left| H \left( \left[ \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k \right) - H \left( \sum_{j=1}^K \mathbf{H}_{i,j} \tilde{\mathbf{X}}_j^N \right) \right| \leq 3M\varepsilon. \quad (228)$$

We next let  $r = 2^{-k}$  and set

$$W_i := \sum_{n=1}^N r^{n-1} \tilde{X}_i^{(n)}, \quad i = 1, \dots, K, \quad (229)$$

where  $\tilde{\mathbf{X}}_i^N = (\tilde{X}_i^{(1)} \dots \tilde{X}_i^{(N)})$ . The single-letter transmit vectors are then constructed as

$$\tilde{X}_i := \sum_{\ell=0}^{\infty} r^{N\ell} W_{i,\ell}, \quad i = 1, \dots, K, \quad (230)$$

where  $\{W_{i,\ell}\}_{\ell \geq 0}$  are i.i.d. copies of the discrete random vector  $W_i$ . By (22) it follows that the  $\tilde{X}_i$  have self-similar homogeneous distributions with similarity ratio  $r^N$ . Note that although Proposition 5 requires the input distributions only to have a singular component, for simplicity, we take them to be purely singular. The essence of this construction is that the noiseless channel output vectors

$$\sum_{j=1}^K \mathbf{H}_{i,j} \tilde{X}_j = \sum_{\ell=0}^{\infty} r^{N\ell} \sum_{j=1}^K \mathbf{H}_{i,j} W_{j,\ell}, \quad i = 1, \dots, K,$$

again have self-similar homogeneous distributions with similarity ratio  $r^N$ . Note that  $W_i$  takes value in  $\mathcal{W} :=$

$\sum_{n=1}^N r^{n-1} \mathcal{V}^M$ . Since multiplication of an element in  $\mathcal{V}$  by a channel coefficient increases the degrees of the involved monomials by 1, we have  $\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M \subseteq \mathcal{V}^M$ . Therefore, it follows that  $m(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M) \geq m(\mathcal{V}^M) = m(\mathcal{V}')$ . Moreover, by (199) and  $\|\mathbf{H}\|_{\infty} \leq 1$  we have  $M(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M) \leq 4KML$ , and thus get

$$\begin{aligned} & \frac{m(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M)}{m(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M) + M(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M)} \\ & \geq \frac{m(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M)}{2M(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M)} \\ & \stackrel{(203)}{\geq} \frac{(2L)^{-L'+1-s}T^{-L'-s}}{8KML} \\ & \stackrel{(207)}{\geq} r. \end{aligned}$$

By Lemma 12 this implies

$$m \left( \sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W} \right) = m \left( \sum_{j=1}^K \mathbf{H}_{i,j} \sum_{n=1}^N r^{n-1} \mathcal{V}^M \right) \quad (231)$$

$$= m \left( \sum_{n=1}^N r^{n-1} \sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M \right) \quad (232)$$

$$\geq r^{N-1} m \left( \sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M \right). \quad (233)$$

For the maximum distance we have  $M(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W}) = M(\sum_{j=1}^K \mathbf{H}_{i,j} \sum_{n=1}^N r^{n-1} \mathcal{V}^M) \leq 8KML$ , by (199) and  $\sum_{n=1}^N r^{n-1} \leq 2$ . We obtain

$$\begin{aligned} r^{N(207)} & \leq r^{N-1} \frac{(2L)^{-L'+1-s}T^{-L'-s}}{16KML} \\ & \stackrel{(203)}{\leq} r^{N-1} \frac{m(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M)}{16KML} \\ & \stackrel{(233)}{\leq} \frac{m(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W})}{2M(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W})} \\ & \leq \frac{m(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W})}{m(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W}) + M(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W})}. \end{aligned}$$

It thus follows from Lemma 8 that the distribution of  $\sum_{j=1}^K \mathbf{H}_{i,j} \tilde{X}_j$  satisfies the open set condition. By Theorem 1, we then get

$$d \left( \sum_{j=1}^K \mathbf{H}_{i,j} \tilde{X}_j \right) = \frac{H(\sum_{j=1}^K \mathbf{H}_{i,j} W_j)}{N \log \frac{1}{r}} \quad (234)$$

$$= \frac{H(\sum_{j=1}^K \mathbf{H}_{i,j} W_j)}{Nk}, \quad (235)$$

and Lemma 12 yields

$$H \left( \sum_{j=1}^K \mathbf{H}_{i,j} W_j \right) = H \left( \sum_{j=1}^K \mathbf{H}_{i,j} \tilde{X}_j^N \right), \quad (236)$$



resulting in

$$d\left(\sum_{j=1}^K \mathbf{H}_{i,j} \tilde{X}_j\right) = H\left(\sum_{j=1}^K \mathbf{H}_{i,j} \tilde{X}_j^N\right). \quad (237)$$

We can apply the same arguments to the quantity  $\sum_{j \neq i}^K \mathbf{H}_{i,j} \mathbf{X}_j^N$  and use the result thereof along with (237) in the sum on the RHS of (193). Collecting the errors from (219) and (227), we finally get

DoF( $\mathbf{H}$ )

$$\begin{aligned} &\leq \varepsilon + 6MK\varepsilon + \sum_{i=1}^K \left[ d\left(\sum_{j=1}^K \mathbf{H}_{i,j} \tilde{X}_j\right) - d\left(\sum_{j \neq i}^K \mathbf{H}_{i,j} \tilde{X}_j\right) \right] \\ &= \varepsilon + 6MK\varepsilon + \text{dof}(\tilde{X}_1, \dots, \tilde{X}_K; \mathbf{H}). \end{aligned}$$

Since  $\varepsilon > 0$  can be chosen arbitrarily small, this completes the proof.  $\blacksquare$

#### D. Proof of Theorem 4

The proof is along the same lines as that of Theorem 3. We therefore detail only the steps that require modification. Specifically, we start from (193) which states that for every  $\varepsilon > 0$  and every  $N_0 > 0$ , there exist  $N, k \geq N_0$  and independent  $\mathbf{X}_1^N, \dots, \mathbf{X}_K^N$  that take value in  $[0, 1]^{M \times N}$  such that

$$\begin{aligned} \text{DoF}(\mathbf{H}) &\leq \varepsilon + \frac{1}{Nk} \\ &\sum_{i=1}^K \left[ H\left(\left[\sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N\right]_k\right) - H\left(\left[\sum_{j \neq i}^K \mathbf{H}_{i,j} \mathbf{X}_j^N\right]_k\right) \right]. \end{aligned} \quad (238)$$

Since we assume all entries of  $\mathbf{H}$  to be rational, we may use the scaling invariance of DoF( $\mathbf{H}$ ) formalized in Lemma 3 to argue that it suffices to consider  $\mathbf{H}$  with integer entries only. This drastically simplifies controlling the distance properties of the noiseless outputs. Specifically, for  $\mathbf{H}$  with all entries rational, we can employ codebooks built from integers, which allows us to control the distances of the noiseless outputs for *all* channel matrices  $\mathbf{H}$ . More concretely, we replace the sets  $\mathcal{V}$  and  $\mathcal{V}'$  in (199), (200) by

$$\mathcal{V} := 2^{-(k-p)} \{0, 1, \dots, 2^{k-p}\}, \quad (239)$$

$$\mathcal{V}' := 2^{-(k-p)} \{-KMH_{\max} 2^{k-p}, -KMH_{\max} 2^{k-p} + 1, \dots, KMH_{\max} 2^{k-p}\}, \quad (240)$$

for  $k \geq p$ , where  $H_{\max} := \|\mathbf{H}\|_{\infty}$  and  $p$  is a positive integer such that

$$2^{-p} \leq \frac{1}{8KMH_{\max}}. \quad (241)$$

Then, we have

$$\sup_{\theta \in [0,1]} |\mathcal{Q}_{\mathcal{V}}(\theta) - \theta| \leq 2^{-(k-p)}, \quad (242)$$

$$\mathfrak{m}(\mathcal{V}') \geq 2^{-(k-p)}, \quad (243)$$

$$\mathfrak{M}(\mathcal{V}') \leq 2KMH_{\max}. \quad (244)$$

Setting  $\tilde{\mathbf{X}}_i^N := \mathcal{Q}_{\mathcal{V}}(\mathbf{X}_i^N)$ , for  $i = 1, \dots, K$ , we get  $\left\| \left[ \sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N \right]_k - \sum_{j=1}^K \mathbf{H}_{i,j} \tilde{\mathbf{X}}_j^N \right\|_{\infty} \leq 2^{-k} + 2^{-(k-p)} KMH_{\max}$  by (242) (cf. (209)–(211)). Through application of Lemma 9 and using (243), we find that

$$\frac{1}{Nk} \left| H\left(\left[\sum_{j=1}^K \mathbf{H}_{i,j} \mathbf{X}_j^N\right]_k\right) - H\left(\sum_{j=1}^K \mathbf{H}_{i,j} \tilde{\mathbf{X}}_j^N\right) \right| \quad (245)$$

$$\leq \frac{M}{k} \log\left(1 + 2 \frac{2^{-k} + 2^{-(k-p)} KMH_{\max}}{2^{-k}}\right) \quad (246)$$

$$\leq \frac{M}{k} \log(3 + 2KMH_{\max} 2^p). \quad (247)$$

Choosing  $N_0$  sufficiently large, we ensure that (247) is as small as desired for all  $k \geq N_0$ . The construction of the single-letter transmit vectors now follows the exact same steps as in the proof of Theorem 3. The only element that differs from what was done there is the choice of the sets  $\mathcal{V}$  and  $\mathcal{V}'$  in (239), (240). We therefore only need to verify the open set condition for the iterated function system constructed according to (229) and (230) using  $\mathcal{V}$  and  $\mathcal{V}'$  in (239), (240). First, we note that since the entries of the matrices  $\mathbf{H}_{i,j}$  are integers, we have  $\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M \subseteq \mathcal{V}^M$  which implies  $\mathfrak{m}(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M) \geq \mathfrak{m}(\mathcal{V}^M) = \mathfrak{m}(\mathcal{V}')$  and  $\mathfrak{M}(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M) \leq \mathfrak{M}(\mathcal{V}^M) = \mathfrak{M}(\mathcal{V}')$ . We set  $r := 2^{-k}$ . Using (243) and (244) we find that

$$\frac{\mathfrak{m}(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M)}{\mathfrak{m}(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M) + \mathfrak{M}(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M)} \quad (248)$$

$$\geq \frac{\mathfrak{m}(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M)}{2\mathfrak{M}(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M)} \quad (249)$$

$$\geq \frac{\mathfrak{m}(\mathcal{V}')}{2\mathfrak{M}(\mathcal{V}')} \quad (250)$$

$$\stackrel{(243),(244)}{\geq} \frac{2^{-(k-p)}}{4KMH_{\max}} \quad (251)$$

$$\stackrel{(241)}{\geq} r. \quad (252)$$

It thus follows by Lemma 12 that for  $\mathcal{W} := \sum_{n=1}^N r^{n-1} \mathcal{V}^M$  (cf. (231)–(233))

$$\mathfrak{m}\left(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W}\right) \geq r^{N-1} \mathfrak{m}\left(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M\right). \quad (253)$$

Moreover note that  $\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W} \subseteq \sum_{n=1}^N r^{n-1} \mathcal{V}^M$  and therefore  $\mathfrak{M}(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W}) \leq 4KMH_{\max}$ , since

$\sum_{n=1}^N r^{n-1} \leq 2$ . This implies

$$r^N \stackrel{(241)}{\leq} r^{N-1} \frac{2^{-(k-p)}}{8KM H_{\max}} \quad (254)$$

$$\stackrel{(243)}{\leq} r^{N-1} \frac{m\left(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{V}^M\right)}{8KM H_{\max}} \quad (255)$$

$$\stackrel{(253)}{\leq} \frac{m\left(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W}\right)}{2M\left(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W}\right)} \quad (256)$$

$$\leq \frac{m\left(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W}\right)}{m\left(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W}\right) + M\left(\sum_{j=1}^K \mathbf{H}_{i,j} \mathcal{W}\right)}. \quad (257)$$

All the remaining steps follow exactly the proof of Theorem 3. ■

### E. Proof of Theorem 5

We first note that the limiting characterization in Proposition 3 holds for general noise distributions satisfying the assumptions of Theorem 5. Next, we conclude that Proposition 11 also extends to noise distributions that satisfy the assumptions of Theorem 5, as the only conditions on the noise distributions used in the proof of Proposition 11 are those needed by Lemma 7, which, in turn, precisely equal the assumptions in Theorem 5. This establishes the claim. ■

### REFERENCES

- [1] D. Stotz and H. Bölcskei, "Degrees of freedom in vector interference channels," *Proc. 50th Ann. Allerton Conf. on Communication, Control, and Computing*, pp. 1755–1760, Oct. 2012.
- [2] —, "ADDENDUM to "Degrees of freedom in vector interference channels"," available online: [http://www.nari.ee.ethz.ch/commth/research/downloads/dof\\_addendum.pdf](http://www.nari.ee.ethz.ch/commth/research/downloads/dof_addendum.pdf), Feb. 2016.
- [3] Y. Wu, S. Shamai (Shitz), and S. Verdú, "A formula for the degrees of freedom of the interference channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 256–279, Jan. 2015.
- [4] A. Guionnet and D. Shlyakhtenko, "On classical analogues of free entropy dimension," *Journal of Functional Analysis*, vol. 251, pp. 738–771, Oct. 2007.
- [5] I. Kontoyiannis and M. Madiman, "Sunset and inverse sunset inequalities for differential entropy and mutual information," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4503–4514, Aug. 2014.
- [6] 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.
- [7] M.-A. Maddah-Ali, A. S. Motahari, and A. K. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3454–3470, Aug. 2008.
- [8] A. S. Motahari, S. O. Gharan, M.-A. Maddah-Ali, and A. K. Khandani, "Real interference alignment: Exploiting the potential of single antenna systems," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4799–4810, Aug. 2014.
- [9] R. H. Etkin and E. Ordentlich, "The degrees-of-freedom of the K-user Gaussian interference channel is discontinuous at rational channel coefficients," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4932–4946, Nov. 2009.
- [10] V. R. Cadambe and S. A. Jafar, "Parallel Gaussian interference channels are not always separable," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3983–3990, Sep. 2009.
- [11] S. A. Jafar, "Interference alignment — A new look at signal dimensions in a communication network," *Foundations and Trends in Communications and Information Theory*, vol. 7, no. 1, 2011.
- [12] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Mathematica Hungarica*, vol. 10, pp. 1–23, Mar. 1959.
- [13] Y. Wu, "Shannon theory for compressed sensing," *Ph.D. dissertation, Princeton University*, Sep. 2011.
- [14] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Sig. Proc.*, vol. 5, pp. 964–967, Apr. 1980.
- [15] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, UK: Cambridge University Press, 2012.
- [16] R. Durrett, *Probability: Theory and Examples*. New York, NY: Cambridge University Press, 2010.
- [17] J. Taylor, *An Introduction to Measure and Probability*. New York, NY: Springer, 1997.
- [18] Y. Wu and S. Verdú, "Rényi information dimension: Fundamental limits of almost lossless analog compression," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3721–3748, Aug. 2010.
- [19] J. E. Hutchinson, "Fractals and self similarity," *Indiana University Mathematics Journal*, vol. 30, pp. 713–747, 1981.
- [20] C. Brandt, N. Viet Hung, and H. Rao, "On the open set condition for self-similar fractals," *Proc. of the American Mathematical Society*, vol. 134, no. 5, pp. 1369–1374, Oct. 2005.
- [21] J. S. Geronimo and D. P. Hardin, "An exact formula for the measure dimensions associated with a class of piecewise linear maps," *Constructive Approximation*, vol. 5, pp. 89–98, Dec. 1989.
- [22] L.-S. Young, "Dimension, entropy and Lyapunov exponents," *Ergod. Th. & Dynam. Sys.*, vol. 2, pp. 109–124, Mar. 1982.
- [23] D. Stotz and H. Bölcskei, "Explicit and almost sure conditions for  $K/2$  degrees of freedom," *Proc. IEEE Int. Symp. on Inf. Theory*, pp. 471–475, Jun. 2014.
- [24] R. Ahlswede, "Multi-way communication channels," *Proc. of the 2nd Int. Symp. on Inf. Theory*, pp. 23–52, Sep. 1971.
- [25] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289–1301, Mar. 2012.
- [26] G. Ungerböck, "Huffman shaping," in *Codes, Graphs, and Systems*, R. E. Blahut and R. Koetter, Eds. Kluwer Academic Publishers, Mar. 2002, pp. 299–313.
- [27] G. Durisi, V. I. Morgenshtern, H. Bölcskei, U. G. Schuster, and S. Shamai (Shitz), "Information theory of underspread WSSUS channels," in *Wireless Communications over Rapidly Time-Varying Channels*, F. Hlawatsch and G. Matz, Eds. Academic Press, 2011, pp. 65–116.
- [28] G. Bresler, D. Cartwright, and D. Tse, "Feasibility of interference alignment for the MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, pp. 5573–5586, Sep. 2014.
- [29] R. B. Ash, *Information Theory*, ser. Dover Books on Mathematics. Mineola, NY: Courier Dover Publications, 1990.
- [30] W. Rudin, *Real and complex analysis*, 3rd ed. McGraw-Hill, 1987.
- [31] S. Bobkov and G. Chistyakov, "Entropy power inequality for the Rényi entropy," *IEEE Trans. Inf. Theory*, vol. 61, no. 2, pp. 708–714, Feb. 2015.
- [32] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY: Wiley-Interscience, 2006.
- [33] V. Beresnevich and S. Velani, "An inhomogeneous transference principle and Diophantine approximation," *Proc. London Math. Soc.*, vol. 101, no. 3, pp. 821–851, Mar. 2010.
- [34] Y. Bugeaud and M. Laurent, "On exponents of homogeneous and inhomogeneous Diophantine approximation," *Moscow Math. Journal*, vol. 5, no. 4, pp. 747–766, Oct. 2005.