Characterizing Degrees of Freedom through Additive Combinatorics

David Stotz and Helmut Bölcskei

Dept. IT & EE, ETH Zurich, Switzerland
Email: {dstotz, boelcskei}@nari.ee.ethz.ch

Abstract

We establish a formal connection between the problem of characterizing degrees of freedom (DoF) in constant single-antenna interference channels (ICs), with general channel matrix, and the field of additive combinatorics. The theory we develop is based on a recent breakthrough result by Hochman in fractal geometry [2]. Our first main contribution is an explicit condition on the channel matrix to admit full, i.e., $K/2$ DoF; this condition is satisfied for almost all channel matrices. We also provide a construction of corresponding DoF-optimal input distributions. The second main result is a new DoF-formula exclusively in terms of Shannon entropies. This formula is more amenable to both analytical statements and numerical evaluations than the DoF-formula by Wu et al. [3], which is in terms of Rényi information dimension. We then use the new DoF-formula to shed light on the hardness of finding the exact number of DoF in ICs with rational channel coefficients, and to improve the best known bounds on the DoF of a well-studied channel matrix.

I. Introduction

A breakthrough finding in network information theory was the result that $K/2$ degrees of freedom (DoF) can be achieved in K-user single-antenna interference channels (ICs) [4], [5]. The corresponding transmit/receive scheme, known as interference alignment, exploits time-frequency selectivity of the channel to align interference at the receivers into low-dimensional subspaces.

Characterizing the DoF in ICs under various assumptions on the channel matrix has since become a heavily researched topic. A particularly surprising result states that $K/2$ DoF can be achieved in single-antenna K-user ICs with constant channel matrix [6], [7], i.e., in channels that do not exhibit

The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Honolulu, HI, June 2014 [1].

The authors would like to thank M. Einsiedler, ETH Zurich, for helpful discussions and for drawing their attention to [2].

any selectivity. This result was shown to hold for (Lebesgue) almost all[1] channel matrices [6, Thm. 1]. Instead of exploiting channel selectivity, here interference alignment happens on a number-theoretic level. The technical arguments—from Diophantine approximation theory—used in the proof of [6, Thm. 1] do not seem to allow an explicit characterization of the "almost-all set" of full-DoF admitting channel matrices. What is known, though, is that channel matrices with all entries rational admit strictly less than $K/2$ DoF [7] and hence belong to the set of exceptions relative to the "almost-all result" in [6].

Recently, Wu et al. [3] developed a general framework, based on (Rényi) information dimension, for characterizing the DoF in constant single-antenna ICs. While this general and elegant theory allows to recover, inter alia, the "almost-all result" from [6], it does not provide insights into the structure of the set of channel matrices admitting $K/2$ DoF. In addition, the DoF-formula in [3] is in terms of information dimension, which can be difficult to evaluate.

Contributions: Our first main contribution is to complement the results in [3], [6], [7] by providing *explicit and almost surely satisfied* conditions on the IC matrix to admit full, i.e., $K/2$ DoF. The conditions we find essentially require that the set of all monomial[2] expressions in the channel coefficients be linearly independent over the rational numbers. The proof of this result is based on a recent breakthrough in fractal geometry [2], which allows us to compute the information dimension of self-similar distributions under conditions much milder than the open set condition [8] required in [3]. For channel matrices satisfying our explicit and almost sure conditions, we furthermore present an explicit construction of DoF-optimal input distributions. The basic idea underlying this construction has roots in the field of additive combinatorics [9] and essentially ensures that the set-sum of signal and interference exhibits extremal cardinality properties. We also show that our sufficient conditions for $K/2$ DoF are not necessary. This is accomplished by constructing examples of channel matrices that admit $K/2$ DoF but do not satisfy the sufficient conditions we identify. The set of all such channel matrices, however, necessarily has Lebesgue measure zero.

Etkin and Ordentlich [7] discovered that tools from additive combinatorics can be applied to characterize DoF in ICs where the off-diagonal entries in the channel matrix are rational numbers and the diagonal entries are either irrational algebraic[3] or rational numbers. Our second main contribution is to establish a formal connection between additive combinatorics and the characterization of DoF in ICs with *arbitrary* channel matrices. Specifically, we show how the DoF-characterization in terms

[1]Throughout the paper "almost all" is to be understood with respect to Lebesgue measure and "almost sure" is with respect to a probability distribution that is absolutely continuous with respect to Lebesgue measure.

[2]A monomial in the variables $x_1, ..., x_n$ is an expression of the form $x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n}$, with $k_i \in \mathbb{N}$.

[3]A real number is called algebraic if it is the zero of a polynomial with integer coefficients. In particular, all rational numbers are algebraic.

of information dimension, discovered in [3], can be translated, again based on [2], into an alternative characterization exclusively involving Shannon entropies. The resulting new DoF-formula is more amenable to both analytical statements and numerical evaluation than the one in [3]. To support this statement, we show how the alternative DoF-formula can be used to explain why determining the exact number of DoF for channel matrices with rational entries, even for simple examples, has remained elusive so far. Specifically, we establish that DoF-characterization for rational channel matrices is equivalent to very hard open problems in additive combinatorics. Finally, we exemplify the quantitative applicability of the new DoF-formula by improving the best-known bounds on the DoF of a particular channel matrix studied in [3].

Notation: Random variables are represented by uppercase letters from the end of the alphabet. Lowercase letters are used exclusively for deterministic quantities. Boldface uppercase letters indicate matrices. Sets are denoted by uppercase calligraphic letters. For $x \in \mathbb{R}$, we write $\lfloor x \rfloor$ for the largest integer not exceeding x. All logarithms are taken to the base 2. $\mathbb{E}[\cdot]$ denotes the expectation operator. $H(\cdot)$ stands for entropy and $h(\cdot)$ for differential entropy. For a measurable real-valued function f and a measure[4] μ on its domain, the push-forward of μ by f is $(f_*\mu)(\mathcal{A}) = \mu(f^{-1}(\mathcal{A}))$ for Borel sets $\mathcal{A}$.

Outline of the paper: In Section II, we introduce the system model for constant single-antenna ICs. Section III contains our first main result, Theorem 1, providing explicit and almost surely satisfied conditions on channel matrices to admit full, i.e., $K/2$ DoF. In Section IV, we review the basic material on information dimension, self-similar distributions, and additive combinatorics needed in the paper. Section V is devoted to sketching the ideas underlying the proof of Theorem 1 in an informal fashion and to introducing the recent result by Hochman [2] that both our main results rely on. In Section VI, we formally prove Theorem 1. Section VII presents a non-asymptotic version of Theorem 1. In Section VIII, we establish that our sufficient conditions for $K/2$ DoF are not necessary. Our second main result, Theorem 3, which provides a DoF-characterization exclusively in terms of Shannon entropies, is presented, along with its proof, in Section IX. Finally, in Section X we discuss the formal connection between DoF and sumset theory, a branch of additive combinatorics, and we apply the new DoF-formula to channel matrices with rational entries.

II. System model

We consider a single-antenna K-user IC with constant channel matrix $\mathbf{H} = (h_{ij})_{1 \leqslant i,j \leqslant K} \in \mathbb{R}^{K \times K}$ and input-output relation

$$Y_i = \sqrt{\mathsf{snr}} \sum_{j=1}^{K} h_{ij} X_j + Z_i, \quad i = 1, ..., K, \tag{1}$$

[4]Throughout the paper, the terms "measurable" and "measure" are to be understood with respect to the Borel σ-algebra.

where $X_i \in \mathbb{R}$ is the input at the i-th transmitter, $Y_i \in \mathbb{R}$ is the output at the i-th receiver, and $Z_i \in \mathbb{R}$ is noise of absolutely continuous distribution such that $h(Z_i) > -\infty$ and $H(\lfloor Z_i \rfloor) < \infty$. The input signals are independent across transmitters and noise is i.i.d. across users and channel uses.

The channel matrix $\mathbf{H}$ is assumed to be known perfectly at all transmitters and receivers. We impose the average power constraint

$$\frac{1}{n}\sum_{k=1}^{n}\left(x_i^{(k)}\right)^2 \leqslant 1$$

on codewords $\left(x_i^{(1)} \, ... \, x_i^{(n)}\right)$ of block-length n transmitted by user $i = 1, ..., K$. The DoF of this channel are defined as

$$\mathsf{DoF}(\mathbf{H}) := \limsup_{\mathsf{snr}\to\infty} \frac{\overline{C}(\mathbf{H};\mathsf{snr})}{\frac{1}{2}\log\mathsf{snr}}, \tag{2}$$

where $\overline{C}(\mathbf{H};\mathsf{snr})$ is the sum-capacity of the IC.

III. Explicit and almost sure conditions for $K/2$ DoF

We denote the vector consisting of the off-diagonal entries of $\mathbf{H}$ by $\check{\mathbf{h}} \in \mathbb{R}^{K(K-1)}$, and let $f_1, f_2, ...$ be the monomials in $K(K-1)$ variables, i.e., $f_i(x_1, ..., x_{K(K-1)}) = x_1^{d_1} \cdots x_{K(K-1)}^{d_{K(K-1)}}$, enumerated as follows: $f_1, ..., f_{\varphi(d)}$ are the monomials of degree[5] not larger than d, where

$$\varphi(d) := \binom{K(K-1)+d}{d}.$$

The following theorem contains the first main result of the paper, namely conditions on $\mathbf{H}$ to admit $K/2$ DoF that are explicit and satisfied for almost all $\mathbf{H}$.

Theorem 1: Suppose that the channel matrix $\mathbf{H}$ satisfies the following condition:

For each $i = 1, ..., K$, the set

$$\{f_j(\check{\mathbf{h}}) \, : \, j \geqslant 1\} \cup \{h_{ii} f_j(\check{\mathbf{h}}) \, : \, j \geqslant 1\} \tag{$*$}$$

is linearly independent over $\mathbb{Q}$.

Then, we have

$$\mathsf{DoF}(\mathbf{H}) = K/2.$$

Proof: See Section VI. ■

We first note that, as detailed in the proof of Theorem 1, Condition $(*)$ implies that all entries of $\mathbf{H}$ must be nonzero, i.e., $\mathbf{H}$ must be fully connected in the terminology of [7]. By [10, Prop. 1] we have $\mathsf{DoF}(\mathbf{H}) \leqslant K/2$ for fully connected channel matrices. The proof of Theorem 1 is constructive in the sense of providing input distributions that achieve this upper bound.

[5]The "degree" of a monomial is defined as the sum of all exponents of the variables involved (sometimes called the total degree).

Let us next dissect Condition ($*$). A set $\mathcal{S} \subseteq \mathbb{R}$ is linearly independent over $\mathbb{Q}$ if, for all $n \in \mathbb{N}$ and all pairwise distinct $v_1, ..., v_n \in \mathcal{S}$, the only solution $q_1, ..., q_n \in \mathbb{Q}$ of the equation

$$q_1 v_1 + \ldots + q_n v_n = 0 \tag{3}$$

is $q_1 = \ldots = q_n = 0$. Thus, if Condition ($*$) is not satisfied, there exists, for at least one $i \in \{1, ..., K\}$, a non-trivial linear combination of a finite number of elements of the set

$$\{f_j(\check{\mathbf{h}}) : j \geqslant 1\} \cup \{h_{ii} f_j(\check{\mathbf{h}}) : j \geqslant 1\}$$

with rational coefficients which equals zero. In fact, this is equivalent to the existence of a non-trivial linear combination that equals zero and has all coefficients in $\mathbb{Z}$. This can be seen by simply multiplying (3) by a common denominator of $q_1, ..., q_n$.

To show that Condition ($*$) is satisfied for almost all channel matrices, we will argue that the condition is violated on a set of Lebesgue measure zero with respect to $\mathbf{H}$. To this end, we first note that for fixed $d \in \mathbb{N}$, fixed $a_1, ..., a_{\varphi(d)}, b_1, ..., b_{\varphi(d)} \in \mathbb{Z}$ not all equal to zero, and fixed $i \in \{1, ..., K\}$,

$$\sum_{j=1}^{\varphi(d)} a_j f_j(\check{\mathbf{h}}) + \sum_{j=1}^{\varphi(d)} b_j h_{ii} f_j(\check{\mathbf{h}}) = 0 \tag{4}$$

is satisfied only on a set of measure zero with respect to $\mathbf{H}$, as the solutions of (4) are given by the set of zeros of a polynomial in the channel coefficients. Since the set of equations (4) is countable with respect to $d \in \mathbb{N}$, $a_1, ..., a_{\varphi(d)}, b_1, ..., b_{\varphi(d)} \in \mathbb{Z}$, and $i \in \{1, ..., K\}$, the set of channel matrices violating Condition ($*$) is given by a countable union of sets of measure zero, which again has measure zero. It therefore follows that Condition ($*$) is satisfied for almost all channel matrices $\mathbf{H}$ and hence Theorem 1 provides conditions on $\mathbf{H}$ that not only guarantee that $K/2$ DoF can be achieved but are also explicit and almost surely satisfied.

We finally note that the prominent example from [7] with all entries of $\mathbf{H}$ rational, shown in [7] to admit strictly less than $K/2$ DoF, does not satisfy Condition ($*$), as two rational numbers are always linearly dependent over $\mathbb{Q}$.

IV. Preparatory Material

This section briefly reviews basic material on information dimension, self-similar distributions, and additive combinatorics needed in the rest of the paper.

A. Information dimension and DoF

Definition 1: Let X be a random variable with arbitrary distribution[6] μ. We define the lower and upper information dimension of X as

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

where $\langle X\rangle_k := \lfloor kX\rfloor/k$. If $\underline{d}(X) = \overline{d}(X)$, we set $d(X) := \underline{d}(X) = \overline{d}(X)$ and call $d(X)$ the information dimension of X. Since $\underline{d}(X), \overline{d}(X)$, and $d(X)$ depend on μ only, we sometimes also write $\underline{d}(\mu), \overline{d}(\mu)$, and $d(\mu)$, respectively.

The relevance of information dimension in characterizing DoF stems from the following relation [11], [3], [12]

$$\limsup_{\mathsf{snr}\to\infty} \frac{h(\sqrt{\mathsf{snr}}X + Z)}{\frac{1}{2}\log \mathsf{snr}} = \overline{d}(X), \tag{5}$$

which holds for arbitrary independent random variables X and Z, with the distribution of Z absolutely continuous and such that $h(Z) > -\infty$ and $H(\lfloor Z\rfloor) < \infty$.

We can apply (5) to ICs as follows. By standard random coding arguments we get that the sum-rate

$$I(X_1;Y_1) + \ldots + I(X_K;Y_K) \tag{6}$$

is achievable, where $X_1, ..., X_K$ are independent input distributions with $\mathbb{E}[X_i^2] \leqslant 1$, $i = 1, ..., K$. Using the chain rule, we obtain

$$I(X_i;Y_i) = h(Y_i) - h(Y_i \,|\, X_i) \tag{7}$$

$$= h\left(\sqrt{\mathsf{snr}}\sum_{j=1}^{K} h_{ij}X_j + Z_i\right) - h\left(\sqrt{\mathsf{snr}}\sum_{j\neq i}^{K} h_{ij}X_j + Z_i\right) \tag{8}$$

for $i = 1, ..., K$. Combining (5)-(8), it now follows that [3]

$$\mathsf{dof}(X_1, ..., X_K; \mathbf{H}) :=$$

$$\sum_{i=1}^{K}\left[d\left(\sum_{j=1}^{K} h_{ij}X_j\right) - d\left(\sum_{j\neq i}^{K} h_{ij}X_j\right)\right] \tag{9}$$

$$\leqslant \mathsf{DoF}(\mathbf{H}), \tag{10}$$

for all independent $X_1, ..., X_K$ with[7] $\mathbb{E}[X_i^2] < \infty$, $i = 1, ..., K$, and such that all information dimension terms appearing in (9) exist. A striking result in [3] shows that inputs of discrete, continuous, or mixed discrete-continuous distribution can achieve no more than 1 DoF irrespective of

[6]We consider general distributions which may be discrete, continuous, singular, or mixtures thereof.

[7]We only need the conditions $\mathbb{E}[X_i^2] < \infty$ as scaling of the inputs does not affect $\mathsf{dof}(X_1, ..., X_K; \mathbf{H})$.

K. For $K > 2$, input distributions achieving $K/2$ (i.e., full) DoF therefore necessarily have a singular component.

Taking the supremum in (10) over all admissible $X_1, \ldots, X_K$ yields

$$\mathsf{DoF}(\mathbf{H}) \geqslant \sup_{X_1,\ldots,X_K} \sum_{i=1}^{K} \left[d\left(\sum_{j=1}^{K} h_{ij} X_j \right) - d\left(\sum_{j \neq i}^{K} h_{ij} X_j \right) \right]. \tag{11}$$

It was furthermore discovered in [3] that equality in (11) holds for almost all channel matrices $\mathbf{H}$; an explicit characterization of this "almost-all set", however, does not seem to be available. The right-hand side (RHS) of (11) can be difficult to evaluate as explicit expressions for information dimension are available only for a few classes of distributions such as mixed discrete-continuous distributions or (singular) self-similar distributions reviewed in the next section.

B. *Self-similar distributions and iterated function systems*

A class of singular distributions with explicit expressions for their information dimension is given by self-similar distributions [13]. What is more, self-similar input distributions can be constructed to retain self-similarity under linear combinations, thereby allowing us to get explicit expressions for the information dimension of the output distributions in (9). For an excellent in-depth treatment of the material reviewed in this section, the interested reader is referred to [14].

We proceed to the definition of self-similar distributions. Consider a finite set $\Phi_r := \{\varphi_{i,r} : i = 1, \ldots, n\}$ of affine contractions $\varphi_{i,r} \colon \mathbb{R} \to \mathbb{R}$, i.e.,

$$\varphi_{i,r}(x) = rx + w_i, \tag{12}$$

where $r \in I \subseteq (0,1)$ and the w_i are pairwise distinct real numbers. We furthermore set $\mathcal{W} := \{w_1, \ldots, w_n\}$. Φ_r is called an iterated function system (IFS) parametrized by the contraction parameter $r \in I$. By classical fractal geometry [14, Ch. 9] every IFS has an associated unique attractor, i.e., a non-empty compact set $\mathcal{A} \subseteq \mathbb{R}$ such that

$$\mathcal{A} = \bigcup_{i=1}^{n} \varphi_{i,r}(\mathcal{A}). \tag{13}$$

Moreover, for each probability vector $(p_1, \ldots, p_n)$, there is a unique (Borel) probability distribution μ_r on $\mathbb{R}$ such that

$$\mu_r = \sum_{i=1}^{n} p_i (\varphi_{i,r})_* \mu_r, \tag{14}$$

where $(\varphi_{i,r})_* \mu_r$ is the push-forward of μ_r by $\varphi_{i,r}$. The distribution μ_r is supported on the attractor set $\mathcal{A}$ in (13) and is referred to as the self-similar distribution corresponding to the IFS Φ_r with underlying probability vector $(p_1, ..., p_n)$. We can give the following explicit expression for a random variable X with distribution μ_r as in (14)

$$X = \sum_{k=0}^{\infty} r^k W_k, \tag{15}$$

where $\{W_k\}_{k \geqslant 0}$ is a set of i.i.d. copies of a random variable W drawn from the set $\mathcal{W}$ according to $(p_1, ..., p_n)$.

C. A glimpse of additive combinatorics

The common theme of our two main results is a formal relationship between the study of DoF in constant single-antenna ICs and the field of additive combinatorics. This connection is enabled by the recent breakthrough result in fractal geometry reported in [2] and summarized in Section V. We next briefly discuss material from additive combinatorics that is relevant for our discussion. For a detailed treatment of additive combinatorics we refer the reader to [9]. Specifically, we will be concerned with sumset theory, which studies, for discrete sets $\mathcal{U}$, $\mathcal{V}$, the cardinality of the sumset $\mathcal{U} + \mathcal{V} = \{u + v \,:\, u \in \mathcal{U}, v \in \mathcal{V}\}$ relative to $|\mathcal{U}|$ and $|\mathcal{V}|$. We begin by noting the trivial bounds

$$\max\{|\mathcal{U}|, |\mathcal{V}|\} \leqslant |\mathcal{U} + \mathcal{V}| \leqslant |\mathcal{U}| \cdot |\mathcal{V}|, \tag{16}$$

for $\mathcal{U}$ and $\mathcal{V}$ finite and non-empty. One of the central ideas in sumset theory says that the left-hand inequality in (16) can be close to equality only if $\mathcal{U}$ and $\mathcal{V}$ have a common algebraic structure (e.g., lattice structures), whereas the right-hand inequality in (16) will be close to equality only if the pairs $\mathcal{U}$ and $\mathcal{V}$ do not have a common algebraic structure, i.e., they are generic relative to each other. Figure 1 illustrates this statement. Algebraic structures relevant in this context are arithmetic progressions, which are sets of the form $\mathcal{S} = \{a, a + d, a + 2d, \dots, a + (n - 1)d\}$ with $a \in \mathbb{Z}$ and $d \in \mathbb{N}$. If $\mathcal{U}$ and $\mathcal{V}$ are finite non-empty subsets of $\mathbb{Z}$, an improvement of the lower bound in (16) to $|\mathcal{U}| + |\mathcal{V}| - 1 \leqslant |\mathcal{U} + \mathcal{V}|$ can be obtained. This lower bound is attained if and only if $\mathcal{U}$ and $\mathcal{V}$ are arithmetic progressions of the same step size d [9, Prop. 5.8].

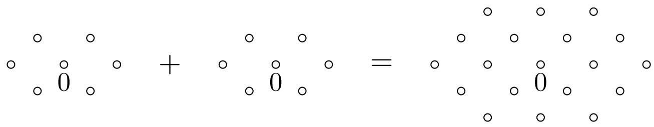
An interesting connection between sumset theory and entropy inequalities was discovered in [15], [16]. This connection revolves around the fact that many sumset inequalities have analogous versions in terms of entropy inequalities. For example, the entropy version of the trivial bounds (16) is

$$\max\{H(U), H(V)\} \leqslant H(U + V) \leqslant H(U) + H(V),$$

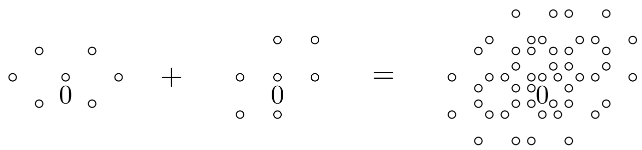
where U and V are independent discrete random variables. Less trivial examples are the sumset inequalities [9], [17]

$$|\mathcal{U} + \mathcal{V}| \cdot |\mathcal{U}| \cdot |\mathcal{V}| \leqslant |\mathcal{U} - \mathcal{V}|^3$$

$$|\mathcal{U} - \mathcal{V}| \leqslant |\mathcal{U} + \mathcal{V}|^{1/2} \cdot (|\mathcal{U}| \cdot |\mathcal{V}|)^{2/3},$$



(a) Sum of two sets with common algebraic structure.



(b) Sum of two sets with different algebraic structures.

Fig. 1: The cardinality of the sum in (a) is 19 and hence small compared to the $7^2 = 49$ pairs summed up, whereas the sum in (b) has cardinality 49.

for finite non-empty sets $\mathcal{U}, \mathcal{V}$, with their entropy counterparts [15], [16]

$$H(U+V) + H(U) + H(V) \leqslant 3H(U-V) \tag{17}$$

$$H(U-V) \leqslant \frac{1}{2}H(U+V) + \frac{2}{3}(H(U)+H(V)) \tag{18}$$

for independent discrete random variables U, V. Note that due to the logarithmic scale of entropy, products in sumset inequalities are replaced by sums in their entropy versions.

V. The cornerstones of the proof of Theorem 1

In this section, we discuss the main ideas and conceptual components underlying the proof of Theorem 1. First, we note that, as already pointed out in Section III, by [10, Prop. 1] we have $\mathsf{DoF}(\mathbf{H}) \leqslant K/2$ for all $\mathbf{H}$ satisfying Condition $(*)$. To achieve this upper bound, we construct self-similar input distributions that yield $\mathsf{dof}(X_1, ..., X_K; \mathbf{H}) = K/2$ for channel matrices satisfying Condition $(*)$. Specifically, we take each input to have a self-similar distribution with contraction parameter r, i.e., $X_i = \sum_{k=0}^{\infty} r^k W_{i,k}$, where, for $i = 1, ..., K$, $\{W_{i,k} : k \geqslant 0\}$ are i.i.d. copies of a discrete random variable[8] W_i with value set $\mathcal{W}_i$, possibly different across i. For the random variables $\sum_j h_{ij} X_j$ appearing in (11) we then have

$$\sum_j h_{ij} X_j = \sum_j \sum_{k=0}^{\infty} r^k h_{ij} W_{j,k} = \sum_{k=0}^{\infty} r^k \sum_j h_{ij} W_{j,k}, \tag{19}$$

and thus $\sum_j h_{ij} X_j$ is again self-similar with contraction parameter r. The "output-$\mathcal{W}$" set, i.e., the value set of $\sum_j h_{ij} W_j$ is then given by $\sum_j h_{ij} \mathcal{W}_j$.

[8]Henceforth "discrete random variable" refers to a random variable that only takes finitely many values.

Next, we discuss conditions on X_j and h_{ij} under which analytical expressions for the information dimension of $\sum_j h_{ij}X_j$ can be given. For general self-similar distributions arising from iterated function systems classical results in fractal geometry impose the so-called open set condition [18, Thm. 2], which requires the existence of a non-empty bounded set $\mathcal{U} \subseteq \mathbb{R}$ such that

$$\bigcup_{i=1}^{n} \varphi_{i,r}(\mathcal{U}) \subseteq \mathcal{U} \tag{20}$$

$$\text{and} \qquad \varphi_{i,r}(\mathcal{U}) \cap \varphi_{j,r}(\mathcal{U}) = \emptyset, \quad \text{for all } i \neq j, \tag{21}$$

for the $\varphi_{i,r}$ defined in (12). Wu et al. [3] ensure that the open set condition is satisfied by imposing an upper bound on the contraction parameter r according to

$$r \leqslant \frac{\mathsf{m}(\mathcal{W})}{\mathsf{m}(\mathcal{W}) + \mathsf{M}(\mathcal{W})}, \tag{22}$$

where $\mathsf{m}(\mathcal{W}) := \min_{i\neq j} |w_i - w_j|$ and $\mathsf{M}(\mathcal{W}) := \max_{i,j} |w_i - w_j|$. The challenge here resides in making (22) hold for the output-$\mathcal{W}$ set. In [3] this is accomplished by building the input sets $\mathcal{W}_i$ from $\mathbb{Z}$-linear combinations (i.e., linear combinations with integer coefficients) of monomials in the off-diagonal channel coefficients and then recognizing that results in Diophantine approximation theory can be used to show that (22) is satisfied for almost all channel matrices. Unfortunately, it does not seem to be possible to obtain an explicit characterization of this "almost-all set". Recent groundbreaking work by Hochman [2] replaces the open set condition by a much weaker condition, which instead of (20), (21) only requires that the IFS must not allow "exact overlap" of the images $\varphi_{i,r}(\mathcal{A})$ and $\varphi_{j,r}(\mathcal{A})$, for $i \neq j$, which we show in Theorem 2 below can be satisfied by "wiggling" with r in an arbitrarily small neighborhood of its original value. This improvement turns out to be instrumental in our Theorem 1 as it allows us to abandon the Diophantine approximation approach and thereby opens the doors to an explicit characterization of an "almost-all set" of full-DoF admitting channel matrices. Specifically, we use the following simple consequence of [2, Thm. 1.8].

Theorem 2: If $I \subseteq (0,1)$ is a non-empty compact interval which does not consist of a single point only, and μ_r is the self-similar distribution from (14) with contraction parameter $r \in I$ and probability vector $(p_1, ..., p_n)$, then[9]

$$d(\mu_r) = \min\left\{\frac{\sum p_i \log p_i}{\log r}, 1\right\}, \tag{23}$$

for all $r \in I \backslash E$, where E is a set of Hausdorff and packing dimension zero.

Proof: For $\mathbf{i} \in \{1, ..., n\}^k$, let $\varphi_{\mathbf{i},r} := \varphi_{i_1,r} \circ \ldots \circ \varphi_{i_k,r}$ and define

$$\Delta_{\mathbf{i},\mathbf{j}}(r) := \varphi_{\mathbf{i},r}(0) - \varphi_{\mathbf{j},r}(0),$$

[9] The "1" in the minimum simply accounts for the fact that information dimension cannot exceed the dimension of the ambient space.

for $\mathbf{i}, \mathbf{j} \in \{1, ..., n\}^k$. Extend this definition to infinite sequences $\mathbf{i}, \mathbf{j} \in \{1, ..., n\}^{\mathbb{N}}$ according to

$$\Delta_{\mathbf{i},\mathbf{j}}(r) := \lim_{k\to\infty} \Delta_{(i_1,...,i_k),(j_1,...,j_k)}(r).$$

Using (12) it follows that

$$\Delta_{\mathbf{i},\mathbf{j}}(r) = \sum_{k=1}^{\infty} r^{k-1}(w_{i_k} - w_{j_k}).$$

Since a power series can vanish on a non-empty open set only if it is identically zero, we get that $\Delta_{\mathbf{i},\mathbf{j}} \equiv 0$ on I if and only if $\mathbf{i} = \mathbf{j}$, as a consequence of the w_i being pairwise distinct and I containing a non-empty open set. This is precisely the condition of [2, Thm. 1.8] which asserts that (23) holds for all $r \in I$ with the exception of a set of Hausdorff and packing dimension zero, and thus completes the proof. ∎

Remark 1: Note that (23) can be rewritten in terms of the entropy of the random variable W, defined in (15), which takes value w_i with probability p_i:

$$d(\mu_r) = \min\left\{\frac{H(W)}{\log(1/r)}, 1\right\}. \tag{24}$$

Remark 2: The concepts of Hausdorff and packing dimension have their roots in fractal geometry [14]. In the proofs of our main results, we will only need the following aspect: For I as in Theorem 2, we can always find an $\widetilde{r} \in I \backslash E$ for which (23) holds. This can be seen as follows: $I \backslash E = \emptyset$ implies that E contains a non-empty open set and therefore would have Hausdorff and packing dimension 1 [14, Sec. 2.2].

Remark 3: The strength of Theorem 2 stems from (23) holding without any restrictions on the $w_i \in \mathcal{W}$. In particular, the elements in the output-$\mathcal{W}$ set $\sum_j h_{ij}\mathcal{W}_j$ may be arbitrarily close to each other rendering (22), needed to satisfy the open set condition, obsolete.

We next show how Theorem 2 allows us to derive explicit expressions for the information dimension terms in (9).

Proposition 1: Let $r \in (0, 1)$ and let $W_1, ..., W_K$ be independent discrete random variables. Then, we have

$$\sum_{i=1}^{K}\left[\min\left\{\frac{H\left(\sum_{j=1}^{K} h_{ij}W_j\right)}{\log(1/r)}, 1\right\} - \min\left\{\frac{H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)}{\log(1/r)}, 1\right\}\right] \leqslant \mathsf{DoF}(\mathbf{H}). \tag{25}$$

Proof: For $i = 1, ..., K$, let $\{W_{i,k} : k \geqslant 0\}$ be i.i.d. copies of W_i. We consider the self-similar inputs $X_i = \sum_{k=0}^{\infty} r^k W_{i,k}$, for $i = 1, ..., K$. Then, the signals

$$\sum_{j=1}^{K} h_{ij}X_j = \sum_{k=0}^{\infty} r^k \sum_{j=1}^{K} h_{ij}W_{j,k}$$

$$\text{and} \quad \sum_{j\neq i}^{K} h_{ij}X_j = \sum_{k=0}^{\infty} r^k \sum_{j\neq i}^{K} h_{ij}W_{j,k}$$

also have self-similar distributions with contraction parameter r. Thus, by Theorem 2, for each $\varepsilon > 0$, there exists an $\widetilde{r}$ in the non-empty compact interval $I_\varepsilon := [r - \varepsilon, r]$ (which does not consist of a single point only for all $\varepsilon > 0$) such that

$$d\left(\sum_{j=1}^{K} h_{ij} X_j\right) = \min\left\{\frac{H\left(\sum_{j=1}^{K} h_{ij} W_j\right)}{\log(1/\widetilde{r})}, 1\right\} \tag{26}$$

$$\text{and} \quad d\left(\sum_{j\neq i}^{K} h_{ij} X_j\right) = \min\left\{\frac{H\left(\sum_{j\neq i}^{K} h_{ij} W_j\right)}{\log(1/\widetilde{r})}, 1\right\}. \tag{27}$$

For $\varepsilon \to 0$ we have $\log(1/\widetilde{r}) \to \log(1/r)$ by continuity of $\log(\cdot)$. Thus, inserting (26) and (27) into (10) and letting $\varepsilon \to 0$, we get (25) as desired. ■

The freedom we exploit in constructing full DoF-achieving X_i lies in the choice of $W_1, ..., W_K$ which thanks to Theorem 2, unlike in [3], is not restricted by distance constraints on the output-$\mathcal{W}$ set. For simplicity of exposition, we henceforth choose the same value set $\mathcal{W}$ for each W_i. We want to ensure that the first term inside the sum (9) equals 1 and the second term equals $1/2$, for all i, resulting in a total of $K/2$ DoF. It follows from (26), (27) that this can be accomplished by choosing the W_i such that

$$H\left(h_{ii} W_i + \sum_{j\neq i}^{K} h_{ij} W_j\right) \approx 2H\left(\sum_{j\neq i}^{K} h_{ij} W_j\right) \tag{28}$$

followed by a suitable choice of the contraction parameter. Resorting to the analogy of entropy and sumset cardinalities sketched in Section IV-C, the doubling condition (28) becomes

$$\left|h_{ii}\mathcal{W} + \sum_{j\neq i}^{K} h_{ij}\mathcal{W}\right| \approx \left|\sum_{j\neq i}^{K} h_{ij}\mathcal{W}\right|^2, \tag{29}$$

which effectively says that the sum of the desired signal and the interference should be twice as "rich" as the interference alone. Note that by the trivial lower bound in (16)

$$|h_{ii}\mathcal{W}| = |\mathcal{W}| \leqslant \left|\sum_{j\neq i}^{K} h_{ij}\mathcal{W}\right|, \tag{30}$$

and, by the trivial upper bound in (16)

$$\left|h_{ii}\mathcal{W} + \sum_{j\neq i}^{K} h_{ij}\mathcal{W}\right| \leqslant |h_{ii}\mathcal{W}| \cdot \left|\sum_{j\neq i}^{K} h_{ij}\mathcal{W}\right|. \tag{31}$$

The doubling condition (29) can therefore be realized by constructing $\mathcal{W}$ such that the inequalities (30) and (31) are close to equality. In particular, this means that (cf. Section IV-C)

A) the terms in the sum $\sum_{j\neq i}^{K} h_{ij}\mathcal{W}$ must have a common algebraic structure and

B) $h_{ii}\mathcal{W}$ and $\sum_{j\neq i}^{K} h_{ij}\mathcal{W}$ must *not* have a common algebraic structure.

The challenge here is to introduce algebraic structure into $\mathcal{W}$ so that A) is satisfied but at the same time to keep the algebraic structures of the sets $h_{ii}\mathcal{W}$ and $\sum_{j\neq i}^{K} h_{ij}\mathcal{W}$ different enough so that B) is met. Before describing the specific construction of $\mathcal{W}$, we note that the answer to the question of whether the sets $h_{ij}\mathcal{W}$ have a common algebraic structure or not depends on the channel coefficients h_{ij}. As we want our construction to be universal in the sense of (29) holding independently of the channel coefficients, a channel-independent choice of $\mathcal{W}$ is out of the question. Inspired by [6], we build $\mathcal{W}$ as a set of $\mathbb{Z}$-linear combinations of monomials (up to a certain degree $d \in \mathbb{N}$) in the off-diagonal channel coefficients, i.e., the elements of $\mathcal{W}$ are given by $\sum_{j=1}^{\varphi(d)} a_j f_j(\check{\mathbf{h}})$, for $a_j \in \{1, ..., N\}$ with $N \in \mathbb{N}$. This construction satisfies A) by inducing the same algebraic structure for $h_{ij}\mathcal{W}$, $j \neq i$, independently of the actual values of the channel coefficients h_{ij}, $j \neq i$. To see this, first note that multiplying the elements $\sum_{j=1}^{\varphi(d)} a_j f_j(\check{\mathbf{h}})$ of $\mathcal{W}$ by an off-diagonal channel coefficient h_{ij}, $j \neq i$, simply increases the degrees of the participating $f_j(\check{\mathbf{h}})$ by 1. For d sufficiently large the number of elements that do not appear both in $h_{ij}\mathcal{W}$ and $\mathcal{W}$ is therefore small, rendering $h_{ij}\mathcal{W}$, $j \neq i$, algebraically "similar" to $\mathcal{W}$, which we denote as $h_{ij}\mathcal{W} \approx \mathcal{W}$. We therefore get $\sum_{j \neq i} h_{ij}\mathcal{W} \approx \mathcal{W} + \ldots + \mathcal{W}$ as the sum of $K-1$ sets with shared algebraic structure and note that the elements of $\mathcal{W} + \ldots + \mathcal{W}$ are given by $\sum_{j=1}^{\varphi(d)} a_j f_j(\check{\mathbf{h}})$ with $a_j \in \{1, ..., (K-1)N\}$. Choosing N to be large relative to K, we finally get $|\sum_{j \neq i} h_{ij}\mathcal{W}| \approx |\mathcal{W}|$. As for Condition B), we begin by noting that h_{ii} does not participate in the monomials $f_j(\check{\mathbf{h}})$ used to construct the elements in $\mathcal{W}$. This means that $\sum_{j \neq i}^{K} h_{ij}\mathcal{W}$ consists of $\mathbb{Z}$-linear combinations of $f_j(\check{\mathbf{h}})$, while $h_{ii}\mathcal{W}$ consists of $\mathbb{Z}$-linear combinations of $h_{ii} f_j(\check{\mathbf{h}})$. By Condition $(*)$ the union of the sets $\{f_j(\check{\mathbf{h}}) : j \geqslant 1\}$ and $\{h_{ii} f_j(\check{\mathbf{h}}) : j \geqslant 1\}$ is linearly independent over $\mathbb{Q}$, which ensures that $h_{ii}\mathcal{W}$ and $\sum_{j \neq i}^{K} h_{ij}\mathcal{W}$ do not share an algebraic structure.

VI. Proof of Theorem 1

Since a set containing 0 is always linearly dependent over $\mathbb{Q}$, Condition $(*)$ implies that all entries of $\mathbf{H}$ must be nonzero, i.e., $\mathbf{H}$ must be fully connected. It therefore follows from [10, Prop. 1] that $\mathsf{DoF}(\mathbf{H}) \leqslant K/2$.

The remainder of the proof establishes the lower bound $\mathsf{DoF}(\mathbf{H}) \geqslant K/2$ under Condition $(*)$. Let N and d be positive integers. We begin by setting

$$\mathcal{W}_N := \left\{ \sum_{i=1}^{\varphi(d)} a_i f_i(\check{\mathbf{h}}) \,:\, a_1, ..., a_{\varphi(d)} \in \{1, ..., N\} \right\} \tag{32}$$

and $r := |\mathcal{W}_N|^{-2}$. Let $W_1, ..., W_K$ be i.i.d. uniform random variables on $\mathcal{W}_N$. By Proposition 1 we then have

$$\sum_{i=1}^{K} \left[\min\left\{ \frac{H\left(\sum_{j=1}^{K} h_{ij} W_j\right)}{2 \log |\mathcal{W}_N|}, 1 \right\} - \min\left\{ \frac{H\left(\sum_{j \neq i}^{K} h_{ij} W_j\right)}{2 \log |\mathcal{W}_N|}, 1 \right\} \right] \leqslant \mathsf{DoF}(\mathbf{H}). \tag{33}$$

Note that the random variable $\sum_{j\neq i} h_{ij}W_j$ takes value in

$$\left\{\sum_{i=1}^{\varphi(d+1)} a_i f_i(\check{\mathbf{h}}) \,:\, a_1,...,a_{\varphi(d+1)} \in \{1,...,(K-1)N\}\right\}. \tag{34}$$

By Condition $(*)$ the set $\{f_j(\check{\mathbf{h}}) \,:\, j \geqslant 1\}$ is linearly independent over $\mathbb{Q}$. Therefore, each element in the set (34) has exactly one representation as a $\mathbb{Z}$-linear combination with coefficients $a_1,...,a_{\varphi(d+1)} \in \{1,...,(K-1)N\}$. This allows us to conclude that the cardinality of the set (34) is given by $((K-1)N)^{\varphi(d+1)}$, which implies $H\left(\sum_{j\neq i} h_{ij}W_j\right) \leqslant \varphi(d+1)\log((K-1)N)$. Similarly, we find that $|\mathcal{W}_N| = N^{\varphi(d)}$ and thus get

$$\frac{H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)}{2\log|\mathcal{W}_N|} \leqslant \frac{\varphi(d+1)\log((K-1)N)}{2\varphi(d)\log N} \tag{35}$$

$$\xrightarrow{d,N\to\infty} \frac{1}{2}, \tag{36}$$

where we used

$$\frac{\varphi(d+1)}{\varphi(d)} = \frac{K(K-1)+d+1}{d+1} \xrightarrow{d\to\infty} 1. \tag{37}$$

We next show that Condition $(*)$ implies that

$$H\left(h_{ii}W_i + \sum_{j\neq i} h_{ij}W_j\right) = H\left(h_{ii}W_i, \sum_{j\neq i} h_{ij}W_j\right). \tag{38}$$

Applying the chain rule twice we find

$$H\left(h_{ii}W_i, \sum_{j\neq i} h_{ij}W_j\right) = H\left(h_{ii}W_i, \sum_{j\neq i} h_{ij}W_j, h_{ii}W_i + \sum_{j\neq i} h_{ij}W_j\right) \tag{39}$$

$$= H\left(h_{ii}W_i + \sum_{j\neq i} h_{ij}W_j\right) + H\left(h_{ii}W_i, \sum_{j\neq i} h_{ij}W_j \,\middle|\, h_{ii}W_i + \sum_{j\neq i} h_{ij}W_j\right), \tag{40}$$

and therefore proving (38) amounts to showing that

$$H\left(h_{ii}W_i, \sum_{j\neq i} h_{ij}W_j \,\middle|\, h_{ii}W_i + \sum_{j\neq i} h_{ij}W_j\right) = 0. \tag{41}$$

In order to establish (41), suppose that $w_1,...,w_K$ and $\widetilde{w}_1,...,\widetilde{w}_K$ are realizations of $W_1,...,W_K$ such that

$$h_{ii}w_i + \sum_{j\neq i} h_{ij}w_j = h_{ii}\widetilde{w}_i + \sum_{j\neq i} h_{ij}\widetilde{w}_j, \tag{42}$$

or equivalently

$$h_{ii}(w_i - \widetilde{w}_i) + \sum_{j\neq i} h_{ij}(w_j - \widetilde{w}_j) = 0. \tag{43}$$

The first term on the left-hand side (LHS) of (43) is a $\mathbb{Z}$-linear combination of elements in $\{h_{ii}f_j(\check{\mathbf{h}}) \,:\, j \geqslant 1\}$, whereas the second term is a $\mathbb{Z}$-linear combination of elements in $\{f_j(\check{\mathbf{h}}) \,:\, j \geqslant 1\}$. Thanks to the linear independence of the union in Condition ($*$), it follows that the two terms in (43) have to equal zero individually and hence $w_i = \widetilde{w}_i$ and $\sum_{j\neq i} h_{ij} w_j = \sum_{j\neq i} h_{ij}\widetilde{w}_j$. This shows that the sum $h_{ii}W_i + \sum_{j\neq i} h_{ij}W_j$ uniquely determines the terms $h_{ii}W_i$ and $\sum_{j\neq i} h_{ij}W_j$ and therefore proves (41). Next, we note that

$$H\left(\sum_{j=1}^{K} h_{ij}W_j\right) = H\left(h_{ii}W_i + \sum_{j\neq i}^{K} h_{ij}W_j\right) \tag{44}$$

$$= H\left(h_{ii}W_i, \sum_{j\neq i}^{K} h_{ij}W_j\right) \tag{45}$$

$$= H(h_{ii}W_i) + H\left(\sum_{j\neq i} h_{ij}W_j\right), \tag{46}$$

where the last equality is thanks to the independence of the W_j, $1 \leqslant j \leqslant K$. Putting the pieces together, we finally obtain

$$\frac{H\left(\sum_{j=1}^{K} h_{ij}W_j\right) - H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)}{2\log|\mathcal{W}_N|} \tag{47}$$

$$= \frac{H(h_{ii}W_i)}{2\varphi(d)\log N} = \frac{\varphi(d)\log N}{2\varphi(d)\log N} = \frac{1}{2}, \tag{48}$$

where we used the scaling invariance of entropy, the fact that W_i is uniform on $\mathcal{W}$, and $|\mathcal{W}| = N^{\varphi(d)}$. This allows us to conclude that, for all d and N, we have

$$\min\left\{\frac{H\left(\sum_{j=1}^{K} h_{ij}W_j\right)}{2\log|\mathcal{W}_N|}, 1\right\} - \min\left\{\frac{H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)}{2\log|\mathcal{W}_N|}, 1\right\} \geqslant 1 - \frac{\varphi(d+1)\log((K-1)N)}{2\varphi(d)\log N}, \tag{49}$$

as either the first minimum on the LHS of (49) coincides with the non-trivial term in which case by (46) the second minimum coincides with the non-trivial term as well, and therefore by (48) the LHS of (49) equals $1/2 \geqslant 1 - \frac{\varphi(d+1)\log((K-1)N)}{2\varphi(d)\log N}$, or the first minimum coincides with 1 in which case we apply $\min\left\{\frac{H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)}{2\log|\mathcal{W}_N|}, 1\right\} \leqslant \frac{H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)}{2\log|\mathcal{W}_N|} \leqslant \frac{\varphi(d+1)\log((K-1)N)}{2\varphi(d)\log N}$, where we used (35) for the second inequality. As, by (36), the RHS of (49) converges to $1/2$ for $d, N \to \infty$, it follows that the LHS of (33) is asymptotically lower-bounded by $K/2$. This completes the proof. ■

VII. Non-asymptotic statement

Given a channel matrix $\mathbf{H}$ verifying Condition ($*$) in theory requires checking infinitely many equations of the form (4). It is therefore natural to ask whether we can say anything about the DoF achievable for a given $\mathbf{H}$ when (4) is known to hold only for finitely many coefficients a_j, b_j and up to a finite degree d. To address this question we consider the same input distributions as in the proof of Theorem 1 and carefully analyze the steps in the proof that employ Condition ($*$). Specifically, there are only two such steps, namely the argument on the uniqueness of the representation of elements in the set (34) and the argument leading to (46). First, as to uniqueness in (34) we need to verify that

$$\sum_{j=1}^{\varphi(d+1)} a_j f_j(\check{\mathbf{h}}) \neq \sum_{j=1}^{\varphi(d+1)} \widetilde{a}_j f_j(\check{\mathbf{h}}) \tag{50}$$

for all $a_j, \widetilde{a}_j \in \{1, ..., (K-1)N\}$ with $(a_1, ..., a_{\varphi(d+1)}) \neq (\widetilde{a}_1, ..., \widetilde{a}_{\varphi(d+1)})$. Note that we have to consider monomials up to degree $d+1$, as the multiplication of W_j by an off-diagonal channel coefficient h_{ij} increases the degrees of the involved monomials by 1, as already formalized in (34). Second, to get (46), we need to ensure that $h_{ii}W_i + \sum_{j\neq i} h_{ij}W_j$ uniquely determines $h_{ii}W_i$ and $\sum_{j\neq i} h_{ij}W_j$, for $i = 1, ..., K$, which amounts to requiring $h_{ii}w_i + \sum_{j\neq i} h_{ij}w_j \neq h_{ii}\widetilde{w}_i + \sum_{j\neq i} h_{ij}\widetilde{w}_j$ whenever $(h_{ii}w_i, \sum_{j\neq i} h_{ij}w_j) \neq (h_{ii}\widetilde{w}_i, \sum_{j\neq i} h_{ij}\widetilde{w}_j)$. Inserting the elements in (32) for $w_i, \widetilde{w}_i$ this condition reads

$$\sum_{j=1}^{\varphi(d+1)} a_j f_j(\check{\mathbf{h}}) + \sum_{j=1}^{\varphi(d)} b_j h_{ii} f_j(\check{\mathbf{h}}) \neq \sum_{j=1}^{\varphi(d+1)} \widetilde{a}_j f_j(\check{\mathbf{h}}) + \sum_{j=1}^{\varphi(d)} \widetilde{b}_j h_{ii} f_j(\check{\mathbf{h}}), \tag{51}$$

for all $a_j, \widetilde{a}_j \in \{1, ..., (K-1)N\}$ and $b_j, \widetilde{b}_j \in \{1, ..., N\}$ with

$$(a_1, ..., a_{\varphi(d+1)}, b_1, ..., b_{\varphi(d)}) \neq (\widetilde{a}_1, ..., \widetilde{a}_{\varphi(d+1)}, \widetilde{b}_1, ..., \widetilde{b}_{\varphi(d)}).$$

Note that (50) is a special case of (51) obtained by setting $b_j = \widetilde{b}_j$, for all j, in (51). Finally, rearranging terms we find that (51) simply says that non-trivial $\mathbb{Z}$-linear combinations of the elements participating in Condition $(*)$ do not equal zero, which in turn is equivalent to (4) restricted to a finite number of coefficients and a finite degree.

Now, assuming that, for a given $\mathbf{H}$, (51) is verified for all $a_j, \widetilde{a}_j, b_j, \widetilde{b}_j$ and fixed d and N, we can proceed as in the proof of Theorem 1 to get the following from (49):

$$\begin{aligned} &\min\left\{\frac{H\left(\sum_{j=1}^{K} h_{ij}W_j\right)}{\log(1/r)}, 1\right\} - \min\left\{\frac{H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)}{\log(1/r)}, 1\right\} \\ &\geqslant 1 - \frac{\varphi(d+1)\log((K-1)N)}{2\varphi(d)\log N} \\ &= 1 - \frac{(K(K-1)+d+1)\log((K-1)N)}{2(d+1)\log N}. \end{aligned}$$

Upon insertion into (33) this yields the DoF lower bound

$$\frac{K}{2}\left[2 - \frac{(K(K-1)+d+1)\log((K-1)N)}{(d+1)\log N}\right].$$

VIII. Condition $(*)$ is not necessary

While Condition $(*)$ is sufficient for $\mathsf{DoF}(\mathbf{H}) = K/2$, we next show that it is not necessary. This will be accomplished by constructing a class of example channel matrices that fail to satisfy Condition $(*)$ but still admit $K/2$ DoF. As, however, almost all channel matrices satisfy Condition $(*)$

this example class is necessarily of Lebesgue measure zero. Specifically, we consider channel matrices that have $h_{ii} \in \mathbb{R} \setminus \mathbb{Q}$, $i = 1, ..., K$, and $h_{ij} \in \mathbb{Q} \setminus \{0\}$, for $i, j = 1, ..., K$ with $i \neq j$. This assumption implies that all entries of $\mathbf{H}$ are nonzero, i.e., $\mathbf{H}$ is fully connected, which, again by [10, Prop. 1], yields $\mathsf{DoF}(\mathbf{H}) \leqslant K/2$. Moreover, as two rational numbers are linearly dependent over $\mathbb{Q}$, these channel matrices violate Condition $(*)$. We next show that nevertheless $\mathsf{DoF}(\mathbf{H}) \geqslant K/2$ and hence $\mathsf{DoF}(\mathbf{H}) = K/2$. This will be accomplished by constructing corresponding DoF-optimal input distributions.

We begin by arguing that we may assume $h_{ij} \in \mathbb{Z}$, for $i \neq j$. Indeed, since $\mathsf{DoF}(\mathbf{H})$ is invariant to scaling of rows or columns of $\mathbf{H}$ by a nonzero constant [12, Lem. 3], we can, without affecting $\mathsf{DoF}(\mathbf{H})$, multiply the channel matrix by a common denominator of the h_{ij}, $i \neq j$, thus rendering the off-diagonal entries integer-valued while retaining irrationality of the diagonal entries h_{ii}.

Let

$$\mathcal{W} := \{0, ..., N-1\}, \tag{52}$$

for some $N > 0$, and take $W_1, ..., W_K$ to be i.i.d. uniformly distributed on $\mathcal{W}$. We set the contraction parameter to

$$r = 2^{-2\log(2h_{\text{max}}KN)}, \tag{53}$$

where $h_{\text{max}} := \max\{|h_{ij}| \, : \, i \neq j\}$. Writing $\sum_{j=1}^{K} h_{ij}W_j = h_{ii} \cdot W_i + 1 \cdot \sum_{j\neq i} h_{ij}W_j$, where $W_i, \sum_{j\neq i} h_{ij}W_j \in \mathbb{Z}$, and realizing that $\{h_{ii}, 1\}$ is linearly independent over $\mathbb{Q}$, we can mimic the arguments leading to (46) to conclude that

$$H\left(\sum_{j=1}^{K} h_{ij}W_j\right) = H(h_{ii}W_i) + H\left(\sum_{j\neq i} h_{ij}W_j\right), \tag{54}$$

for $i = 1, ..., K$. In fact, it is precisely the linear independence of $\{h_{ii}, 1\}$ over $\mathbb{Q}$ that makes this example class work. Next, we note that

$$\sum_{j\neq i}^{K} h_{ij}W_j \in \{-h_{\text{max}}(K-1)N, ..., 0, ..., h_{\text{max}}(K-1)N\}$$

and hence $H\left(\sum_{j\neq i} h_{ij}W_j\right) \leqslant \log(2h_{\text{max}}KN)$. Since the W_j, $1 \leqslant j \leqslant K$, are identically distributed, we have $H(h_{ii}W_i) = H(h_{ij}W_j)$, for all i, j, and therefore $H(h_{ii}W_i) \leqslant H(\sum_{j\neq i} h_{ij}W_j)$ as a consequence of the fact that the entropy of a sum of independent random variables is greater than the entropy of each participating random variable [19, Ex. 2.14]. Thus (54) implies that

$$H\left(\sum_{j=1}^{K} h_{ij}W_j\right) \leqslant 2H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right) \leqslant 2\log(2h_{\text{max}}KN)\,.$$

With (53) we therefore obtain

$$\min\left\{\frac{H\left(\sum_{j=1}^{K} h_{ij}W_j\right)}{\log(1/r)}, 1\right\} = \frac{H\left(\sum_{j=1}^{K} h_{ij}W_j\right)}{\log(1/r)},$$

and since

$$H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right) \leqslant H\left(\sum_{j=1}^{K} h_{ij}W_j\right), \tag{55}$$

again by [19, Ex. 2.14], we also have

$$\min\left\{\frac{H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)}{\log(1/r)}, 1\right\} = \frac{H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)}{\log(1/r)}.$$

Applying Proposition 1 with (54) and using $H(h_{ii}W_i) = \log N$, we finally obtain

$$\mathsf{DoF}(\mathbf{H}) \geqslant \frac{\sum_{i=1}^{K} H(h_{ii}W_i)}{\log(1/r)} = \frac{K\log N}{\log(1/r)} = \frac{K\log N}{2\log(2h_{\max}KN)}. \tag{56}$$

Since (56) holds for all N, in particular for $N \to \infty$, this establishes that $\mathsf{DoF}(\mathbf{H}) \geqslant K/2$ and thereby completes our argument.

Recall that in the case of channel matrices satisfying Condition $(*)$ the value set $\mathcal{W}$ in (32) is channel-dependent. Here, however, the assumption of the diagonal entries of $\mathbf{H}$ being irrational and the off-diagonal entries rational already induces enough algebraic structure for our arguments to work. In the case of channel matrices satisfying Condition $(*)$ we induce an algebraic structure that is shared by all participating channel matrices through the choice of the channel-dependent set $\mathcal{W}$ and by enforcing Condition $(*)$. We conclude by noting that the example class studied here was investigated before in [7, Thm. 1] and [3, Thm. 6]. In contrast to [3], [7] our proof of DoF-optimality is, however, not based on arguments from Diophantine approximation theory.

IX. DoF-characterization in terms of Shannon entropy

To put our second main result, reported in this section, into context, we first note that the DoF-characterization [3, Thm. 4], see also (11) and the statement thereafter, is in terms of information dimension. As already noted, information dimension is, in general, difficult to evaluate. Now, it turns out that the DoF-lower bound in Proposition 1 can be developed into a full-fledged DoF-characterization in the spirit of [3, Thm. 4], which, however, will be entirely in terms of Shannon entropies.

Theorem 3: Achievability: For all channel matrices $\mathbf{H}$, we have

$$\sup_{W_1,\dots,W_K} \frac{\sum_{i=1}^{K}\left[H\left(\sum_{j=1}^{K} h_{ij}W_j\right) - H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)\right]}{\max_{i=1,\dots,K} H\left(\sum_{j=1}^{K} h_{ij}W_j\right)} \leqslant \mathsf{DoF}(\mathbf{H}), \tag{57}$$

where the supremum in (57) is taken over all independent discrete $W_1, ..., W_K$ such that the denominator in (57) is nonzero.[10]

Converse: We have equality in (57) for almost all $\mathbf{H}$ including channel matrices with all off-diagonal entries algebraic numbers and arbitrary diagonal entries.

Proof: We begin with the proof of the achievability statement. The idea of the proof is to apply Proposition 1 with a suitably chosen contraction parameter r. Specifically, let $W_1, ..., W_K$ be independent discrete random variables such that the denominator in (57) is nonzero, and apply Proposition 1 with

$$r := 2^{-\max_{i=1,\dots,K} H\left(\sum_{j=1}^{K} h_{ij} W_j\right)},$$

which ensures that all minima in (25) coincide with the respective non-trivial terms. Specifically, for $i = 1, ..., K$, we have

$$\min\left\{\frac{H\left(\sum_{j=1}^{K} h_{ij} W_j\right)}{\log(1/r)}, 1\right\} = \frac{H\left(\sum_{j=1}^{K} h_{ij} W_j\right)}{\max_{i=1,\dots,K} H\left(\sum_{j=1}^{K} h_{ij} W_j\right)}$$

$$\text{and}\quad \min\left\{\frac{H\left(\sum_{j\neq i}^{K} h_{ij} W_j\right)}{\log(1/r)}, 1\right\} = \frac{H\left(\sum_{j\neq i}^{K} h_{ij} W_j\right)}{\max_{i=1,\dots,K} H\left(\sum_{j=1}^{K} h_{ij} W_j\right)},$$

where the latter follows from $H\left(\sum_{j=1}^{K} h_{ij} W_j\right) \geqslant H\left(\sum_{j\neq i}^{K} h_{ij} W_j\right)$ (cf. (55)). Proposition 1 now yields

$$\frac{\sum_{i=1}^{K} \left[H\left(\sum_{j=1}^{K} h_{ij} W_j\right) - H\left(\sum_{j\neq i}^{K} h_{ij} W_j\right)\right]}{\max_{i=1,\dots,K} H\left(\sum_{j=1}^{K} h_{ij} W_j\right)} \leqslant \mathsf{DoF}(\mathbf{H}). \tag{58}$$

Finally, the inequality (57) is obtained by supremization of the LHS of (58) over all admissible $W_1, ..., W_K$.

To prove the converse, we begin by referring to the proof of [3, Thm. 4], where the following is shown to hold for almost all $\mathbf{H}$ including channel matrices $\mathbf{H}$ with all off-diagonal entries algebraic numbers and arbitrary diagonal entries: For every $\delta > 0$, there exist independent discrete random variables $W_1, ..., W_K$ and an $r \in (0, 1)$ satisfying[11]

$$\log(1/r) \geqslant \max_{i=1,\dots,K} H\left(\sum_{j=1}^{K} h_{ij} W_j\right) \tag{59}$$

[10]This condition only excludes the cases where all W_i that appear with nonzero channel coefficients are chosen as deterministic. In fact, such choices yield $\mathsf{dof}(X_1, ..., X_K; \mathbf{H}) = 0$ (irrespective of the choice of the contraction parameter r) and are thus not of interest.

[11]This statement is obtained from the proof of [3, Thm. 4] as follows. The W_i and r here correspond to the W_i and r^n defined in [3, Eq. (146)] and [3, Eq. (147)], respectively. The relation in (59) is then simply a consequence of [3, Eq. (153)] and the cardinality bound for entropy.

such that

$$\mathsf{DoF}(\mathbf{H}) \leqslant \delta + \frac{\sum_{i=1}^{K}\left[H\left(\sum_{j=1}^{K} h_{ij}W_j\right) - H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)\right]}{\log(1/r)}. \tag{60}$$

By (59) it follows that

$$\frac{\sum_{i=1}^{K}\left[H\left(\sum_{j=1}^{K} h_{ij}W_j\right) - H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)\right]}{\log(1/r)} \leqslant \frac{\sum_{i=1}^{K}\left[H\left(\sum_{j=1}^{K} h_{ij}W_j\right) - H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)\right]}{\max_{i=1,\ldots,K} H\left(\sum_{j=1}^{K} h_{ij}W_j\right)}.$$

Finally, letting $\delta \to 0$ and taking the supremum over all admissible $W_1, \ldots, W_K$, we get

$$\mathsf{DoF}(\mathbf{H}) \leqslant \sup_{W_1,\ldots,W_K} \frac{\sum_{i=1}^{K}\left[H\left(\sum_{j=1}^{K} h_{ij}W_j\right) - H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)\right]}{\max_{i=1,\ldots,K} H\left(\sum_{j=1}^{K} h_{ij}W_j\right)}$$

for almost all $\mathbf{H}$ including channel matrices $\mathbf{H}$ with all off-diagonal entries algebraic numbers and arbitrary diagonal entries. This completes the proof. ■

Remark 4: In the achievability part of Theorem 3, we have actually shown that for all $\mathbf{H}$

$$\begin{aligned}&\sup_{W_1,\ldots,W_K} \frac{\sum_{i=1}^{K}\left[H\left(\sum_{j=1}^{K} h_{ij}W_j\right) - H\left(\sum_{j\neq i}^{K} h_{ij}W_j\right)\right]}{\max_{i=1,\ldots,K} H\left(\sum_{j=1}^{K} h_{ij}W_j\right)} \\ &\leqslant \sup_{X_1,\ldots,X_K} \sum_{i=1}^{K}\left[d\left(\sum_{j=1}^{K} h_{ij}X_j\right) - d\left(\sum_{j\neq i}^{K} h_{ij}X_j\right)\right],\end{aligned} \tag{61}$$

which combined with (11) yields (57). The LHS of (61) is obtained by reasoning along the same lines as in the proof of Proposition 1, namely by applying the RHS of (61) to self-similar $X_1, \ldots, X_K$ with suitable contraction parameter r, invoking Theorem 2, and noting that the supremization is then carried out over a smaller set of distributions. By Theorem 3 we know that our alternative DoF-characterization is equivalent to the original DoF-characterization in [3, Thm. 4], i.e., (61) holds with equality, for almost all $\mathbf{H}$ including $\mathbf{H}$-matrices with all off-diagonal entries algebraic numbers and arbitrary diagonal entries, since in all these cases we have a converse for both DoF-characterizations. As shown in the next section, this includes cases where $\mathsf{DoF}(\mathbf{H}) < K/2$. Moreover, the two DoF-characterizations are equivalent on the "almost-all set" characterized by Condition $(*)$, as in this case the LHS of (61) equals $K/2$ and therefore by (11) and $\mathsf{DoF}(\mathbf{H}) \leqslant K/2$ [10, Prop. 1], we get that the RHS of (61) equals $K/2$ as well. What we do not know is whether (61) is always satisfied with equality, but certainly the set of channel matrices where this is not the case is of Lebesgue measure zero.

Remark 5: Compared to the original DoF-characterization [3, Thm. 4] the alternative expression in Theorem 3 exhibits two advantages. First, the supremization has to be carried out over discrete random variables only, whereas in [3, Thm. 4] the supremum is taken over general input distributions. Second, Shannon entropy is typically much easier to evaluate than information dimension. Our alternative characterization is therefore more amenable to both analytical statements and numerical evaluations.

This is demonstrated in the next section, where we put the new DoF-characterization to work to explain why determining the exact number of DoF for channel matrices with rational entries has remained elusive so far, even for simple examples. In addition, we will exemplify the quantitative applicability of our DoF-formula by improving upon the best-known bounds on the DoF of a particular channel matrix studied in [3].

X. DoF CHARACTERIZATION AND ADDITIVE COMBINATORICS

In this section, we apply our alternative DoF-characterization in Theorem 3 to establish a formal connection between the characterization of DoF for arbitrary channel matrices and sumset problems in additive combinatorics. We also show how Theorem 3 can be used to improve the best known bounds on the DoF of a particular channel matrix studied in [3].

We begin by noting that according to [7, Thm. 2] channel matrices with all entries rational admit strictly less than $K/2$ DoF, i.e.,

$$\mathsf{DoF}(\mathbf{H}) < \frac{K}{2}.$$

However, finding the exact number of DoF for rational $\mathbf{H}$, even for simple examples, turns out to be a very difficult problem. Based on our alternative DoF-characterization (57) in Theorem 3, which here holds with equality as all entries of $\mathbf{H}$ are rational, we will be able to explain why this problem is so difficult. Specifically, we establish that characterizing the DoF for $\mathbf{H}$ with all entries rational is equivalent to solving very hard problems in sumset theory. As noted before, however, finding the exact number of DoF is difficult only on a set of channel matrices of Lebesgue measure zero, since $\mathsf{DoF}(\mathbf{H}) = K/2$ for almost all $\mathbf{H}$.

The simplest non-trivial example is the 3-user case with

$$\mathbf{H} = \begin{pmatrix} h_1 & 0 & 0 \\ h_2 & h_3 & 0 \\ h_4 & h_5 & h_6 \end{pmatrix},$$

where $h_1, ..., h_6 \in \mathbb{Q} \setminus \{0\}$. Since $\mathsf{DoF}(\mathbf{H})$ is invariant to scaling of rows or columns of $\mathbf{H}$ by a nonzero constant [12, Lem. 3], we can transform this channel matrix as follows:

$$\begin{pmatrix} h_1 & 0 & 0 \\ h_2 & h_3 & 0 \\ h_4 & h_5 & h_6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ h_2 & h_3 & 0 \\ 1 & \frac{h_5}{h_4} & \frac{h_6}{h_4} \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ h_2 & h_3 & 0 \\ 1 & \frac{h_5}{h_4} & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 1 & \frac{h_3 h_4}{h_2 h_5} & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

We can therefore restrict ourselves to the analysis of channel matrices of the form

$$\mathbf{H}_\lambda = \begin{pmatrix} 1 & 0 & 0 \\ 1 & \lambda & 0 \\ 1 & 1 & 1 \end{pmatrix}, \tag{62}$$

where $\lambda \in \mathbb{Q}\backslash\{0\}$. This example class was studied before in [3], [7]. In particular, using the DoF-characterization in terms of information dimension (11), Wu et al. showed that [3, Thm. 11]

$$\mathsf{DoF}(\mathbf{H}_\lambda) = 1 + \sup_{X_1, X_2} \left[d(X_1 + \lambda X_2) - d(X_1 + X_2) \right], \tag{63}$$

where the supremum is taken over all independent X_1, X_2 such that $\mathbb{E}[X_1^2], \mathbb{E}[X_2^2] < \infty$ and the appearing information dimension terms exist. Based on (63) one can lower-bound $\mathsf{DoF}(\mathbf{H}_\lambda)$ through concrete choices for the input distributions X_1 and X_2. If one is interested in analytical expressions, these choices are, however, restricted to input distributions that allow analytical expressions for the information dimension terms appearing in (63). Upper bounds on $\mathsf{DoF}(\mathbf{H}_\lambda)$ can be established by employing general upper and lower bounds on information dimension. However, there is not much one can get beyond what basic inequalities deliver.

By applying Theorem 3 to the channel matrix (62), we next develop an alternative characterization to (63). The resulting expression for $\mathsf{DoF}(\mathbf{H}_\lambda)$ involves the minimization of the ratio of entropies of linear combinations of discrete random variables and is analytically and numerically more tractable than (63).

Theorem 4: For

$$\mathbf{H}_\lambda = \begin{pmatrix} 1 & 0 & 0 \\ 1 & \lambda & 0 \\ 1 & 1 & 1 \end{pmatrix},$$

we have

$$\mathsf{DoF}(\mathbf{H}_\lambda) = 2 - \inf_{U,V} \frac{H(U+V)}{H(U+\lambda V)}, \tag{64}$$

where the infimum is taken over all independent discrete random variables U, V such that[12] $H(U + \lambda V) > 0$.

Proof: As the off-diagonal entries of $\mathbf{H}_\lambda$ are all rational and therefore algebraic numbers, we have equality in (57), which upon insertion of $\mathbf{H}_\lambda$ yields

$$\mathsf{DoF}(\mathbf{H}_\lambda) = \sup_{U,V,W} \frac{H(U+\lambda V) + H(U+V+W) - H(U+V)}{\max\{H(U), H(U+\lambda V), H(U+V+W)\}}, \tag{65}$$

where the supremum is taken over all independent discrete random variables U, V, W such that the denominator in (65) is nonzero. Now, again using [19, Ex. 2.14], we have $H(U) \leqslant H(U + \lambda V)$,

[12]Again, this condition simply prevents the denominator in (64) from being zero. The case $H(U+\lambda V) = 0$ is equivalent to U and V deterministic. This choice would, however, yield $\mathsf{dof}(X_1, ..., X_K; \mathbf{H}) \leqslant 1$ and is thus not of interest.

which when inserted into (65) yields

$$\mathsf{DoF}(\mathbf{H}_\lambda) = \sup_{U,V,W} \frac{H(U+\lambda V) + H(U+V+W) - H(U+V)}{\max\{H(U+\lambda V), H(U+V+W)\}} \tag{66}$$

$$\leqslant 1 + \sup_{U,V,W} \frac{H(U+\lambda V) - H(U+V)}{\max\{H(U+\lambda V), H(U+V+W)\}} \tag{67}$$

$$\leqslant 1 + \sup_{U,V} \frac{H(U+\lambda V) - H(U+V)}{H(U+\lambda V)} \tag{68}$$

$$= 2 - \inf_{U,V} \frac{H(U+V)}{H(U+\lambda V)}, \tag{69}$$

where we used the fact that the supremum in (67) is non-negative (as seen, e.g., by choosing U to be non-deterministic and V deterministic) and hence invoking $\max\{H(U+\lambda V), H(U+V+W)\} \geqslant H(U+\lambda V)$ in the denominator of (67) yields the upper bound (68).

For the converse part, let U, V be independent discrete random variables such that $H(U+\lambda V) > 0$. We take W to be discrete, independent of U and V, and to satisfy

$$H(W) \geqslant H(U+\lambda V), \tag{70}$$

e.g., we may simply choose W to be uniformly distributed on a sufficiently large finite set. Applying Proposition 1 with $W_1 = U$, $W_2 = V$, $W_3 = W$, and $r := 2^{-H(U+\lambda V)}$, we obtain

$$\min\left\{\frac{H(U)}{H(U+\lambda V)}, 1\right\} + \min\left\{\frac{H(U+\lambda V)}{H(U+\lambda V)}, 1\right\} - \min\left\{\frac{H(U)}{H(U+\lambda V)}, 1\right\} \\ + \min\left\{\frac{H(U+V+W)}{H(U+\lambda V)}, 1\right\} - \min\left\{\frac{H(U+V)}{H(U+\lambda V)}, 1\right\} \leqslant \mathsf{DoF}(\mathbf{H}_\lambda). \tag{71}$$

Since $H(U+V+W) \geqslant H(W) \geqslant H(U+\lambda V)$, where the first inequality is by [19, Ex. 2.14] and the second by the assumption (70), we get from (71) that

$$2 - \min\left\{\frac{H(U+V)}{H(U+\lambda V)}, 1\right\} \leqslant \mathsf{DoF}(\mathbf{H}_\lambda). \tag{72}$$

We treat the cases $H(U+V) > H(U+\lambda V)$ and $H(U+V) \leqslant H(U+\lambda V)$ separately. If $H(U+V) > H(U+\lambda V)$, then

$$2 - \frac{H(U+V)}{H(U+\lambda V)} < 1 = 2 - \min\left\{\frac{H(U+V)}{H(U+\lambda V)}, 1\right\} \leqslant \mathsf{DoF}(\mathbf{H}_\lambda). \tag{73}$$

On the other hand, if $H(U+V) \leqslant H(U+\lambda V)$, (72) becomes

$$2 - \frac{H(U+V)}{H(U+\lambda V)} \leqslant \mathsf{DoF}(\mathbf{H}_\lambda). \tag{74}$$

Combining (73) and (74), we finally get

$$2 - \frac{H(U+V)}{H(U+\lambda V)} \leqslant \mathsf{DoF}(\mathbf{H}_\lambda), \tag{75}$$

for all independent U, V such that $H(U+\lambda V) > 0$. Taking the supremum in (75) over all admissible U and V completes the proof. ∎

Through Theorem 4 we reduced the DoF-characterization of $\mathbf{H}_\lambda$ to an optimization of the ratio of the entropies of two linear combinations of discrete random variables. This optimization problem has a counterpart in additive combinatorics, namely the following sumset problem: find finite sets $\mathcal{U}, \mathcal{V} \subseteq \mathbb{R}$ such that the relative size

$$\frac{|\mathcal{U} + \mathcal{V}|}{|\mathcal{U} + \lambda\mathcal{V}|} \tag{76}$$

of the sumsets $\mathcal{U} + \mathcal{V}$ and $\mathcal{U} + \lambda\mathcal{V}$ is minimal. The additive combinatorics literature provides a considerable body of useful bounds on (76) as a function of $|\mathcal{U}|$ and $|\mathcal{V}|$ [17]. A complete answer to this minimization problem does, however, not seem to be available. Generally, finding the minimal value of sumset quantities as in (76) or corresponding entropic quantities, i.e., $H(U+V)/H(U+\lambda V)$ in this case, appears to be a very hard problem, which indicates why finding the exact number of DoF of channel matrices with rational entries is so difficult.

The formal relationship between DoF characterization and sumset theory, by virtue of Theorem 3, goes beyond $\mathbf{H}$ with rational entries and applies to general $\mathbf{H}$. The resulting linear combinations one has to deal with, however, quickly lead to very hard optimization problems.

We finally show how our alternative DoF-characterization can be put to use to improve the best known bounds on $\mathsf{DoF}(\mathbf{H}_\lambda)$ for $\lambda = -1$. Similar improvements are possible for other values of λ. For brevity we restrict ourselves, however, to the case $\lambda = -1$.

Proposition 2: We have

$$1.13258 \leqslant \mathsf{DoF}(\mathbf{H}_{-1}) \leqslant \frac{4}{3}.$$

Proof: For the lower bound, we choose U and V to be independent and distributed according to

$$\mathbb{P}[U = 0] = \mathbb{P}[V = 0] = (0.08)^3$$

$$\mathbb{P}[U = 1] = \mathbb{P}[V = 1] = (0.08)^2$$

$$\mathbb{P}[U = 2] = \mathbb{P}[V = 2] = 0.08$$

$$\mathbb{P}[U = 3] = \mathbb{P}[V = 3] = 1 - 0.08 - (0.08)^2 - (0.08)^3.$$

This choice is motivated by numerical investigations, not reported here. It then follows from (64) that

$$\mathsf{DoF}(\mathbf{H}_{-1}) \geqslant 2 - \frac{H(U+V)}{H(U-V)} = 1.13258. \tag{77}$$

A more careful construction of U and V should allow improvements of this lower bound.

For the upper bound, let U and V be independent discrete random variables such that $H(U-V) > 0$ as required in the infimum in (64). Recall the entropy inequalities (17) and (18) stating that

$$H(U - V) \leqslant 3H(U + V) - H(U) - H(V) \tag{78}$$

$$H(U - V) \leqslant \frac{1}{2}H(U + V) + \frac{2}{3}(H(U) + H(V)). \tag{79}$$

Multiplying (78) by $2/3$ and adding the result to (79) yields

$$\frac{5}{3}H(U-V) \leqslant \frac{5}{2}H(U+V),$$

and hence

$$\frac{H(U+V)}{H(U-V)} \geqslant \frac{2}{3}. \tag{80}$$

Using (80) in (64), we then obtain

$$\mathsf{DoF}(\mathbf{H}_{-1}) = 2 - \inf_{U,V} \frac{H(U+V)}{H(U-V)} \leqslant \frac{4}{3},$$

which completes the proof. ■

The bounds in Proposition 2 improve on the best known bounds obtained in [3, Thm. 11][13] as $1.0681 \leqslant \mathsf{DoF}(\mathbf{H}_{-1}) \leqslant \frac{7}{5}$.

References

[1] 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, June 2014.

[2] M. Hochman, "On self-similar sets with overlaps and inverse theorems for entropy," *Annals of Mathematics*, Vol. 180, No. 2, pp. 773–822, Sep. 2014.

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

[5] 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.

[6] 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, June 2014.

[7] 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.

[8] C. Brandt, N. Viet Hung, and H. Rao, "On the open set condition for self-similar fractals," *Proc. of the AMS*, Vol. 134, No. 5, pp. 1369–1374, Oct. 2005.

[9] T. Tao and V. Vu, *Additive Combinatorics*, ser. Cambridge Studies in Advanced Mathematics. New York, NY: Cambridge University Press, 2006, Vol. 105.

[10] A. Høst-Madsen and A. Nosratinia, "The multiplexing gain of wireless networks," *Proc. IEEE Int. Symp. on Inf. Theory*, pp. 2065–2069, Sep. 2005.

[11] A. Guionnet and D. Shlyakhtenko, "On classical analogues of free entropy dimension," *Journal of Functional Analysis*, Vol. 251, pp. 738–771, Oct. 2007.

[12] D. Stotz and H. Bölcskei, "Degrees of freedom in vector interference channels," *Submitted to IEEE Trans. Inf. Theory, arXiv:1210.2259v2*, Vol. cs.IT, Sep. 2014.

[13] The lower bound stated in [3, Thm. 11] is actually 1.10. Note, however, that in the corresponding proof [3, p. 273], the term $H(U-V) - H(U+V)$ needs to be divided by $\log 3$, which seems to have been skipped and when done leads to the lower bound 1.0681 stated here.

[13] J. E. Hutchinson, “Fractals and self similarity,” *Indiana University Mathematics Journal*, Vol. 30, pp. 713–747, 1981.

[14] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 2nd ed. John Wiley & Sons, 2004.

[15] I. Ruzsa, “Sumsets and entropy,” *Random Structures & Algorithms*, Vol. 34, No. 1, pp. 1–10, Jan. 2009.

[16] T. Tao, “Sumset and inverse sumset theory for Shannon entropy,” *Combinatorics, Probability & Computing*, Vol. 19, No. 4, pp. 603–639, July 2010.

[17] I. Z. Ruzsa, “Sums of finite sets,” in *Number Theory: New York Seminar 1991–1995*, D. V. Chudnovsky, G. V. Chudnovsky, and M. B. Nathanson, Eds. Springer US, 1996, pp. 281–293.

[18] 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.

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