Information Dimension and the Degrees of Freedom of the Interference Channel

Yihong Wu, Member, IEEE, Shlomo Shamai (Shitz), Fellow, IEEE, and Sergio Verdú, Fellow, IEEE

Abstract—The degrees of freedom (DoFs) of the K-user Gaussian interference channel determine the asymptotic growth of the maximal sum rate as a function of the signal-to-noise ratio. Subject to a very general sufficient condition on the cross-channel gains, we give a formula for the DoFs of the scalar interference channel as a function of the deterministic channel matrix, which involves maximization of a sum of information dimensions over K scalar input distributions. Known special cases are recovered, and even generalized in certain cases with unified proofs.

Index Terms—Multiuser information theory, interference channel, information dimension, Shannon theory, sumset entropy inequalities, degrees of freedom, high-SNR channels.

I. INTRODUCTION

Consider a K-user real-valued memoryless Gaussian interference channel (IC) with a fixed deterministic channel matrix \( \mathbf{H} = [h_{ij}] \) (known at the encoder and decoder), where at each symbol epoch the \( i \)th user transmits \( X_i \) and the \( i \)th decoder receives

\[
Y_i = \sum_{j=1}^{K} \sqrt{\text{snr}} h_{ij} X_j + N_i,
\]

where \( \{X_i, N_i\}_{i=1}^{K} \) are independent with \( \mathbb{E}[X_i^2] \leq 1 \) and \( N_i \sim \mathcal{N}(0,1) \).

Denote the capacity region of (1) by \( C(\mathbf{H}, \text{snr}) \) and the sum-rate capacity by

\[
\overline{C}(\mathbf{H}, \text{snr}) \triangleq \max \left\{ \sum_{i=1}^{K} R_i : \mathbf{R} \in C(\mathbf{H}, \text{snr}) \right\}.
\]

The degrees of freedom (DoF) or the multiplexing gain is the \( \text{pre-log}^1 \) of the sum-rate capacity in the high-SNR regime,

\[
\text{DoF} = \limsup_{\text{snr} \rightarrow \infty} \frac{\overline{C}(\mathbf{H}, \text{snr})}{\frac{1}{2} \log \text{snr}}.
\]

Note that the normalization in the definition of (3) is such that \( \text{DoF} = 1 \) in the single-user case. Despite its suggestive terminology, \( \text{DoF} \) need not be an integer in the multi-user case.

Determining the degrees of freedom has been an active and challenging research endeavor, in view of the fact that the maximal sum rate of the Gaussian interference channel \( \overline{C}(\mathbf{H}, \text{snr}) \) remains unknown. In [2] it is shown that

\[
\text{DoF} \leq \frac{K}{2}
\]

for fully-connected \( \mathbf{H} \), i.e., \( \mathbf{H} \) with no zero entries. Using Diophantine approximations, (4) is shown to be achievable for Lebesgue-almost every \( \mathbf{H} \) [3]. The almost sure achievability of \( \frac{K}{2} \) for the vector interference channel with varying channel gains has been shown in [4] using the technique of interference alignment. Sufficient conditions on individual \( \mathbf{H} \) that guarantee equality in (4) are also given in [5] and [6]. However, equality does not hold in general: for example, if all the entries in \( \mathbf{H} \) are rational, then (4) is strict as shown in [5] based on additive-combinatorial results and deterministic channel approximations. In addition to theoretical investigations, various communication and interference alignments schemes have been proposed (see [7], for a comprehensive review).

The goal of this paper is to give a single-letter formula for \( \text{DoF}(\mathbf{H}) \) via the maximization of a functional involving the information dimension (also known as Rényi dimension or entropy dimension) [8] under a general sufficient condition on the off-diagonals of \( \mathbf{H} \), which, in particular, is satisfied almost surely or if the off-diagonals are algebraic. Moreover, the maximization gives a lower bound on \( \text{DoF}(\mathbf{H}) \) for any channel matrix \( \mathbf{H} \). This unified approach allows us to uncover new results, as well as recover previously known results obtained via different methods. We also give results for the more general case in which the rates are, unlike (2), not equally weighted.

The rest of the paper is organized as follows: Section II gives the necessary background on information dimension, which is the cornerstone of the main results presented in Section III. Our results on the degrees of freedom region are given in Section III-F. In Section IV we discuss self-similar distributions, which play an important role in constructing DoF-achieving input distributions. Proofs of the main results are given in Section V, while several technical lemmas are...
relegated to the appendix. The converse results for rational channel matrices are based on new sumset inequalities for linear combinations of independent group-valued random variables in Appendix E, which might be of independent interests.

II. INFORMATION DIMENSION

In this section we summarize the main properties of information dimension relevant to the current paper. In particular, the connection between the information dimension and the high-SNR asymptotics of mutual information with additive noise is presented in Section II-B. Other properties and further discussions can be found in [8]–[11].

A. Definitions

A key concept in fractal geometry, the information dimension of a random variable measures the rate of growth of the entropy of its successively finer discretizations [8].

Definition 1: Let $X$ be a real-valued random variable. For $m \in \mathbb{N}$, denote the uniformly quantized version of $X$ by

$$\langle X \rangle_m \triangleq \frac{[mX]}{m}. \quad (5)$$

The information dimension of $X$ is defined as

$$d(X) = \lim_{m \to \infty} \frac{H(\langle X \rangle_m)}{\log m}. \quad (6)$$

Useful if the limit in (6) does not exist, the lim inf and lim sup are called lower and upper information dimensions of $X$, respectively, denoted by $\underline{d}(X)$ and $\overline{d}(X)$.

Definition 1 can be readily extended to random vectors, where the floor function $\lfloor \cdot \rfloor$ is taken componentwise [8, p. 208]. Since $d(X)$ only depends on the distribution of $X$, we also denote $d(P_X) = d(X)$. Similar convention is traditionally followed with entropy and other information measures.

The information dimension of $X$ is finite if and only if the mild condition

$$H(\langle X \rangle) < \infty \quad (7)$$

is satisfied [8], [10]. A sufficient condition for finite information dimension is

$$\mathbb{E} \left[ \log(1 + |X|) \right] < \infty, \quad (8)$$

which is milder than the existence of $\mathbb{E}[X]$. Therefore (7) is satisfied for all random variables considered in this paper.

Equivalent definitions of information dimension include:

- For an arbitrary integer $M \geq 2$, write the $M$-ary expansion of $X$ as

$$X = [X] + \sum_{i \in \mathbb{N}} (X)_i M^{-i}. \quad (9)$$

where the $i^{\text{th}}$ digit $(X)_i \triangleq \lfloor M^i X \rfloor - M \lfloor M^{i-1} X \rfloor$ is a discrete random variable taking values on $\{0, \ldots, M - 1\}$. Then $\underline{d}(X)$ and $\overline{d}(X)$ coincide with the normalized lower and upper entropy rates of the process $(X)_i : j \in \mathbb{N}$.

- Denote by $B(x, \epsilon)$ the open ball$^3$ of radius $\epsilon$ centered at $x$. Then (see [12, Definition 4.2] and [10, Appendix A])

$$d(X) = \lim_{\epsilon \to 0} \frac{\mathbb{E} \left[ \log P_X(B(X, \epsilon)) \right]}{\log \epsilon}. \quad (10)$$

With Shannon entropy replaced by Rényi entropy of order $\alpha$ in Definition 1, the generalized notion of dimension of order $\alpha$ is defined similarly [9], [14]:

Definition 2 (Information Dimension of Order $\alpha$): Let $\alpha \in [0, \infty]$. The information dimension of $X$ of order $\alpha$ is defined as

$$d_\alpha(X) = \lim_{m \to \infty} \frac{H_\alpha(\langle X \rangle_m)}{\log m}, \quad (11)$$

where $H_\alpha(Y)$ denotes the Rényi entropy of order $\alpha$ of a discrete random variable $Y$ with probability mass function $P_Y$ defined on $\mathcal{Y}$, defined as

$$H_\alpha(Y) = \begin{cases} \frac{1}{\alpha} \log \max_{y \in \mathcal{Y}} P_Y(y) \alpha, & \alpha = \infty, \\ 1 - \frac{1}{\alpha} \log \left( \sum_{y \in \mathcal{Y}} P_Y(y)^\alpha \right), & \alpha \neq 1, \infty. \end{cases} \quad (12)$$

The lim inf and lim sup of the ratio in (11) are called lower and upper information dimensions of $X$ of order $\alpha$ respectively, denoted by $d_\alpha(X)$ and $\overline{d_\alpha}(X)$.

Note that $d_1(X) = d(X)$. As a consequence of the monotonicity of Rényi entropy, $d_\alpha$ decreases with $\alpha$. Moreover, $\alpha \mapsto d_\alpha(X)$ is continuous except possibly at $\alpha = 1$ [15]. For instance, if $X$ has a discrete-continuous mixed distribution, then $d_\alpha(X) = d(X) = 1$ for all $\alpha < 1$, while $d(X)$ equals to the weight of the continuous part (see Theorem 1). If $X$ has a Cantor distribution (see [16] or Section IV), $d_\alpha(X) = \log_3 2$ for all $\alpha \in [0, \infty]$.

B. Properties

The following are basic properties of information dimension [8], [10], the last three of which are inherited from Shannon entropy.

Lemma 1:

- $0 \leq d(X^n) \leq n$. \quad (13)
- Scale-Invariance: For all $\alpha \neq 0$,

$$d(\alpha X^n) = d(X^n). \quad (14)$$

- If $X^n$ and $Y^n$ are independent, then$^5$

$$\max\{d(X^n), d(Y^n)\} \leq d(X^n + Y^n) \leq d(X^n) + d(Y^n). \quad (15)$$

- If $X^n$, $Y^n$, $Z^n$ are independent, then

$$d(X^n + Y^n + Z^n) \leq d(X^n + Y^n) + d(Y^n+Z^n). \quad (17)$$

$^3$By the equivalence of norms on finite-dimensional space, the value of the limit on the right side of (10) is independent of the underlying norm.

$^4$Also known as the $L^n$ dimension or spectrum [13]. In particular, $d_2$ is also called the correlation dimension.

$^5$The upper bound (16) can be improved by defining conditional information dimension.
Proof: Appendix A.

To calculate the information dimension in (6) or (11), it is sufficient to restrict to an exponential subsequence, as a result of the following lemma.

Lemma 2: Let $\lambda > 1$. If $d_\alpha(X)$ exists, then
\[
    d_\alpha(X) = \lim_{l \to \infty} \frac{H_\alpha((X)_{gl})}{l \log \lambda}.
\]

As shown in [8], the information dimension for the mixture of discrete and absolutely continuous distributions can be determined as follows (see [8, Ths. 1 and 3] or [9, Th. 1, pp. 588–592]):

Theorem 1: Assume that the distribution of a scalar random variable $X$ is given by
\[
    v = (1 - \rho)v_d + \rho v_c,
\]

where $v_d$ is a discrete probability measure, $v_c$ is an absolutely continuous probability measure and $0 \leq \rho \leq 1$. Then
\[
    d(X) = \rho.
\]

In particular, if $X$ has a density (with respect to Lebesgue measure), then $d(X) = 1$; if $X$ is discrete, then $d(X) = 0$.

When the distribution of $X$ has a singular component, its information dimension does not admit a simple formula in general. However, for the important class of self-similar singular distributions, the information dimension can be explicitly determined, as we show in Section IV-B.

C. Information Dimension and High-SNR Mutual Information

The high-SNR asymptotics of mutual information with additive noise is governed by the input information dimension, as shown by the following result due to Guionnet and Shlyakhtenko [17]. In this paper we give a considerably simplified and self-contained proof for this result based on a non-asymptotic bound on mutual information in Section V-A, from which Theorem 1 is an immediate consequence.

Theorem 2: Let $X$ be independent of $N$ which is standard normal. Denote
\[
    I(X, \text{snr}) \triangleq I(X; \sqrt{\text{snr}}X + N),
\]

Then
\[
    \lim_{\text{snr} \to \infty} \frac{I(X, \text{snr})}{\frac{1}{2} \log \text{snr}} = d(X).
\]

Moreover, (22) holds verbatim in the vector case.

In view of Theorem 2, the information dimension $d(X)$ represents the single-user degrees of freedom when the input distribution is constrained to be $P_X$. Naturally, information dimension, as we will see, also appears in the characterization of degrees of freedom in the multi-user case. In the single-user case, (22) indicates that mutual information is maximized

\[
    \text{asymptotically by any absolutely continuous input distribution, in view of Theorem 1 in Section II-B. In the interference channel, as we will see in Section III, the situation is far more intricate since analog inputs are deleterious from the viewpoint of interference.}
\]

D. Information Dimension Under Projection

Let $A \in \mathbb{R}^{m \times n}$ with $m \leq n$. Then for any $X^n$,
\[
    d(AX^n) \leq \min\{d(X^n), \text{rank}(A)\}.
\]

Understanding how the dimension of a measure behaves under projections is a basic problem in fractal geometry. It is well-known that almost every projection preserves the dimension, be it Hausdorff dimension (Marstrand’s projection theorem [20, Ch. 9]) or information dimension [12, Ths. 1.1 and 4.1]. However, computing the dimension for individual projections is, in general, difficult.

The preservation of information dimension of order $\alpha \in (1, 2]$ under typical projections is shown in [12, Th. 1], which is relevant to our investigation of the almost sure behavior of degrees of freedom.

Lemma 3: [12] Let $\alpha \in (1, 2]$ and $m \leq n$. Then for almost every $A \in \mathbb{R}^{m \times n}$,
\[
    d_a(AX^n) = \min\{d_a(X^n), m\}.
\]

It is easy to see that (24) fails for information dimension ($\alpha = 1$) [12, p. 1041]: Let $P_{X^n} = (1 - \rho)d_0 + \rho Q$, where $Q$ is an absolutely continuous distribution on $\mathbb{R}^n$, $n > m$ and $\rho < 1$. Note that almost surely, $AX^n$ also has a mixed distribution with an atom at zero of mass $1 - \rho$. Then $d(AX^n) = \rho m < \min\{d(X^n), m\} = \min\{\rho n, m\}$. Examples of nonpreservation of $d_a$ for $0 \leq \alpha < 1$ or $\alpha > 2$ are given in [12, Sec. 5].

A problem closely related to determining DoF$(H)$ is to find the dimension difference of a product measure (i.e., distribution of a pair of independent random variables) under two projections (c.f. Remark 5 in Section III-G). Let $p, q, p', q'$ be non-zero real numbers. Then
\[
    d(pX + qY) - d(p'X + q'Y) \\
\leq \frac{1}{2}(d(pX + qY) + d(pX + qY) - d(X) - d(Y)) \leq \frac{1}{2},
\]

where (25) and (26) follow from (15) and (16), respectively. Therefore, the dimension of a two-dimensional product measure under two projections can differ by at most one half. Moreover, the next result shows that if the coefficients are rational, then the dimension difference is strictly less than one half. The proof is based on new entropy inequalities for linear combinations of independent random variables developed in Appendix E.

Theorem 3: Let $p, p', q, q'$ be non-zero integers. Then
\[
    d(pX + qY) - d(p'X + q'Y) \leq \frac{1}{2} - \epsilon(p', q, pq'),
\]

where
\[
    \epsilon(a, b) \triangleq \frac{1}{56(|\log |a|| + |\log |b|) + 16}.
\]
In the special cases of \( p = 2, q = 1 \) and \( p = 1, q = -1 \), the following improved bounds hold
\[
d(2X + Y) - d(X + Y) \leq \frac{3}{7},
\]
and
\[
d(X - Y) - d(X + Y) \leq \frac{2}{5}.
\]

Proof: Appendix E. \( \square \)

Remark 1: The constants in (28) are, of course, not optimal. However, the logarithmic dependence on the coefficients is sharp, as shown by the following example: For any \( p \in \mathbb{N} \), there exists independent \( X \) and \( Y \), such that
\[
d(pX + Y) - d(X + Y) \geq \frac{1}{p} - \Theta \left( \frac{1}{\log(p)} \right). \tag{30}
\]
See the proof of Theorem 11 in Section V-E for the construction.

III. MAIN RESULTS

A. A General Formula

The degrees of freedom of the \( K \)-user interference channel with channel matrix \( \mathbf{H} \) defined in (3) is given by the following result.

Theorem 4: Let
\[
dof(X^K, \mathbf{H}) \triangleq \sum_{i=1}^{K} d \left( \sum_{j=1}^{K} h_{ij} X_{j} \right) - d \left( \sum_{j \neq i} h_{ij} X_{j} \right). \tag{31}
\]
Then, for any \( \mathbf{H} \),
\[
\text{DoF}(\mathbf{H}) \geq \sup_{X^K} \text{dof}(X^K, \mathbf{H}), \tag{32}
\]
where the supremum is over independent \( X_1, \ldots, X_K \) such that
\[
H(\{X_i\}) < \infty \tag{33}
\]
and all information dimensions appearing in (31) exist. Equality holds in (32) except possibly for \( \mathbf{H} \) whose off-diagonals are non-algebraic and belong to a set of zero Lebesgue measure.

Proof: Section V-B. \( \square \)

For a fixed channel matrix \( \mathbf{H} \), we have been able to show that equality holds in (32) under the sufficient condition that all cross-channel gains are algebraic (e.g., rationals, quadratic irrationals, etc.).

In fact, the lower bound (32) holds even if \( \text{DoF}(\mathbf{H}) \) in (3) is defined with a limit \( \lim_{\text{snr} \to \infty} \). Therefore the limit exists whenever equality holds in (32).

In Section V-B, we give a deterministic sufficient condition on the cross-channel gains, which is more general than algebraicity and holds almost surely. See (112).

Remark 2: Condition (33) is much weaker than the average power constraint, since \( \mathbb{E} [X^2] \leq 1 \) implies that \( H(\{X\}) \leq \mathbb{E} [1 + \|\{X\}\|] \leq 2 \). In fact, as we show in Section V-B, in order for (32) to hold, the inputs need not satisfy any power constraint. As long as (33) is satisfied, \( \text{DoF}(\mathbf{H}) \) does not change even if we allow infinite input power, in which case \( \text{snr} \) is simply a scale parameter no longer carrying the operational meaning of signal-to-noise ratio.

We proceed to offer some insight on the formula in Theorem 4. By the limiting characterization of the interference channel capacity region \( [21] [22], \text{eq. (2)} \), the sum-rate capacity is given by
\[
\mathcal{C}(\mathbf{H}, \text{snr}) = \lim_{n \to \infty} \frac{1}{n} \sup_{X_1^n, \ldots, X_K^n} \sum_{i=1}^{K} I(X_i^n; Y_i^n),
\]
where \( X_i^n = [X_{i1}, \ldots, X_{in}] \) is the input of the \( i \)-th user, and the supremum is over independent \( X_1^n, \ldots, X_K^n \) satisfying the corresponding average cost constraints. Then,
\[
I(X_i^n; Y_i^n) = I(X_1^n, \ldots, X_K^n; Y_i^n) - I(X_1^n, \ldots, X_{i-1}^n; Y_i^n | X_i^n)
\]
\[
= I \left( \sum_{j=1}^{K} h_{ij} X_{j}^n, \text{snr} \right) - I \left( \sum_{j \neq i} h_{ij} X_{j}^n, \text{snr} \right),
\]
where \( I(\cdot, \text{snr}) \) is defined in (21). Therefore the degrees of freedom admit the following limiting characterization:
\[
\text{DoF}(\mathbf{H}) = \lim_{\text{snr} \to \infty} \limsup_{n \to \infty} \sup \frac{2}{n \log \text{snr}} \sum_{i=1}^{K} I \left( \sum_{j=1}^{K} h_{ij} X_{j}^n, \text{snr} \right) - I \left( \sum_{j \neq i} h_{ij} X_{j}^n, \text{snr} \right),
\]
where \( I(\cdot, \text{snr}) \) is defined in (21). Therefore the degrees of freedom admit the following limiting characterization:

Our main result is the single-letterization of (39). Note that if the limit over \( \text{snr} \) were inside the supremum, then the desired result (32) would follow immediately in view of the high-SNR scaling of mutual information in Theorem 2. Indeed, employing i.i.d. input distributions yields (32). Thus the main difficulty in the converse proof lies in exchanging the supremum with the limits in (39), which amounts to proving that varying the input distribution with increasing SNR does not improve the degrees of freedom. To this end, we need a non-asymptotic version of Theorem 2 given in Section V-A.

In the rest of the section, we assume that the cross-channel gains are algebraic and we discuss some of the benefits of the information dimension characterization in Theorem 4.

B. Some Simple Special Cases

As we illustrate next, the degrees of freedom of various channels can be obtained by specializing Theorem 4.

- Two-User IC:

\[
\text{DoF} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{cases} 0 & a = d = 0 \\ 2 & a \neq 0, d \neq 0, b = c = 0 \\ 1 & \text{otherwise} \end{cases}
\]

\text{The second limit in (39) can be replaced by supremum over } n \in \mathbb{N}. \]


- **Many-to-One IC [23]:** if \( h_{ii} \neq 0 \) for all \( i \) and \( h_{1i} \neq 0 \) for some \( i \), then

\[
\text{DoF} \left( \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1K} \\ 0 & h_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & h_{KK} \end{bmatrix} \right) = K - 1. \quad (41)
\]

To see this, assuming \( h_{12} \neq 0 \), we have

\[
\text{dof}(X^K, \mathbf{H}) = d \left( \sum_{j} h_{1j}X_j \right) - d \left( \sum_{j \neq 1} h_{1j}X_j \right) + \sum_{j} d(X_j) \leq d \left( \sum_{j} h_{1j}X_j \right) + \sum_{j=3}^{K} d(X_j) \leq K - 1,
\]

where (42) is due to (15). The upper bound \( K - 1 \) is attained by choosing \( X_2, \ldots, X_K \) to be absolutely continuous.

- **One-to-Many IC [23]:** if \( h_{ii} \neq 0 \) for all \( i \), then

\[
\text{DoF} \left( \begin{bmatrix} h_{11} & 0 & \cdots & 0 \\ h_{21} & h_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{K1} & 0 & \cdots & h_{KK} \end{bmatrix} \right) = K - 1 \quad (44)
\]

To verify (44) note that

\[
\text{dof}(X^K, \mathbf{H}) = \sum_{j \neq 1} d \left( h_{1j}X_1 + h_{jj}X_j \right) - (K - 2)d(X_1) \leq \sum_{j \neq 1} d \left( h_{1j}X_1 + h_{jj}X_j \right) \leq K - 1,
\]

attained by choosing \( X_1 \) discrete and the rest absolutely continuous.

- **Multiple-Access Channel (MAC):** If \( \mathbf{H} \) is the all-one matrix, then \( \text{DoF}(\mathbf{H}) = 1 \), because

\[
\text{dof}(X^K, \mathbf{H}) = K d \left( \sum_{j=1}^{K} X_j \right) - K d \left( \sum_{j \neq i} X_j \right) \leq d \left( \sum_{j=1}^{K} X_j \right) \leq 1,
\]

where we have used the following additive-combinatorial result [24, p. 3]:

\[
(K - 1) H \left( \sum_{i=1}^{K} U_j \right) \leq K H \left( \sum_{j \neq i} U_j \right)
\]

with \( \{U_i\} \) taking values on an arbitrary group. In view of Lemma 9, setting \( U_j = \{X_j\}_m \) in (51) and sending \( m \to \infty \) yields (49). More generally, \( \text{DoF}(\mathbf{H}) = 1 \) if \( \text{rank}(\mathbf{H}) = 1 \). Another way to see that \( \text{DoF}(\mathbf{H}) = 1 \) is to invoke the last property of \( \text{DoF}(\cdot) \) in Section III-D:

\[
\text{DoF}(\mathbf{H}) \leq \text{DoF}(\mathbf{H}') = \sup_{X^K} d \left( \sum_{j} X_j \right) \leq 1.
\]

- **C. Suboptimality of Discrete-Continuous Mixtures**

Mixed discrete-continuous input distributions are strictly suboptimal in general. In fact, they achieve at most one degree of freedom in the fully-connected case. To see this, denote for brevity \( d(X_j) = \rho_i \). Assume that \( \mathbf{H} \) is fully-connected. For each \( i \), \( \sum_{j \neq i} h_{ij}X_j \) has a discrete-continuous distribution, where the weight of the discrete component is equal to \( \prod_{j=1}^{K} (1 - \rho_j) \).

\[
\text{dof}(X^K, \mathbf{H}) = \sum_{i=1}^{K} \left( 1 - \prod_{j=1}^{K} (1 - \rho_j) \right) \left( 1 - \prod_{j \neq i} (1 - \rho_j) \right)
\]

\[
= \sum_{i=1}^{K} \rho_i \prod_{j \neq i} (1 - \rho_j) \leq 1,
\]

Therefore to obtain more than one degree of freedom, it is necessary to employ input distributions with singular components.

As we will show later, singular distributions of information dimension one half are crucial in achieving the maximal degrees of freedom. In Section IV-C we give a family of such distributions \( \{\mu_N\}_{N \geq 2} \) which are self-similar [16] and \( \text{DoF} \)-achieving in many cases. In fact, \( \mu_N \) can be understood as the distribution of a random variable whose \( N \)-ary expansion has equiprobable even digits and zero odd digits.

\[
d(\mu_N) = \frac{1}{2} \quad \text{in view of (9)}.
\]

- **D. Properties of DoF(\mathbf{H})**

The following are immediate consequences of Theorem 4 combined with the elementary properties of information dimension in Lemma 1:

- \( \text{DoF}(\mathbf{H}) \) is invariant under row or column scaling [5, Lemma 1], in view of (14). To be more precise, for any \( X^K \), \( \text{dof}(X^K, \mathbf{H}) \) is invariant under row scaling by (14), but \( \text{dof}(X^K, \mathbf{H}) \) is not invariant under column scaling. However, by replacing each \( X_i \) with its scaled version, \( \sup_{X^K} \text{dof}(X^K, \mathbf{H}) \) is invariant under column scaling.

- \( \text{DoF}(\mathbf{H}) \leq K \), with equality if and only if \( \mathbf{H} \) is a diagonal matrix with no diagonal entry equal to zero. To see this, note that \( \text{dof}(X^K, \mathbf{H}) = K \) if and only if

\[
d \left( \sum_{j=1}^{K} h_{ij}X_j \right) = 1,
\]

and

\[
d \left( \sum_{j \neq i} h_{ij}X_j \right) = 0.
\]
for all $i$. On one hand, (54)–(55) imply $h_{ii} \neq 0$. On the other hand, since
\[
d(X_i) \geq d \left( \sum_j h_{ij} X_j \right) - d \left( \sum_{j \neq i} h_{ij} X_j \right),
\]
we have $d(X_i) = 1$ for all $i$, and (55) implies $h_{ij} = 0$ for all $i \neq j$.

- Let $\text{diag}(\mathbf{H}) = [h_{11}, \ldots, h_{KK}]$. Then $\text{DoF}(\mathbf{H}) \geq 0$, with equality if and only if $\text{diag}(\mathbf{H}) = 0$. If $\text{diag}(\mathbf{H}) \neq 0$, then $\text{DoF}(\mathbf{H}) \geq 1$, which can be achieved by choosing one of the inputs to be absolutely continuous (unit dimension) and the rest discrete (zero dimension).

-Removing cross-links increases the degrees of freedom: Let $\mathbf{H}'$ be obtained from $\mathbf{H}$ by setting some of the off-diagonal entries to zero. By (17), for any independent $X^K$, $\text{dof}(X^K, \mathbf{H}) \leq \text{dof}(X^K, \mathbf{H}')$. Therefore $\text{DoF}(\mathbf{H}) \leq \text{DoF}(\mathbf{H}')$. This can also be seen using a genie-aided converse argument by providing the interfering messages to the relevant receivers.

\section*{E. Bounds and Exact Expressions}

Next we prove that the number of degrees of freedom is upper bounded by $\frac{K}{2}$ under more general sufficient conditions than the fully-connected assumption in [2].

\textbf{Theorem 5:} Suppose that $\mathbf{H}$ satisfies the assumption in Theorem 4. Let $\pi$ be a fixed-point-free permutation on $\{1, \ldots, K\}$, i.e., $\pi(i) \neq i$ for all $i$. If $h_{\pi(i),i} \neq 0$ for each $i$, then
\[
\text{DoF}(\mathbf{H}) \leq \frac{K}{2}.
\]

Moreover, if $K$ is odd and $\pi$ is cyclic, i.e., there exist distinct elements $\{n_1, \ldots, n_K\}$ of $\{1, \ldots, K\}$, such that $\pi(n_i) = n_{i+1}$ for $i = 1, \ldots, K - 1$ and $\pi(n_K) = n_1$, then $\text{dof}(X^K, \mathbf{H}) = \frac{K}{2}$, if and only if for each $i$,
\[
d(X_i) = \frac{1}{2},
\]
\[
d \left( \sum_{j \neq i} h_{ij} X_j \right) = \frac{1}{2},
\]
\[
d \left( \sum_{j=1}^{K} h_{ij} X_j \right) = 1.
\]

\textbf{Proof:} Section V-C. □

\textbf{Remark 3:} In the second part of Theorem 5, the assumption that $\pi$ is cyclic is not superfluous. To see this, consider $K = 5$, $\pi(1) = 2$, $\pi(2) = 1$, $\pi(3) = 4$, $\pi(4) = 5$, $\pi(5) = 3$ and $h_{ij} = 0$ except for $h_{ii}$ and $h_{\pi(i), i}$. Such a channel consists of a two-user IC decoupled with a three-user IC. Therefore as the sum of the respective DoFs, $\text{DoF}(\mathbf{H}) = 5/2$, which can be achieved by choosing $X_1$ to be absolutely continuous, $X_2$ discrete and $X_3, X_4, X_5$ to be the corresponding DoF-achieving distributions for the three-user IC. Clearly in this case (58) is not necessary. The same also applies to the case of even number of users.

Next we give various sufficient conditions on $\mathbf{H}$ that guarantee $\text{DoF}(\mathbf{H}) = \frac{K}{2}$.

\textbf{Theorem 6 (}[3]\textbf{):} $\text{DoF}(\mathbf{H}) = \frac{K}{2}$ for Lebesgue-a.e. $\mathbf{H}$.

\textbf{Proof:} Section V-D. □

Recently, Stotz and Bölcskei [25], following up on an earlier version of this paper, obtained deterministic sufficient conditions on $\mathbf{H}$ to guarantee that $\text{DoF}(\mathbf{H}) = \frac{K}{2}$, which imply Theorem 6 as a special case. Using the achievability bound (32) via information dimension and the self-similar input distribution in Section IV, the proof in [26] invokes a recent result by Hochman [27] in fractal geometry, in lieu of the Diophantine approximation results used in the present paper as well as [5] and [3]. In particular, using results from [27] it is shown in [25, Th. 2] that for the self-similar construction in (73), the information dimension formula (81) holds for almost every contraction ratio $r$ without requiring the open set condition, which can be difficult to verify. This shortcut offers simpler proofs of achievability results such as the following (see [25, Sec. VIII]):

\textbf{Theorem 7:} If the off-diagonal entries of $\mathbf{H}$ are rational and the diagonal entries are irrational non-zeros, then $\text{DoF}(\mathbf{H}) = \frac{K}{2}$.

Theorem 7 generalizes the following known results, previously obtained using different Diophantine approximation results:

- [5, Th. 1], which relies on the Thue-Siegel-Roth theorem [28] and requires the diagonal entries to be irrational algebraic numbers. Note that the set of real algebraic numbers is dense but countable. Therefore, almost every real number is transcendental.

- [6, Th. 1(2)], which assumes that the diagonal and off-diagonal entries are equal to one and $h$ respectively, with $h$ irrational. Upon scaling, this is equivalent to a channel matrix with unit off-diagonal entries and irrational diagonal entries $h^{-1}$.

\textbf{Remark 4:} Theorem 7 requires all diagonal entries to be irrational. Otherwise $\frac{K}{2}$ degrees of freedom might be unattainable. For instance, consider the $3 \times 3$ channel matrix $\mathbf{H}_3$ in (65). Since $\text{DoF}$ is invariant under row or column scaling, $\text{DoF}(\mathbf{H}_3)$ does not depend on $h_{11}$ or $h_{33}$ as long as they are both non-zero. Then Theorem 11 implies that $\text{DoF}(\mathbf{H}_3) < \frac{3}{2}$ if $h_{22}$ is rational, even if $h_{11}$ and $h_{33}$ are both irrational.

For $\mathbf{H}$ with non-zero rational coefficients it is known that $\text{DoF}(\mathbf{H}) < \frac{K}{2}$ [5]. Next we give a more general sufficient condition for the strict inequality in the three-user case.

\textbf{Theorem 8:} Let $K = 3$ and let the off-diagonals of $\mathbf{H}$ be algebraic. If there exist distinct $i, j, k$, such that $h_{ij}, h_{ii}, h_{kj}$ and $h_{ik}$ are non-zero rationals, then $\text{DoF}(\mathbf{H}) < \frac{3}{2}$.

\textbf{Proof:} Section V-E. □

Examples of strict improvement of $\text{DoF}(\mathbf{H}) < \frac{3}{2}$ for certain $\mathbf{H}$ can be found in Section III-G. The following example, which is proved in Appendix F, illustrates that the condition in Theorem 8 is not superfluous:

\[
\text{DoF} \left( \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 2 \\ 2 & 0 & 1 \end{bmatrix} \right) = \frac{3}{2}.
\]
F. Degrees of Freedom Region

The degrees of freedom region of the interference channel (1) is defined as the high-SNR limit of normalized capacity regions:

\[
\text{DoF}(\mathbf{H}) \triangleq \left\{ r^K : \lim_{\text{SNR} \to \infty} \rho \left( r^K, \frac{C(\mathbf{H}, \text{SNR})}{\frac{\pi}{2} \log \text{SNR}} \right) = 0 \right\},
\]

where \( \rho(x, E) \triangleq \inf_{y \in E} \|x - y\|_2 \) denotes the distance between the point \( x \) and the set \( E \) belonging to an arbitrary Euclidean space.

The degrees of freedom region exists and is characterized by the following result, whose proof is almost identical to that of Theorem 4:

**Theorem 9:** Suppose that \( \mathbf{H} \) satisfies the assumption in Theorem 4. \( \text{DoF}(\mathbf{H}) \) is the collection of all \( r^K \in [0, 1]^K \), such that for any probability vector \( w^K \),

\[
(r^K, w^K) \leq \sup_{K} \left\{ \sum_{i=1}^{K} w_i d \left( \sum_{j=1}^{K} h_{ij} X_j \right) - w_i d \left( \sum_{j \neq i} h_{ij} X_j \right) \right\},
\]

where the supremum is over independent \( X_1, \ldots, X_K \) such that (33) is satisfied and all information dimensions appearing in (63) exist.

Analogous to Theorems 5 and 6, the following result shows that the degrees of freedom region coincides with a polyhedron for almost every channel matrix. See Fig. 1 for an illustration in the three-user case.

**Theorem 10:** Let \( \mathbf{H} \) satisfy the assumption in Theorem 5. Then

\[
\text{DoF}(\mathbf{H}) \subset \text{co} \left\{ \mathbf{e}_1, \ldots, \mathbf{e}_K, \frac{1}{2} \mathbf{1}, \mathbf{0} \right\},
\]

where \( \{\mathbf{e}_1, \ldots, \mathbf{e}_K\} \) is the standard basis, and \( \text{co} \) denotes the convex hull. Moreover, (64) holds with equality for Lebesgue-a.e. \( \mathbf{H} \).

**Proof:** Section V-C.

\[\Box\]

G. An Example of Lower-Triangular Channel Matrix

In this section we consider a prototypical example of the following lower-triangular channel matrix [5, Sec. V], which captures the essential difficulty of the three-user interference channel:

\[
\mathbf{H}_\lambda \triangleq \begin{bmatrix}
1 & 0 & 0 \\
1 & \lambda & 0 \\
1 & 1 & 1
\end{bmatrix}.
\]

Since the off-diagonals of \( \mathbf{H}_\lambda \) are integers, equality holds in (32), where the maximization can be further simplified as follows:

**Theorem 11:**

\[
\text{DoF}(\mathbf{H}_\lambda) = 1 + \sup_{X_1, X_2} d(X_1 + \lambda X_2) - d(X_1 + X_2). \tag{66}
\]

Moreover,

1) \( \text{DoF}(\mathbf{H}_1) = \text{DoF}(\mathbf{H}_1^T) \).

2) For all \( \lambda \neq 0 \),

\[
\text{DoF}(\mathbf{H}_\lambda) = \text{DoF}(\mathbf{H}_{\lambda^{-1}}). \tag{67}
\]

3) \( \text{DoF}(\mathbf{H}_{\frac{1}{2}}) \geq 1 \) with equality if and only if \( \lambda = 0 \) or \( 1 \);

4) \( \text{DoF}(\mathbf{H}_{\lambda}) \leq \frac{3}{\lambda} \) with equality if and only if \( \lambda \) is irrational.

5) For integers \( \lambda = 2, 3, 4, \ldots \),

\[
\frac{3}{2} - \frac{1}{\lambda} \log \lambda \sum_{i=1}^{\lambda-1} \frac{\log \lambda}{i} \leq \text{DoF}(\mathbf{H}_\lambda) \leq \frac{3}{2} - \frac{1}{56 \log \lambda + 16} \tag{69}
\]

Therefore as \( \lambda \to \infty \) on \( \mathbb{N} \),

\[
\text{DoF}(\mathbf{H}_\lambda) = \frac{3}{2} + \Theta \left( \frac{1}{\log \lambda} \right). \tag{70}
\]

For \( \lambda = 2, (68), (69) \) can be sharpened to

\[
1.27 \approx 1 + \log_5 \phi \leq \text{DoF}(\mathbf{H}_2) \leq \frac{10}{7} \approx 1.43, \tag{71}
\]

where \( \phi = \frac{1 + \sqrt{5}}{2} \) denotes the golden ratio, while for \( \lambda = -1 \) we obtain

\[
1.10 \approx 1 + \frac{25 \log 5 - 13 \log 13 - 8}{18 \log 3} \leq \text{DoF}(\mathbf{H}_{-1}) \leq \frac{7}{5}. \tag{72}
\]

**Proof:** Section V-F.

**Remark 5:** Note that (66) deals with the maximal information dimension difference between projections of a production measure to the lines of angle \( \frac{\pi}{5} \) and of an arbitrary angle \( \tan^{-1}(\lambda) \). In view of this interpretation and the fact that \( \tan^{-1}(\lambda^{-1}) = \frac{\pi}{2} - \tan^{-1}(\lambda) \), (67) becomes apparent. Since \( \lambda \) is the channel gain of the direct link for the second user, one might expect at first glance that \( \text{DoF}(\mathbf{H}_{\lambda}) \) should be increasing in \( |\lambda| \). However, (67) shows that this is not the case.
IV. SELF-SIMILAR DISTRIBUTIONS

In this section we discuss self-similar distributions, an important subclass of singular distributions, which play the central role of DoF-achieving input distributions in the proofs shown in Section V. As shown in Theorem 5, input distributions with information dimension one half are necessary for achieving the maximal degrees of freedom $\frac{k}{2}$. Although, in view of (20), equal mixtures of discrete and absolutely continuous components also have information dimension one half, they can achieve at most one degree of freedom as proved in Section III-C. In order to overcome this bottleneck, it is necessary to employ singular input distributions.

A. Definitions

In general, a self-similar distribution is defined as the invariant measure of iterated function systems, see [29], [30]. For our purpose in this paper we focus on the special case of homogeneous self-similar distributions defined via nonlinear iterative function systems as well.

Definition 3: A probability measure $\mu$ is called homogeneous self-similar with similarity ratio $r \in (0, 1)$ if $\mu$ is the distribution of

$$X = \sum_{k \geq 1} W_k r^{k-1},$$

(73)

where $\{W_k\}$ are i.i.d. copies of a real-valued nondeterministic random variable $W$ taking a finite number of values.

The distribution of $X$ is self-similar in view of the fact that

$$X \overset{(d)}{=} rX + W$$

(74)

where $W$ is independent of $X$. It is easy to see that (73) and (74) are equivalent. When $W_k$ is binary, (73) is known as the Bernoulli convolution, an important object in fractal geometry and ergodic theory [31]. As a special case, (73) encompasses distributions defined by $M$-ary expansion with independent digits, in which case $r = \frac{1}{M}$ and $W$ is $\{0, \ldots, M-1\}$-valued. For example, the Cantor distribution [32], supported on the middle-third Cantor set, can be defined by (73) via its ternary expansion, which consists of independent digits taking value 0 or 2 equiprobably.

B. Information Dimension

It is shown in [13] that the information dimension (or dimension of order $\alpha$) of self-similar measures always exists. More general results on the information dimension of self-similar distributions are summarized in [18, Sec. 2.8].

Theorem 12 ([13, Th. 1 and Lemma 2.21]): Let the distribution of $X$ be homogeneous self-similar. Then $d(X)$ exists and satisfies

$$d(X) = \sup_{m \geq 1} \frac{H([X]_m)}{m},$$

(75)

where

$$[x]_m \triangleq (x)_{2^m}.$$

Moreover, $d_a(X)$ exists for all $\alpha \geq 0$.

In spite of the general existence result in [13], there is no known formula for the information dimension of self-similar distributions unless certain separation conditions are satisfied, the most common of which is the open set condition [16, p. 129]: There exists a nonempty bounded open set $U \subset \mathbb{R}$, such that $\bigcup F_i(U) \subset U$ and $F_i(U) \cap F_j(U) = \emptyset$ for $i \neq j$, where we have denoted $F_i(x) = rx + w_j$ for $j = 1, \ldots, m$. The following lemma (proved in Appendix B) gives a sufficient condition for a homogeneous self-similar distribution (or random variables of the form (73)) to satisfy the open set condition.

Lemma 4: Let $X$ be defined in (73) with $0 < r < 1$ and $\mathcal{W} \triangleq \{w_1, \ldots, w_m\} \subset \mathbb{R}$. Then the open set condition is satisfied if

$$r \leq \frac{m(\mathcal{W})}{m(\mathcal{W}) + M(\mathcal{W})},$$

(77)

where the minimum distance and the diameter of $\mathcal{W}$ are defined by

$$m(\mathcal{W}) \triangleq \min_{w_i \neq w_j \in \mathcal{W}} |w_i - w_j|,$$

(78)

and

$$M(\mathcal{W}) \triangleq \max_{w_i, w_j \in \mathcal{W}} |w_i - w_j|$$

(79)

respectively.

The next result gives an upper bound on the information dimension of a self-similar homogeneous measure $\mu$, which holds with equality if the open set condition is satisfied.

Theorem 13: For homogeneous self-similar measure $\mu$ with similarity ratio $r$ defined in Definition 3,

$$d_a(\mu) \leq \frac{H_2(W)}{\log \frac{1}{r}}, \quad \alpha \geq 0.$$  

(80)

Moreover, if $\mu$ satisfies the open set condition, then

$$d_a(\mu) = \frac{H_2(W)}{\log \frac{1}{r}}, \quad \alpha \geq 0.$$  

(81)

Proof: See Appendix C. □

C. Self-Similar Distributions With Half Dimension

In this subsection, we construct a family of scalar self-similar distributions which play an important role in the achievability proof of Theorems 7 and 11. Let $N \geq 2$ be an integer. Let $\mu_N$ be the distribution of (73) with $r = \frac{1}{N^2}$, $w_j = N(j - 1)$ and $p_j = \frac{1}{N}$, $j = 1, \ldots, N$. Then $\mu_N$ is homogeneous self-similar and satisfies the open set condition, in view of Lemma 4. By Theorem 13, $d(\mu_N) = \frac{1}{2}$ for all $N$.

The DoF-achieving distributions in Theorems 7 and 11 are slight variants of $\mu_N$.

Alternatively, $\mu_N$ can be defined by the $N$-ary expansions as follows: Let $X$ be distributed according to $\mu_N$. Then the odd digits in the $N$-ary expansion of $X$ are independent and equiprobable, while the even digits are set to zero. In view of (9), the information dimension of $\mu_N$ is the normalized entropy rate of the digits, which is equal to $\frac{1}{2}$.
D. Convolutions of Self-Similar Measures

Formula (32) for DoF(H) deals with convolutions of input distributions. It is therefore of interest to investigate the behavior of self-similar measures under convolutions. The next lemma shows that convolutions of homogeneous self-similar measures with common similarity ratio are also self-similar (see also [33, p. 1341] and the references therein), but with possible overlaps which might violate the open set conditions. Therefore (81) does not necessarily apply to convolutions.

**Lemma 5**: Let X and X′ be independent with homogeneous self-similar distributions of identical similarity ratio r. Then for any a, a′ ∈ ℝ, the distribution of aX + a′X′ is also homogeneous self-similar with similarity ratio r.

**Proof**: According to (74), X and X′ satisfy X(d) = rX + W and X′(d) = rX′ + W′, respectively, where W and W′ are both finitely valued and {X, W, X′, W′} are independent. Therefore aX + a′X′(d) = (aX + a′X′) + aW + a′W′, which implies the desired conclusion.

V. PROOFS

A. Auxiliary Results

Before proceeding to the proof of Theorem 4, we present the following non-asymptotic bounds on mutual information, which, in view of the alternative definition of information dimension in (10), implies the high-SNR scaling law of mutual information in Theorem 2 as an immediate corollary. This argument is considerably simpler and more general than the original result in [17]. This result also allows us to conclude that the capacity region of the interference channel (1) for non-Gaussian noise can differ only by absolute additive constants from the Gaussian counterpart if the noise has a well-behaved density such as uniform or Laplace distribution (see Remark 6).

**Lemma 6**: Let B(xn, ε) denote the ℓ∞-ball of radius ε centered at xn. For any Xn and any snr > 0,

\[ -8 ≤ −\frac{1}{n}I(X^n, \text{snr}) - \frac{1}{n}E\left[\log \frac{1}{P_X(B(x^n, 1/\sqrt{\text{snr}}))}\right] ≤ 2. \]  

**Proof**: Appendix D.

To gain some insight on Lemma 6, we note that the bracketed term in (82) is close to the mutual information between X and itself contaminated by uniformly distributed noise. Consider the scalar case and let U be uniformly distributed on [−\frac{1}{2}, \frac{1}{2}]. Then the density of Y = X + εU is given by pY(y) = \frac{1}{2} p_X(B(y, \epsilon/2)) and it is easy to verify that I(X; Y) = E[\log \frac{1}{P_Y(B(X, \epsilon/2))}], which is provably close to E[\log \frac{1}{P_X(B(Y, \epsilon))}]. The desired result for Gaussian noise follows from the fact that the Gaussian density can be approximated by piecewise constant functions incurred at most an additive constant difference in the mutual information.

Next we present several auxiliary results regarding the Shannon and Rényi entropy of quantized random variables. The first result states that if X is close to Y and the alphabet of X is well-separated, then the entropy of X is also close to that of Y.

**Lemma 7**: Let X and Y be real-valued discrete random variables such that |X − Y| ≤ ε almost surely. Suppose X takes values in a set X with minimum distance m(X) ≥ δ. Then

\[ H(X) - H(Y) ≤ \log \left(1 + \frac{2ε}{δ}\right). \]  

In particular, if 2ε < δ, then H(X) ≤ H(Y). Moreover, (84) also holds for Rényi entropy H_α for all α ≥ 0.

**Proof**: Inequality (84) follows from

\[ H(X) - H(Y) ≤ H(X, Y) - H(Y) \]

\[ = H(X|Y) = H(X - Y|Y) \]

\[ ≤ \log \left(1 + \frac{2ε}{δ}\right). \]

since conditioned on Y the random variable X takes at most 1 + \left[\frac{2ε}{δ}\right] values. However, (86) in the above argument does not directly generalize to Rényi entropy due to the lack of conditioning property of Rényi entropy. Instead, (84) can be deduced from the following: First note that by Jensen’s inequality, if a distribution Q has support cardinality m, then \[ \sum_u Q^α(u) ≤ m^{1-α} \] (resp. \[ m^{1-α} \]) of α < 1 (resp. α > 1). Therefore,

\[ H_a(X) ≤ H_a(X, Y) \]

\[ = \frac{1}{1 - α} \log \sum_y p^n_Y(y) \sum_x p^n_{X|Y=y}(x) \]

\[ ≤ H_a(Y) + \log \left(1 + \left[\frac{2ε}{δ}\right]\right). \]

The following lemma admits the same proof as Lemma 7.

**Lemma 8**: Let U^n and V^n be integer-valued random vectors. If −B_i ≤ U_i − V_i ≤ A_i almost surely, then

\[ |H_U^n − H_V^n| ≤ \sum_{i=1}^{n} \log(1 + A_i + B_i). \]

which also holds for H_a for all α ≥ 0.

The next lemma shows that the entropy of quantized linear combinations is close to the entropy of linear combinations of quantized random variables, as long as the coefficients are integers. The proof is a simple application of Lemma 8.

**Lemma 9**: For any a^K ∈ ℤ^K and any X^n_1, . . . , X^n_K,

\[ |H\left(\sum_{j=1}^{K} a_j X^n_j\right)_{m} - H\left(\left[\sum_{j=1}^{K} a_j X^n_j\right]_m\right)\]

\[ ≤ n \log \left(2 + 2 \sum_{j=1}^{K} |a_j|\right). \]

where [·]_m is defined in (76).

**Lemma 10**: For any h^K ∈ ℝ^K and any X^n_1, . . . , X^n_K,

\[ H\left(\sum_{j=1}^{K} h_j X^n_j\right)_m ≤ H\left(\left[\sum_{j=1}^{K} h_j X^n_j\right]_m\right) + \sum_{j=1}^{K} H\left(X^n_j\right) + 1, \]

\[ H\left(\sum_{j=1}^{K} h_j X^n_j\right)_m ≤ H\left(\left[\sum_{j=1}^{K} h_j X^n_j\right]_m\right) + \sum_{j=1}^{K} H\left(X^n_j\right) + 1, \]
where \((X^n_j) \triangleq X^n_j - \left\lfloor X^n_j \right\rfloor\) is the fractional part of \(X^n_j\).

**Proof:** Define
\[
V = 2^m \sum_{j=1}^K h_j \langle X^n_j \rangle, \quad Z = 2^m \sum_{j=1}^K h_j \left\lfloor X^n_j \right\rfloor.
\] (94)

Then
\[
H(V + Z) - H(V) \leq H(V + Z|V) \leq 1 + H(Z) \leq 1 + \sum_{j=1}^K H(X^n_j).
\] (95)

\[
1 \leq 1 + \sum_{j=1}^K H(X^n_j).
\] (96)

We state a Diophantine approximation result on manifolds due to Kleinbock and Margulis [34, Th. A] which is useful for proving Theorems 4 and 6. The version below is stated in terms of the Diophantine approximation exponent introduced by Mahler [35], which is defined as follows:

**Definition 4:** Let \(m, n \in \mathbb{N}\). For \(x \in \mathbb{R}^m\), denote by
\[
(x) \triangleq \arg \min_{z \in \mathbb{Z}^m} \|x - z\|_\infty
\] (99)

denote the \(\ell_\infty\)-distance of \(x\) to the nearest integer vector. For \(A \in \mathbb{R}^{m \times m}\), define \(\hat{\omega}(A)\) to be the supremum of \(w > 0\) such that for all sufficiently large real number \(Q > 0\), the set
\[
\{q \in \mathbb{Z}^m \setminus \{0\} : (AQ) \leq Q^{-w}, \|q\|_\infty \leq Q\}
\] (100)
is non-empty.

For any matrix \(A \in \mathbb{R}^{m \times m}\), the Dirichlet approximation theorem [36, Th. VI, p. 13] states that
\[
\hat{\omega}(A) \geq \frac{m}{n},\tag{101}
\]
which holds with equality for Lebesgue-a.e. \(A \in \mathbb{R}^{m \times m}\) due to the Borel-Cantelli lemma [36, Ch. VII]. The Kleinbock-Margulis theorem generalizes this result to manifolds.

**Lemma 11** ([37, p. 823]): Let \(M\) be a non-degenerate manifold of \(\mathbb{R}^m\). Then \(\omega(a) = \frac{1}{k}\) and \(\omega(a^1) = k\) for almost every \(a \in M\).

Lemma 11 deals with **homogeneous** diophantine approximation (approximating zero by linear forms with integer coefficients). In the course of proving Theorem 4, we will also need results on **inhomogeneous** diophantine approximation (approximating an arbitrary point by linear forms), which are connected to the homogeneous case via the so-called **transference theorems** (see [38], [39]). The following transference result is from [39, Lemma 3, p. 750], which is a direct consequence of [38, Th. XVII.B, p. 99] (see also [36, Th. VI, p. 318] for a more general version). Note that this result holds for any fixed matrix \(G\).

**Lemma 12:** Let \(G \in \mathbb{R}^{n \times m}\). Let \(k = 2^{1-m-n((m + n)/2)}\).

Let \(X, Y > 0\). Suppose the set
\[
\{y \in \mathbb{Z}^n \setminus \{0\} : \|G^\dagger y\|_\infty < kX^{-1}, \|y\|_\infty \leq Y\}
\] (102)
is empty. Then for all \(z \in \mathbb{R}^n\), the set
\[
\{x \in \mathbb{Z}^m : \|Gx + z\|_\infty \leq kY^{-1}, \|x\|_\infty \leq X\}
\] (103)
is non-empty.

In the sequel we adopt the following conventional notation of sumsets and dilations:
\[
A + B = \{a + b : a \in A, b \in B\},\tag{104}
\]
\[
\lambda A = \{\lambda a : a \in A\}.	ag{105}
\]

We need the following lemma for minimum distances:

**Lemma 13:** Let \(W \triangleq \{w_1, \ldots, w_m\} \subset \mathbb{R}\) and let \(r > 0\) satisfy (77). Then for any \(n \in \mathbb{N}\),
\[
\text{m}(W + rW + \cdots + r^{n-1}W) \geq r^{n-1}\text{m}(W).
\] (106)

Moreover, \(W + rW + \cdots + r^{n-1}W\) and \(W^n\) are in one-to-one correspondence.

**Proof:** Let \(a^n, b^n\) be distinct elements of \(W^n\). Then \(k \triangleq \min\{i : a_i \neq b_i, 1 \leq i \leq n\}\) is well-defined. Note that \(|\sum_{i=1}^n(a_i - b_i)r^{i-1}| \geq |a_k - b_k|r^{k-1} - \sum_{i=k+1}^n r^{i-1}|a_i - b_i|\). Consequently,
\[
\text{m}(W + rW + \cdots + r^{n-1}W) \geq \min_{0 \leq k \leq n-1} \left\{r^k\text{m}(W) - \text{M}(W) \sum_{i=k+1}^{n-1} r^i\right\} = r^{n-1}\text{m}(W).
\] (107)

To see that the above minimum is attained at \(k = n - 1\), note that the assumption that \(r \leq \frac{\text{M}(W)}{\text{m}(W)}\) implies that \(\frac{r}{m} \geq \frac{1}{r^n}\), which is equivalent to \(r^k\text{m}(W) - \text{M}(W) \sum_{i=k+1}^{n-1} r^i \geq r^{n-1}\text{m}(W)\) for all \(0 \leq k \leq n - 1\). \(\square\)

**B. General Formula**

In this section we prove Theorem 4. As a side product of the proof, we also establish that the capacity regions of the Gaussian interference channel and a particular deterministic interference channel differ by at most universal constants (See Remark 6).

**Step 1:** We start by proving the achievability side of the theorem, which holds for any channel matrix \(H\). Choosing i.i.d. input distribution, we have
\[
\tilde{C}(H, \text{snr}) \geq \sum_{i=1}^K \mathbb{I}(X_i; Y_i)
\]
\[
= \sum_{i=1}^K \left(\sum_{j=1}^K h_{ij} X_j, \text{snr}\right) - \mathbb{I}\left(\sum_{j \neq i} h_{ij} X_j, \text{snr}\right).
\] (108)

In view of Theorem 2, dividing both sides by \(\frac{1}{2} \log \text{snr}\), sending \(\text{snr} \to \infty\) and optimizing over independent \(X_1, \ldots, X_K\) subject to (33) yield (32).

**Step 2:** The rest of the subsection is devoted to proving the converse side of the theorem except possibly for \(H\) whose off-diagonals are non-algebraic and belong to a set of zero Lebesgue measure, regardless of the diagonal entries. To this end, first we state a number-theoretic condition in terms of the Diophantine approximation exponents, which are satisfied in either of the two cases.
Let \( \ell \triangleq K(K-1) \). Given a matrix \( \mathbf{H} \in \mathbb{R}^{K \times K} \), denote by \( \bar{\mathbf{H}} \) the \( \ell \)-dimensional vector consisting of the off-diagonals of \( \mathbf{H} \). Let \( b \in \mathbb{N} \) and define
\[
L_b \triangleq \binom{\ell + b}{b}\]  

(109)

Let
\[
\mathcal{P}_{\ell,b} = \{f_1, \ldots, f_{L_b}\}
\]

(110)
denote the collection of all \( \ell \)-variant monomials of degree at most \( b \), where \( f_1 \equiv 1 \). Define the column vector \( \mathbf{p}_b(\mathbf{H}) = (f_2(\bar{\mathbf{H}}), \ldots, f_{L_b}(\bar{\mathbf{H}}))^T \in \mathbb{R}^{L_b \times 1} \).

Recalling the Diophantine approximation exponent \( \hat{w}(\cdot) \) in Definition 4, we define
\[
w_b(\mathbf{H}) \triangleq \hat{w}(\mathbf{p}_b(\mathbf{H})), \quad w'_b(\mathbf{H}) \triangleq \hat{w}(\mathbf{p}_b(\mathbf{H})^T).
\]

(111)

We will assume the following condition on the cross-channel gains for the converse proof:
\[
w_b(\mathbf{H})w'_{b+1}(\mathbf{H}) \xrightarrow{b \to \infty} 1.
\]

(112)

This assumption warrants some explanation. First of all, as a consequence of (101), for any matrix \( \mathbf{H} \), we have \( w_b(\mathbf{H}) \geq \frac{1}{L_b-1} \) and \( w'_{b+1}(\mathbf{H}) \geq L_{b+1} - 1 \). Note that \( L_{b+1} = \frac{\ell + b + 1}{b+1}L_b \) and \( \ell = K(K-1) \). Then
\[
\liminf_{b \to \infty} w_b(\mathbf{H})w'_{b+1}(\mathbf{H}) \geq \lim_{b \to \infty} L_{b+1} - 1 = 1.
\]

(113)

Therefore the assumption (112) amounts to requiring that the limit sup, hence the limit, are also equal to one.

As mentioned earlier, sufficient conditions for (112) include the following:

1) **Algebraic Case:** For \( i \neq j \), let \( h_{ij} \) be algebraic of degree \( r_{ij} \). Consider the finite set
\[
S = \left\{ \prod_{i \neq j} h_{ij}^{k_{ij}} : k_{ij} \in \{0, \ldots, r_{ij} - 1\}, i \neq j \right\}.
\]

(114)

Then there exists \( S' \subset S \) such that any point in \( S' \) can be written as a linear combination of elements of \( S' \) with rational coefficients. Let \( r \) denote the rational dimension of \( S \), i.e., the smallest cardinality of such \( S' \). Then \( r \leq \prod_{i \neq j} r_{ij} \). Therefore, for all degree \( b \geq \max r_{ij} \), \( \mathbf{p}_b(\mathbf{H}) \) has rational dimension \( r \). As a consequence of the Schmidt subspace theorem, we have [40, p. 2166]:
\[
w_b(\mathbf{H}) = \frac{1}{\min(L_b - 1, r)},
\]

(115)

\[
w'_{b+1}(\mathbf{H}) = \min(L_{b+1} - 1, r).
\]

(116)

Hence \( w_b(\mathbf{H})w'_{b+1}(\mathbf{H}) = 1 \) for all sufficiently large \( b \).

2) **Almost Sure Case:** Since \( \{\mathbf{p}_b(\mathbf{G}) : \mathbf{G} \in \mathbb{R}^{K \times K}\} \) is a non-degenerate manifold of \( \mathbb{R}^{L_b - 1} \) (see [37, p. 822]), Lemma 11 implies that
\[
w_b(\mathbf{H}) = \frac{1}{L_b - 1},
\]

(117)

\[
w'_{b+1}(\mathbf{H}) = L_{b+1} - 1.
\]

(118)

hold for a.e. \( \mathbf{H} \in \mathbb{R}^{K \times K} \), which satisfy (112).

As mentioned in Section III-A, in the three-user case the channel matrix can be reduced to the standard form (34), where the off-diagonals are either one or \( h \). Then the polynomials of the off-diagonals are in fact univariate polynomials of \( h \). Consequently, we can abbreviate \( w_b(\mathbf{H}) \) as \( w_b(h) = w((h, \ldots, h^b)^T) \), which is the exponent for simultaneous approximation of powers of a real number, and \( w'_b(\mathbf{H}) \) as \( w'_b(h) = w((h, \ldots, h^b)) \), which is the exponent for polynomial approximation. These are the central objects studied in [39]–[41]. In particular, by (101) and [41, Ths. 20 and 27], we have \( \frac{1}{b} \leq w'_b(h) \leq \frac{2}{b} \) and \( b \leq w'_b(h) \leq 2b - 1 \). Therefore for any \( h \),
\[
1 \leq \liminf_{b \to \infty} w_b(h)w'_{b+1}(h) \leq \limsup_{b \to \infty} w_b(h)w'_{b+1}(h) \leq 4,
\]

and the condition (112) requires that the limit is equal to one.

**Step 3:** We proceed to show that (32) holds with equality by assuming the condition (112).

By the monotonicity of \( \text{snr} \mapsto \bar{C}(\mathbf{H}, \text{snr}) \), we may restrict the limit in (3) to any increasing unbounded subsequence \( \{\text{snr}_m\}_{m \geq 1} \) as long as the growth is no faster than exponential, i.e.,
\[
\sup_{m \geq 1} \frac{\text{snr}_{m+1}}{\text{snr}_m} < \infty.
\]

(119)

In particular, we choose \( \text{snr}_m = 4^m \). Consequently, we can restrict to this subsequence in the limiting characterization in (39).

Define
\[
g(X^n, \epsilon) \triangleq \mathbb{E} \left[ \log \frac{1}{P_{X^n}(B(X^n, \epsilon))} \right] = \mathbb{E} \left[ \log \frac{1}{\mathbb{P} \{X^n \in B(X^n, \epsilon)\} \mid X^n \} \right],
\]

(120)

where \( \tilde{X}^n \) is an independent copy of \( X^n \). By Lemma 6, for any \( X^n \) and any \( \text{snr} > 0 \),
\[
-8n \leq I(X^n, \text{snr}) - g \left( X^n, \text{snr}^{-\frac{1}{2}} \right) \leq 2n.
\]

(122)

By [10, eq. (294) and (296)] (or [13, Lemma 2.3]), for any \( m \in \mathbb{N} \),
\[
0 \leq H([X^n]_m) - g(X^n, 2^{-m}) \leq 3n.
\]

(124)

(125)

Plugging (122)–(125) into (39), we have the following equivalent limiting characterization of \( \text{DoF}(\mathbf{H}) \):
\[
\text{DoF}(\mathbf{H}) = \limsup_{m \to \infty} \limsup_{n \to \infty} \sup_{X^n_1, \ldots, X^n_K} \frac{1}{nm} \sum_{i=1}^{K} H \left( \sum_{j} h_{ij} X^j_n \right) - H \left( \sum_{j \neq i} h_{ij} X^j_n \right).
\]

(126)

where \( X^n_1, \ldots, X^n_K \) are independent, such that \( H((X_{ik})_1) \leq C \) for some \( C \geq 0 \) and for all \( i = 1, \ldots, K, k = 1, \ldots, n \).

In view of Lemma 10, the right side of (126) is unchanged
if we replace all inputs by their fractional parts. Therefore, it
suffices to consider \([0, 1]\)-valued inputs in the maximization
of (126). Moreover, by the scale-invariance of the DoF(·),
we can assume, without loss of generality, that \(|H|_{\infty} \leq 1,
\text{ i.e., } |h_{ij}| \leq 1 \text{ for all } i, j.

For simplicity, in the remainder of the proof we abbreviate
\(L_b \) and \(L_{b+1} \) by \( L \) and \( L' \), respectively, and suppress
the dependence of \( w_b(H) \) and \( w_{b+1}(H) \) on \( H \). By Definition
4, for any \( \delta > 0 \), there exists \( Q_0 = Q_0(\delta) > 0 \) such that for all
\( Q \geq Q_0 , \)

\[ \{ q \in \mathbb{Z} \setminus \{0\} : (p_b(H)q) \leq \mathcal{Q}^{-(1+\delta)w_b}, \ |q| \leq \mathcal{Q} \} = \emptyset \quad (127) \]

and

\[ \{ q \in \mathbb{Z}^{L'-1} \setminus \{0\} : (p_{b+1}(H)^Tq) \leq \mathcal{Q}^{-(1+\delta)w_{b+1}}, \ |q|_{\infty} \leq Q \} = \emptyset . \quad (128) \]

Next we assume that \(|H|_{\infty} \leq 1 \). Any finite set \( A \) induces
a quantizer \( Q_A : \mathbb{R} \rightarrow A \) defined by

\[ Q_A(x) = \arg \min_{a \in A} |x - a|, \quad (129) \]

with ties broken arbitrarily. For an arbitrary positive integer \( N \),
define the following sets (subsets of a lattice)

\[ \mathcal{V} \triangleq \frac{1}{N} \left\{ \sum_{i=1}^{L} q_i f_i(\tilde{H}) : q_i \in \mathbb{Z}, |q_i| \leq LN, \ \max_{2 \leq i \leq L} |q_i| \leq N \right\}, \quad (130) \]

\[ \mathcal{V}' \triangleq \frac{1}{N} \left\{ \sum_{i=1}^{L'} q_i f_i(\tilde{H}) : q_i \in \mathbb{Z}, \ |q_i| \leq LN, i = 1, \ldots, L' \right\}. \quad (131) \]

As a consequence of (127), we can upper bound the approxi-
mation error of the quantizer \( Q_\mathcal{V} \) as follows. Recall the
constant \( \kappa \) defined in Lemma 12, which, in the special case
of \( m + n = L \), is given by

\[ \kappa = \kappa(b) = 2^{1-L(L!)} . \quad (132) \]

Let \( N_0 = N_0(b, \delta) = \kappa Q_0^{w_b(0+\delta)} \) and \( Q = \left( \frac{N}{\kappa} \right)^{\frac{1}{1+w_{b+1}}} \).
Let \( N \geq N_0 \) and, equivalently, \( Q \geq Q_0 \). Hence for
all integer \(|q| \leq Q \), we have \((p_b(H)q) \leq Q^{-(1+\delta)w_b}\).
Applying Lemma 12 with \( X = N, Y = Q, m = LN - 1, n = 1 \)
yields the following: For all \( \theta \in [0, 1] \), there exists \( x \in \mathbb{Z}^{L-1} \), such that
\(|x|_\infty \leq N \) and

\[ (p_b(H)x + N \theta) \leq \kappa Y^{-1} \leq \kappa(b, \delta) . \quad (133) \]

By the assumption that \(|H|_{\infty} \leq 1 \), we have \( |f_i(\tilde{H})| \leq 1 \)
for all \( i \), hence \((p_b(H)x + N \theta) \leq LN \). Recall that \( (x) = \min_{x \in \mathbb{Z}} |x - a| \).
Therefore (133) implies that there exists \( q \in \mathbb{R}^{L'} \), such that
\(|q| \leq LN, |q_i| \leq N, 2 \leq i \leq L \), such that
\(|N^{-1} \sum_{i=1}^{L} q_i f_i(\tilde{H}) + \theta| \leq (\kappa(b, \delta))^{1+w_{b+1}} \). In other words, for all
\( N \geq N_0 \), we have

\[ \sup_{\theta \in [0, 1]} |Q_\mathcal{V}(\theta) - \theta| \leq (\kappa(b, \delta))^{1+w_{b+1}} . \quad (134) \]

Using (128), we can lower bound the minimum distance
of the set \( \mathcal{V}' \) as follows. For all \( Q \geq Q_0 , (128) \) implies that

for any \( q \in \mathbb{Z}^{L'} \) such that \( |q_i| \leq Q, 2 \leq i \leq L' \), we have
\(|\sum_{i=1}^{L'} q_i f_i(\tilde{H})| \geq Q^{-(1+\delta)w_{b+1}} . \)

\[ \mathcal{V}' - \mathcal{V}' = \frac{1}{N} \left\{ \sum_{i=1}^{L'} q_i f_i(\tilde{H}) : q_i \in \mathbb{Z}, |q_i| \leq 2LN \right\}, \quad (135) \]

then as long as \( N \geq \frac{Q_0}{(1+\delta)w_{b+1}} \), we have

\[ \text{m}(\mathcal{V}') \geq \frac{1}{N}(2LN)^{-(1+\delta)w_{b+1}} . \quad (136) \]

Therefore we conclude that (134) and (136) hold simultane-
ously for all \( N \geq \max\{N_0, \frac{Q_0}{(1+\delta)w_{b+1}}\} \).

By the multi-letter characterization (126), for any \( \epsilon > 0 \) and
sufficiently large \( m \), there exist \( n \) and independent \( X_1^n, \ldots, X_K^n \)
with \( X_j^n = (X_jt)_t \in [0, 1]^n \), such that

\[ \text{DoF}(H) \leq \epsilon + \frac{1}{mn} \sum_{j=1}^{K} H\left( \sum_{j \neq i} h_{ij} X_j^n \right) \]

\[ - H\left( \sum_{j \neq i} h_{ij} X_j^n \right) \] \quad (137)

The goal is to construct single-letter input distributions
\( \{\hat{X}_1, \ldots, \hat{X}_K\} \) based on the multi-letter inputs \( X_1^n, \ldots, X_K^n \),
such that \text{DoF}(\hat{X}_K, H) \) dominates the right side of (137). To
this end, fix \( N \) to be chosen depending on \( m \) and \( b \) later. Set
\( \hat{X}_j = Q_\mathcal{V}(X_jt) \in \mathcal{V} \) for all \( 1 \leq t \leq n \). In view of (134), we have
\(|\|X_j^n - X_j^n\|_{\infty} \leq (\kappa(b, \delta))^{1+w_{b+1}} \).
Now we see that \( H(\sum_{j \neq i} h_{ij} X_j^n) \) and \( H(\|\sum_{j \neq i} h_{ij} X_j^n\|_m) \) are close. First note that

\[ \left| \sum_{j \neq i} h_{ij} X_j^n - \left[ \sum_{j \neq i} h_{ij} X_j^n \right]_{m \infty} \right| \leq 2^{-m} + K(\kappa/b)^{(1+w_{b+1})} \] \quad (138)

Moreover, for all \( t \), \( \sum_{j \neq i} h_{ij} \hat{X}_jt \) takes values in
\( \sum_{j \neq i} h_{ij} \mathcal{V} \subset \mathcal{V}' \). We conclude from (136) that the minimum
distance satisfy

\[ \text{m}\left( \sum_{j \neq i} h_{ij} \mathcal{V} \right) \geq \text{m}(\mathcal{V}') \geq \frac{1}{N}(2LN)^{-(1+\delta)w_{b+1}} . \quad (139) \]

Set

\[ N = \left\lfloor \frac{1}{2L} \left( \frac{2^m}{4K} \right)^{(1+w_{b+1})} \right\rfloor . \quad (140) \]

Applying Lemma 7 yields

\[ \frac{1}{n} H\left( \sum_{j \neq i} h_{ij} \hat{X}_j^n \right) - \frac{1}{n} H\left( \left[ \sum_{j \neq i} h_{ij} X_j^n \right]_m \right) \leq \log \left( 1 + 2^{-m} + K(\kappa/b)^{(1+w_{b+1})} \right) \quad (141) \]

\[ \leq \log \left( 3 + 2KN(\kappa/b)^{(1+w_{b+1})} \right) \quad (142) \]

\[ \leq C(b, \epsilon) + \left( 1 + \delta \right)w_{b+1} - \frac{1}{1+\delta}w_{b} \log N , \quad (143) \]
where (142) follows from (140) since $2^{-m} \leq \frac{1}{\sqrt{N}}(2LN)^{-(1+\delta)w_{b+1}}$, and in (143) we have defined

$$C(b, \delta) \triangleq \log 3 + \log(16K^2L\kappa)^{(1+\delta)w_{b+1}}(2L)^{(1+\delta)w_{b+1}}$$

which only depends on $b$ and $\delta$, since $\kappa$ and $L$ are functions of $b$ only. Analogously, a similar application of Lemma 7 yields

$$\frac{1}{n}H\left(\left(\sum_{j} h_{ij} X_{ij}^{n}\right)_{m}\right) - \frac{1}{n}H\left(\sum_{j} h_{ij} \tilde{X}_{ij}^{n}\right)$$

$$\leq \log\left(1 + 2^{2^{-m} + (K/\kappa)N^{1+\delta \theta_0}}\right)$$

$$\leq C(b, \delta) + \left((1 + \delta)w_{b+1} - \frac{1}{(1 + \delta)w_b}\right)\log N.$$  (144)

Now we construct scalar input distributions which are self-similar (see Section IV) and DoF-achieving. First put

$$W_j = \sum_{k=1}^{n} \tilde{X}_{ij} r^{k-1},$$  (146)

where

$$r = 2^{-m}.$$  (147)

Let $\{W_{ji} : t \geq 0\}$ be a sequence of i.i.d. copies of $W_j$. Set

$$\tilde{X}_j = \sum_{t \geq 0} W_{ji} r^{nt}.$$  (148)

Note that each $\tilde{X}_j$ has a homogeneous self-similar distribution with similarity ratio $r^n$. In view of Theorem 12, $d(\tilde{X}_j)$ exists for all $j$. Moreover, by (148),

$$\sum_{j=1}^{K} h_{ij} \tilde{X}_j = \sum_{t \geq 0} \left(\sum_{j=1}^{K} h_{ij} W_{ji}\right) r^{nt}$$

is also homogeneously self-similar with similarity ratio $r^n$, in view of Lemma 5. Note that $\mathcal{M}(\sum_{i} h_{ij} \mathcal{V}) \leq KM(\mathcal{V}) \leq KM((-2L, 2L)) = 4KL$. Combining (139), (140) and (147), we have

$$r \leq \frac{m(\sum_{i} h_{ij} \mathcal{V})}{2M(\sum_{i} h_{ij} \mathcal{V})}.$$  (150)

Set $\mathcal{W} = \sum_{i=1}^{n} r^{k-1} \mathcal{V}$. Since $\tilde{X}_{ji} \in \mathcal{V}$, we have

$$\sum_{i} h_{ij} W_{ji} \in \sum_{i} h_{ij} \mathcal{W} = \sum_{k=1}^{n} r^{k-1} \left(\sum_{i} h_{ij} \mathcal{V}\right).$$  (151)

In view of (150), applying Lemma 13 yields

$$m\left(\sum_{i} h_{ij} \mathcal{W}\right) \geq r^{n-1} m\left(\sum_{i} h_{ij} \mathcal{V}\right).$$  (152)

On the other hand, the diameter satisfies $\mathcal{M}(\sum_{i} h_{ij} \mathcal{W}) \leq KM(\mathcal{V}) \leq K \sum_{i=1}^{n} r^{k-1} \mathcal{M}(\mathcal{V}) \leq 8KL$. Therefore, in view of (150) and (152), our choice of $r$ in (147) satisfies

$$r^n \leq \frac{m(\sum_{i} h_{ij} \mathcal{W})}{2M(\sum_{i} h_{ij} \mathcal{W})}.$$  (153)

Invoking the sufficient conditions in Lemma 4, we conclude that $\sum_{i} h_{ij} \tilde{X}_j$ satisfies the open set condition. Hence, by Theorem 13, the information dimension is given by

$$d\left(\sum_{i} h_{ij} \tilde{X}_j\right) = \frac{H\left(\sum_{i} h_{ij} W_{ji}\right)}{\log \frac{1}{r}} = \frac{H\left(\sum_{i} h_{ij} W_{ji}\right)}{nm},$$

where (154) follows from the fact that $\sum_{i=1}^{n} r^{k-1} (\sum_{i} h_{ij} \mathcal{V})$ and $(\sum_{i} h_{ij} \mathcal{V})^n$ are in one-to-one correspondence with each other, in view of (146) and Lemma 13. Entirely analogously, we have

$$d\left(\sum_{j \neq i} h_{ij} \tilde{X}_j\right) = \frac{H\left(\sum_{j \neq i} h_{ij} \tilde{X}_j\right)}{nm}.$$  (155)

Assembling (140), (145) and (154), we obtain

$$d\left(\sum_{i} h_{ij} \tilde{X}_j\right) \geq \frac{1}{nm} H\left[\left(\sum_{i} h_{ij} X_{ij}^{n}\right)_{m}\right] - \frac{C(b, \delta)}{m}$$

$$- \left((1 + \delta)w_{b+1} - \frac{1}{(1 + \delta)w_b}\right)\log N.$$  (156)

Combining (140), (143) and (155) yields

$$d\left(\sum_{j \neq i} h_{ij} \tilde{X}_j\right) \leq \frac{1}{nm} H\left[\left(\sum_{j \neq i} h_{ij} X_{ij}^{n}\right)_{m}\right] + \frac{C(b, \delta)}{m}$$

$$+ \left((1 + \delta)w_{b+1} - \frac{1}{(1 + \delta)w_b}\right)\log N.$$  (157)

Recall the notation $\text{dof}$ defined in (31). Plugging (156), (157) into (137) and summing over $i$, we obtain

$$\text{DoF}(\mathcal{H}) \leq \text{dof}(\tilde{X}^{K}, \mathcal{H}) + \epsilon + \frac{2KC(b, \delta)}{m}$$

$$+ \left((1 + \delta)w_{b+1} - \frac{1}{(1 + \delta)w_b}\right)\frac{2K\log N}{m}. $$  (158)

Note that according to (140), $\frac{\log N}{m} \to \infty$ as $m \to \infty$. Sending $m \to \infty$, then $b \to \infty$, and $\delta \downarrow 0$, and finally $\epsilon \downarrow 0$, we have

$$\text{DoF}(\mathcal{H}) \leq \frac{1}{1+(1+\delta)w_{b+1}} \leq 1.$$  (159)

$\text{dof}(X^{K}, \mathcal{H})$ where the last step follows from our assumption (112). This completes the proof of the converse.

Remark 6: Deterministic channel approximation has been employed to determine the capacity region of various network
communication problems within constant gaps, see [42], [43].

As a side product of the proof of Theorem 4, we establish a connection between the Gaussian interference channel

\[ Y^K = \sqrt{\text{snr}} H X^K + N^K \]  

(159)

and the following deterministic Gaussian interference channel,

\[ Y^K = [HX^K]_m \]  

(160)

where \([\cdot]_m\), defined in (76), is taken componentwise, and \(E[X^2] \leq 1\). It can be shown that for \(m \approx \frac{1}{2} \log \text{snr}\), the capacity regions of (1) and (160) differ by at most universal constants. To see this, denote the capacity region of (160) by \(\tilde{C}_m(H)\). Assembling (122)–(125) as well as the multi-letter characterization (36) and (38), we conclude that there exist a universal constant \(c\) (which is at most 22 bits), such that

\[ d_H(C(H, \text{snr}), \tilde{C}_m(H)) \leq c \]  

(161)

holds uniformly for any \(\text{snr} > 0\) and any \(H\), where \(m\) and \(\text{snr}\) are related through \(m = \lfloor \frac{1}{2} \log \text{snr} \rfloor\), and

\[ d_H(A, B) \overset{\Delta}{=} \max \left\{ \sup_{b \in A} \| a - b \|_\infty, \sup_{a \in A} \| a - b \|_\infty \right\}. \]

Moreover, (161) also holds for quantized interference channel with (160) replaced by \(Y^K = [H]_m X^K\) or \(Y^K = [H]_m X^K\) _\_m_. Note that a related one-sided bound was given in [5, Lemma 4].

Furthermore, we note that for non-Gaussian additive noise with well-behaved densities such as uniform or Laplace distribution, the capacity region is only a constant gap away from the Gaussian one. This follows from the multi-letter characterization and the uniform bound of mutual information in Lemma 6 (see also Remark 10 in Appendix D).

Remark 7 (Rational Channel Matrices): For channel matrices with rational entries, the general formula can be proved without appealing to diophantine approximation arguments. Essentially, we can simply construct \(X^K\) based on a quantized version of \(X^K\). An elementary proof is as follows: Since DoF is scale-invariant, we assume that \(h_{ij} \in \mathbb{Z}\) without loss of generality. The gist of constructing \(X_j\) is by concatenating the bits from \(X^n_j\). Let \(L\) be an integer such that \(L > 1 + \max_{1 \leq i \leq k} \log(1 + \sum_{j=1}^{k} |h_{ij}|)\) and let \(M = n(m + L)\). Define

\[ \hat{X}_j = \sum_{l \geq 0} Z_{jl} 2^{-Ml}, \]  

(162)

where \([Z_{jl}]_{l \geq 1}\) are independent copies of

\[ Z_j = \sum_{k=1}^{n} 2^{-L} X_{jk} 2^{-(m + L)(k - 1)}. \]  

(163)

Note that \(Z_j\) is a discrete random variable whose binary expansion has \(M\) bits, formed by concatenating the first \(m\) bits of each \(X_{jk}\) preceded by \(L\) zeros. Therefore, \(Z_j\) takes values in \(2^{-M} [0, \ldots, 2^{m - L} - 1]\). Note that each \(\hat{X}_j\) has a homogeneous self-similar distribution with similarity ratio \(2^{-M}\). In view of Theorem 12, \(d(\hat{X}_j)\) exists for all \(j\). Moreover, by (162),

\[ \sum_{j=1}^{K} h_{ij} \hat{X}_j = \sum_{l \geq 0} \left( \sum_{j=1}^{K} h_{ij} Z_{jl} \right) 2^{-Ml} \]  

(164)

is also homogeneous self-similar with similarity ratio \(2^{-M}\), in view of Lemma 5. It is easy to verify that the distribution of (164) also satisfies the open set condition: Let

\[ \mathcal{W} = 2^{-M} \sum_{j=1}^{K} h_{ij} [0, \ldots, 2^{M-L} - 1] \]  

(165)

denote the alphabet of \(\sum_{j=1}^{K} h_{ij} Z_j\). Since \(h_{ij} \in \mathbb{Z}\), the minimum distance of \(\mathcal{W}\) is given by \(m(\mathcal{W}) = 2^{-M}\), while the diameter is upper bounded by \(M(\mathcal{W}) \leq (2^L - 2^{-M}) \sum_{j=1}^{K} |h_{ij}| \leq \frac{1}{2}\). By Lemma 4, (164) satisfies the open set condition. Applying Theorem 13, we have

\[ M \cdot d \left( \sum_{j=1}^{K} h_{ij} \hat{X}_j \right) \]

\[ = H \left( \sum_{j=1}^{K} h_{ij} Z_j \right) \]  

(166)

\[ = H \left( \sum_{k=1}^{n} \sum_{j=1}^{K} 2^{-L} h_{ij} X_{jk} 2^{-(m + L)(k - 1)} \right) \]  

(167)

\[ = H \left( \sum_{j=1}^{K} h_{ij} X_{ij} 2^{-(m + L)(k - 1)} \right) \]  

(168)

\[ = H \left( \sum_{j=1}^{K} h_{ij} X^n_j \right) \]  

(169)

where (167) is due to (163) and (168) follows from the fact that each \((m + L)\)-bit block in the summation of (167) does not overlap. Since \(h_{ij} \in \mathbb{Z}\), by Lemma 9,

\[ \left| H \left( \sum_{j=1}^{K} h_{ij} [X^n_j]_m \right) - H \left( \left[ \sum_{j=1}^{K} h_{ij} X^n_j \right]_m \right) \right| \]

\[ \leq n \log \left( 2 + 2 \sum_{j=1}^{K} |h_{ij}| \right) \]  

(170)

\[ \leq nL. \]  

(171)

Combining (169) and (171), we have

\[ (m + L) d \left( \sum_{j=1}^{K} h_{ij} \hat{X}_j \right) - \frac{1}{n} H \left( \left[ \sum_{j=1}^{K} h_{ij} X^n_j \right]_m \right) \leq L. \]  

(172)

Similarly, (172) holds if the summation over \(j\) excludes \(i\). Plugging (172) into (137) and summing over \(i\), we have

\[ \text{DoF}(H) \leq \frac{m + L}{N} \sum_{i=1}^{K} d \left( \sum_{j=1}^{K} h_{ij} \hat{X}_j \right) - d \left( \sum_{j \neq i} h_{ij} \hat{X}_j \right) \]

\[ + \frac{2KL}{m}. \]  

(173)

By the arbitrariness of \(\epsilon\) and \(m\), the proof for rational \(H\) is complete.
C. The Upper Bound

Proof of Theorem 5: Counting in different ways, we have

\[ \text{2 dof}(X^K, \mathbf{H}) = \sum_{i=1}^{K} \left( \sum_{j} h_{ij} X_j \right) + d \left( \sum_{j} h_{ij} X_j \right) - d \left( \sum_{j \neq i} h_{ij} X_j \right) \leq K, \]

where (175) follows from (13)–(15), and (176) is due to (16). If (174)–(176) hold with equality, then for all \( i \), we have

\[ 1 = d \left( \sum_{j} h_{ij} X_j \right) = d \left( \sum_{j \neq i} h_{ij} X_j \right) + d(X_i) = d(X_{\pi(i)}) + d(X_i). \]

Since \( \pi \) is cyclic, the system of linear equations

\[ d(X_{\pi(i)}) + d(X_i) = 1, \quad i = 1, \ldots, K \]

has a unique solution: \( d(X_i) = \frac{1}{2} \) for all \( i \), yielding the desired (58)–(60).

Remark 8: Based on the fact that, with nonzero off-diagonal entries, the degrees of freedom of the two-user IC is at most one, (57) can be shown, alternatively, using a similar method to [2, Proposition 1]. Consider user \( i \) and user \( \pi(i) \). Since \( h_{\pi(i),i} \neq 0 \), we have

\[ R_i + R_{\pi(i)} \leq \frac{1}{2} \log \text{snr} + o(\log \text{snr}) \]

even in the absence of interference from users other than \( i \) and \( \pi(i) \). Summing (181) over \( i \) gives (57).

Proof of Theorem 10: To prove (64), it is equivalent to show that for any \( d^K \in \text{Doff} \{\mathbf{H}\} \) and any probability vector \( w^K \),

\[ \langle d^K, w^K \rangle \leq \max \left\{ w_1, \ldots, w_K, \frac{1}{2} \right\}. \]

Repeating (174)–(176), we have

\[ \langle d^K, w^K \rangle \leq \frac{1}{2}. \]

Therefore (182) is indeed satisfied if \( w_i \leq \frac{1}{2} \) for all \( i \).

Otherwise, there exists a unique \( w_1 \), say \( w_1 \) for simplicity, such that

\[ w_1 > \frac{1}{2} > \sum_{j \neq 1} w_j. \]

Then for any \( X^K \),

\[ \sum_{i=1}^{K} w_id \left( \sum_{j} h_{ij} X_j \right) - w_id \left( \sum_{j \neq i} h_{ij} X_j \right) \leq w_1 - w_1d \left( \sum_{j=2}^{K} h_{1j} X_j \right) + w_2d(X_i) \]

\[ \leq w_1 - \sum_{i=2}^{K} w_i \left[ d \left( \sum_{j=2}^{K} h_{1j} X_j \right) - d(X_i) \right], \]

\[ \leq w_1, \]

where (185)–(187) follow from (16), (184) and (15), respectively. Note that the necessary and sufficient condition for (187) to hold with equality is \( d(X_j) = 0 \) for all \( j \neq 1 \).

The proof of the almost sure equality of (64) is analogous to the proof of Theorem 6 in Section V-D.

D. Almost Sure Results on Degrees of Freedom

Proof of Theorem 6: The outline of the proof is as follows: According to Theorem 4, we construct input distributions depending only on the off-diagonal entries of \( \mathbf{H} \) such that (58) and (59) are (approximately) satisfied. Using the projection theorem [12, Th. 1] with respect to the diagonal entries, the condition (60) (approximately) holds for Lebesgue-almost every diagonal elements.

Let \( \ell \equiv K(K - 1) \). Recall that \( \mathbf{h} \) denotes the \( \ell \)-dimensional vector formed by the off-diagonal entries of \( \mathbf{H} \). Let \( N, b \in \mathbb{N} \) and \( r \in (0, 1) \) be parameters to be specified later. We construct an input distribution \( \mu \) which depends on \( \mathbf{h} \) and parameters \((N, b, r)\) as follows: Recall the set of monomials \( \mathcal{P}_{\ell,b} = \{ f_1, \ldots, f_L \} \) defined in (110), where \( L = L_b \) is defined in (109). Let \( \mu \) be the homogeneous self-similar distribution of similarity ratio \( r \) as per Definition 3, with \( p_j = \frac{1}{N^r} \) and \( \mathcal{W} = \{ w_1, \ldots, w_m \} \) given by

\[ \mathcal{W} \triangleq \frac{1}{N} \left\{ \sum_{i=1}^{L} q_i f_i(\mathbf{h}) : q_i \in \{0, \ldots, N - 1\} \right\}, \]

i.e., the set of all \( \ell \)-variate polynomials evaluated at \( \mathbf{h} \), with degrees not exceeding \( b \) and coefficients valued in the arithmetic progression \( \frac{1}{N}[0, \ldots, N - 1] \). Let \( X \) be distributed according to \( \mu \). Then \( X \) is of the self-similar form in (73), where \( \mathcal{W}_b \) are i.i.d. and equiprobable on \( \mathcal{W} \).

Next we choose \( r \) properly so that \( \mu \) satisfies the open set condition, which renders the formula (81) applicable. This part is analogous to the second step in the converse proof of Theorem 4. In view of Lemma 4, we proceed by giving lower and upper bounds on the minimum distance \( m(\mathcal{W}) \) and the diameter \( M(\mathcal{W}) \), which are defined in (78) and (79) respectively.

Recall that \( \mathbf{p}_b(\mathbf{H}) = (f_2(\mathbf{h}), \ldots, f_L(\mathbf{h}))^T \in \mathbb{R}^{L-1} \) denotes the vector of non-constant monomials of \( \mathbf{h} \) with degree not exceeding \( b \). By Lemma 11, \( w(\mathbf{p}_b(\mathbf{H})^T) = L - 1 \) for a.e. \( \mathbf{h} \).

Therefore, for a.e. \( \mathbf{h} \), there exists \( N_0 = N_0(b, \mathbf{h}) > 0 \) such
that for all \(N \geq N_0\),
\[
\{q \in \mathbb{Z}^{L-1}\setminus\{0\} : (p_b(H)^T q) \leq N^{-2L}, \quad \|q\|_\infty \leq N\} = \emptyset,
\]
(189)
Therefore as long as \(N \geq N_0\), we have
\[
m(W) > N^{-2L},
\]
(190)
which implies that \(|W| = N^L\) in particular. On the other hand,
\[
M(W) \leq 2 \sum_{i=1}^L |f_i(h)|.
\]
(191)
Defining
\[
C = C(b, h) \triangleq 4 \sum_{i=1}^L |f_i(h)|,
\]
(192)
we choose
\[
r = \frac{1}{CN^{2L}},
\]
(193)
which, in view of (190)–(192), satisfies
\[
r \leq \frac{m(W)}{2M(W)} \leq \frac{m(W)}{m(W) + M(W)}.
\]
(194)
By Lemma 4, \(\mu\) satisfies the open set condition. By Theorem 13, for all \(\alpha \geq 0\), we have
\[
d_\alpha(\mu) = \frac{\log |W|}{\log r} = \frac{L \log N}{2L \log N + \log C}.
\]
(195)
Next we upper bound the information dimension of the interference. Let \(X_1, \ldots, X_K\) be i.i.d. according to \(\mu\). By Lemma 5, the distribution of \(\sum_{j \neq i} h_{ij} X_j\) is homogeneous self-similar with similarity ratio \(r\). By Theorem 13,
\[
d\left(\sum_{j \neq i} h_{ij} X_j\right) \leq \frac{H}{\log r} \left(\sum_{j \neq i} h_{ij} W_j\right),
\]
(196)
where the \(W_j\)'s are i.i.d. copies of \(W\). Next we upper bound the alphabet size of \(\sum_{j \neq i} h_{ij} W_j\): In view of (188), we have
\[
\sum_{j \neq i} h_{ij} W \subseteq \left\{ \frac{1}{N} \sum_{i=1}^L q_i f_i(h) : q_i \in \{0, \ldots, K(N-1)\} \right\},
\]
(197)
where \(L' = L_{b+1} = \frac{\ell + b + 1}{\ell} \) per (109). Combining (196), (197) and (193) gives
\[
d\left(\sum_{j \neq i} h_{ij} X_j\right) \leq \frac{L' \log(K+1) + \log N}{2L \log N + \log C}.
\]
(198)
By the dimension preservation result in Lemma 3, for any fixed \(h\) and any \(\alpha \in (1, 2)\),
\[
d_\alpha\left(\sum_j h_{ij} X_j\right) = \min \left\{ 1, d_\alpha(X_i) + d_\alpha\left(\sum_{j \neq i} h_{ij} X_j\right) \right\}
\]
(200)
holds for Lebesgue-almost every \(h_{ii}\). Then
\[
d\left(\sum_j h_{ij} X_j\right) \geq d_\alpha\left(\sum_j h_{ij} X_j\right)
\]
(201)
\[
= \min \left\{ 1, d_\alpha(X_i) + d_\alpha\left(\sum_{j \neq i} h_{ij} X_j\right) \right\}
\]
(202)
\[
\geq \min \{1, 2d_\alpha(\mu)\} = \frac{2L \log N}{2L \log N + \log C},
\]
(203)
where
- (201): by the monotonicity of \(\alpha \mapsto d_\alpha\);
- (202): by (200);
- (203): by (15);
- (204): by (195).
Subtracting (199) from (204) then summing over \(i\), we conclude that
\[
\text{DoF}(H) \geq 2LK \log N - L' K (\log(K+1) + \log N) \over 2L \log N + \log C
\]
(205)
holds for almost every \(H\) and every \(b, N \in \mathbb{Z}\). Recall that \(C\) is given by (192), which does not depend on \(N\). Therefore, sending \(N \to \infty\) first and then \(b \to \infty\) in (205), we have
\[
\text{DoF}(H) = \lim_{b \to \infty} \frac{K}{2} - \frac{2(b+1)K}{\ell + b + 1} \left(\frac{2(b+1)}{\ell + b + 1}\right)
\]
(206)
\[
= K/2,
\]
(207)
where in (206) we used the fact that \(L' = \frac{\ell + b + 1}{\ell} L\).

**Remark 9:** A variant of the Kleinbock-Margulis diophantine approximation result (Lemma 11) is also used in [3] to construct a communication scheme, called real interference alignment, that achieves \(\frac{K}{2}\) degrees of freedom almost surely. The diophantine approximation theorem is then invoked to lower bound the minimum distance in the multi-layered signal constellation. In a similar spirit, in our proof the role of Lemma 11 is to prove that \(\mu\) satisfies the open set condition almost surely, which leads to (195), i.e., the information dimension of each user is close to \(1/2\). To upper bound the information dimension of the interference, no separation conditions are needed.

### E. Channel Matrices With Rational Entries

**Proof of Theorem 8:** Since the off-diagonals of \(H\) are algebraic, (32) holds with equality. Without loss of generality, we assume that \((i, j, k) = (2, 1, 3)\), i.e., \(h_{21}, h_{22}, h_{31}, h_{32}\) are...
non-zero rationals. Then

\begin{align*}
2 \text{ dof}(X^3, H) \\
\leq d(h_{11}X_1 + h_{13}X_3) - d(X_3) + d(h_{21}X_1 + h_{22}X_2) - d(X_1) + d(h_{32}X_2 + h_{33}X_3) - d(X_2) \\
+ \sum_{i=1}^3 d \left( \sum_j h_{ij} X_j \right) - d \left( \sum_{j \neq i} h_{ij} X_j \right) \\
\leq d(h_{11}X_1 + h_{13}X_3) - d(h_{21}X_1 + h_{23}X_3) \\
+ d(h_{21}X_1 + h_{22}X_2) - d(h_{31}X_1 + h_{32}X_2) \\
+ 2 + [d(h_{32}X_2 + h_{33}X_3) - d(X_2) - d(X_3)] \\
+ [d(h_{11}X_1 + h_{12}X_2 + h_{13}X_3) - d(h_{12}X_2 + h_{13}X_3) - d(X_1)] \\
< 3,
\end{align*}

where

- (208): by (17);
- (209)–(212): \( d \left( \sum_j h_{2j} X_j \right) \leq 1 \), \( d \left( \sum_j h_{3j} X_j \right) \leq 1 \);
- (213): the dimension differences in (209) and (210) are \( \leq \frac{1}{2} \) and \( \leq \frac{1}{2} \) respectively, in view of (26) and Theorem 3. By (16), the bracketed terms in (211) and (212) are both negative. □

\[ F. \text{ Lower-Triangular Channel Matrix} \]

\textbf{Proof of Theorem 11:} In view of the definition of \( H_{ij} \) in (65), we have

\begin{align*}
\text{dof}(X^3, H_{ij}) \\
\geq d(X_1) + d(X_1 + \lambda X_2) - d(X_1) \\
+ d(X_1 + X_2 + X_3) - d(X_1 + X_2) \\
= d(X_1 + X_2 + X_3) + d(X_1 + \lambda X_2) - d(X_1 + X_2). 
\end{align*}

Note that the convolution between an absolutely continuous and an arbitrary distribution is absolutely continuous [32]. To maximize (215), choosing an absolutely continuous \( P_{X_3} \) yields \( d(X_1 + X_2 + X_3) = 1 \), regardless of \( P_{X_1} \) or \( P_{X_2} \). This completes the proof of (66).

It remains to prove the listed properties of DoF(H_{ij}):

1) The proof is completely analogous to the derivation of (66) by exchanging \( X_1 \) and \( X_3 \).
2) Identity (67) follows from the scale-invariance of information dimension and exchanging \( X_1 \) and \( \lambda X_2 \).
3) Using (26) instead of Theorem 5, which does not apply to \( H_{ij} \), we obtain the upper bound

\[ \text{DoF}(H_{ij}) \leq \frac{3}{2}. \] (216)

Next we prove equality in (216) for irrational \( \lambda \).

By Dirichlet’s theorem [44, Th. 1A, p. 34], there exists arbitrarily large \( N \in \mathbb{N} \) and \( p \in \mathbb{Z} \) such that \( (p, N) = 1 \) and

\[ \left| \lambda - \frac{p}{N} \right| \leq \frac{1}{2N^2}. \] (217)

Let \( X_1 \) and \( X_2 \) be i.i.d. homogeneous self-similar given by

\begin{align*}
X_1 &= \sum_{i \geq 1} Z_i r^i, \\
X_2 &= \sum_{i \geq 1} W_i r^i,
\end{align*}

where \( \{Z_i\} \) and \( \{W_i\} \) are both i.i.d. and equiprobable on \( N\{0, \ldots, N - 1\} \) and

\[ r = \frac{1}{4N^2(1 + |\lambda|)}. \] (220)

Then \( X_1 + X_2 \) and \( X_1 + \lambda X_2 \) both have homogeneous self-similar distributions with similarity ratio \( r \). Now we verify that they both satisfy the open set condition. Since \( X_1 + X_2 = \sum_{i \geq 1} (Z_i + W_i)r^i \), where \( Z_i + W_i \) takes values in \( N\{0, 2(N - 2)\} \). Applying Lemma 4, we conclude that \( X_1 + X_2 \) satisfies the open set condition, and, in view of Theorem 13,

\[ d(X_1 + X_2) = \frac{H(Z_1 + W_1)}{\log \frac{1}{r}} \leq \frac{\log(2N - 1)}{2 \log N + \log(4(1 + |\lambda|))}. \] (221)

Next we turn to \( X_1 + \lambda X_2 = \sum_{i \geq 1} (Z_i + W_i)r^i \), where \( Z_i + \lambda W_i \) takes values in

\[ \forall \mathcal{V} = N\{0, N - 1\} + \lambda\{0, N - 1\}. \] (222)

Therefore the diameter satisfies

\[ \mathcal{M}(\mathcal{V}) \leq (|\lambda| + 1)N(N - 1). \] (223)

To lower bound the minimum distance, let \( a, b \in \{-N - 1, \ldots, N - 1\} \) not all zero. Since \( p \) and \( N \) are coprime and \( |b| \leq N - 1 \), we have \( |aN + bp| \geq 1 \). Put \( \epsilon = \lambda - \frac{p}{N} \). Then \( |\epsilon| \leq \frac{1}{2\sqrt{N}} \) in view of (217), and

\[ |a + b\lambda| = \left| \frac{aN + bp}{q} + b\epsilon \right| \geq \left| \frac{aN + bp}{N} \right| - |b||\epsilon| \] (224)

\[ \geq \frac{1}{N} - \frac{N - 1}{2N^2} = \frac{N + 1}{2N^2}. \] (225)

Therefore \( \mathcal{M}(\mathcal{V}) \geq \frac{N + 1}{2N^2} \). Consequently, \( r \leq \frac{\mathcal{M}(\mathcal{V})}{\mathcal{M}(\mathcal{V})} \) in view of (220). By Lemma 4, \( X_1 + \lambda X_2 \) also satisfies the open set condition. Applying Theorem 13 yields

\[ d(X_1 + \lambda X_2) = \frac{H(Z_1 + \lambda W_1)}{\log \frac{1}{r}} \leq \frac{2 \log N}{2 \log N + \log(4(1 + |\lambda|))}. \] (226)

where the last inequality follows from the irrationality of \( \lambda \): Since \( Z_1 \) and \( W_1 \) are integer-valued, \( Z_1 + \lambda W_1 \) and the pair \((Z_1, W_1)\) are in one-to-one correspondence, hence \( H(Z_1 + \lambda W_1) = H(Z_1, W_1) \). Since \( N \) can be chosen arbitrarily large, plugging (221) and (226) into (66) and sending \( N \to \infty \), we obtain the equality of (216).
For rational $\lambda = \frac{p}{q}$, applying Theorem 3 we obtain
\[
\text{DoF}(H_2) \leq \frac{3}{2} - \epsilon(p, q) < \frac{3}{2},
\]
(227)

4) If $\lambda = 0$, then $d(X_1) \leq d(X_1 + X_2)$ hence $\text{DoF}(H_0) \leq 1$, with equality attained by choosing $X_2 = 0$. Next we assume that $\lambda \neq 0$. We prove that the range of the mapping
\[ (P_{X_1}, P_{X_2}) \mapsto d(X_1 + \lambda X_2) - d(X_1 + X_2) \] 
(228)
is symmetric. To see this, assume that $d(X_1 + \lambda X_2) - d(X_1 + X_2) = \gamma$. Then $d(X'_1 + \lambda X'_2) - d(X'_1 + X'_2) = \gamma$ with $X'_1 = \lambda X_2$ and $X'_2 = X_1$. Consequently the supremum in (69) is non-negative, with equality if and only if $d(X_1 + \lambda X_2) = d(X_1 + X_2)$ for all independent $X_1$ and $X_2$, which, as we show next, is equivalent to $\lambda = 1$.

Assuming $\lambda \neq 1$ or 0, we show that $\text{DoF}(H_3) > 1$. In view of (67), we may assume that $|\lambda| \leq 1$. Consider the following two cases separately:

(i) $\lambda \in (-1, 0) \cap (0, 1)$: In view of the fact that $\text{DoF}(H_4) = \frac{3}{2}$ for irrational $\lambda$ proved previously, it suffices to consider rational $\lambda = \frac{p}{q}$, where $q \in \mathbb{N}$ and $p \in \{1, \ldots, q - 1\}\setminus\{0\}$. Observing that
\[
\sup_{X_1, X_2} d\left(\frac{X_1 + \frac{p}{q} X_2}{q}ight) \leq d\left(X_1 + X_2\right)
\]
\[
\sup_{X_1, X_2} d(qX_1 + X_2) - d(pX_1 + X_2),
\]
(229)
we construct $X_1$ and $X_2$ such that
\[
d(pX_1 + X_2) = 1,
\]
(230)
d\[
d(X_1 + X_2) < 1,
\]
(231)
which, in view of (69), implies that $\text{DoF}(H_4) > 1$.

To this end, let
\[
X_1 = \sum_{i \geq 1} U_i q^{-2i},
\]
(232)
\[
X_2 = \sum_{i \geq 1} V_i q^{-2i},
\]
(233)
where $\{(U_i, V_i)\}$ are i.i.d. copies of $(U, V)$, which are i.i.d. equiprobable on $\{0, \ldots, q-1\}$. It is straightforward to check that $X_1 + q X_2$ is uniformly distributed on $[0, 1]$, hence (230) holds in view of Theorem 1. Next we establish (231). According to Lemma 5, $X_1 + p X_2 = \sum_{i \geq 1} (U_i + pV_i) q^{-2i}$ has a homogeneous self-similar distribution with similarity ratio $q^{-2}$. The alphabet of $U_i + pV_i$ is $\mathcal{W} = \{0, \ldots, q - 1\} + p\{0, \ldots, q - 1\}$ respectively. Therefore the minimum distance and diameter satisfy $\mathbb{M}(\mathcal{W}) = 1$ and $\mathbb{M}(\mathcal{W}) \leq (\lfloor p \rfloor + 1)(q - 1) \leq q(q - 1)$. By Lemma 4, the open set condition is satisfied for $pX_1 + X_2$. By Theorem 13, $d(pX_1 + X_2) = \frac{H(pU + V)}{2 \log q}$. It remains to show that $H(pU + V) < 2 \log q$, which is equivalent to that $(U, V) = f(pU + V)$ for some deterministic function $f$. This is an obvious contradiction, because, in view of the fact that $|p| \leq q - 1$, there exist distinct pairs $(u, v)$ and $(u', v')$ in $\{0, \ldots, q - 1\}^2$ such that $pu + v = pu' + v'$.

(ii) $\lambda = -1$: let $X_1 = \sum_{i \geq 1} U_i 3^{-i}$ and $X_2 = \sum_{i \geq 1} V_i 3^{-i}$, where $\{(U_i, V_i)\}$ are i.i.d. copies of $(U, V)$. Here $U$ takes values 0 or 1 with probability $p$ and $1 - p$ respectively, while $V$ is an independent copy of $U$. It is easy to check that $X_1 - X_2$ and $X_1 + X_2$ both satisfy the open set condition. Therefore $d(X_1 - X_2) = \frac{H(U - V)}{\log 3}$ and $d(X_1 + X_2) = \frac{H(U + V)}{\log 3}$.

It is straightforward to check that if, for example, $p = \frac{1}{2}$, then $H(U - V) - H(U + V) \approx 0.11$. Hence $\text{DoF}(H_{-1}) > 1.10$, which is the lower bound in (72).

5) To achieve (71) for $\lambda = 2$, consider the following singular input distributions:
\[
X_1 = \sum_{i \geq 1} U_i 6^{-i},
\]
(234)
\[
X_2 = \sum_{i \geq 1} V_i 6^{-i},
\]
(235)
where $\{(U_i, V_i)\}$ are i.i.d. copies of $(U, V)$, with $U$ and $V$ independently valued on $\{0, 1\}$ and $\{0, 1, 2\}$ respectively. Then
\[
X_1 + 2X_2 = \sum_{i \geq 1} (U_i + 2V_i) 6^{-i},
\]
(236)
\[
X_1 + 2X_2 = \sum_{i \geq 1} (U_i + 2V_i) 6^{-i},
\]
(237)
where $U + V$ and $U + 2V$ are valued on $\{0, 1, 2, 3\}$ and $\{0, 1, 2, 3, 4, 5\}$ respectively. By the entropy-rate definition of information dimension in (9), we have
\[
d(X_1 + X_2) = \frac{H(U + V)}{\log 6},
\]
(238)
\[
d(X_1 + 2X_2) = \frac{H(U + 2V)}{\log 6},
\]
(239)
\[
= \frac{H(U) + H(V)}{\log 6}.
\]
(240)

Next we maximize
\[ H(U) + H(V) - H(U + V) = H(U|U + V). \]
(241)

It can be shown that $H(U|U + V)$ is concave in $P_U$ and $P_V$ individually. Moreover, $H(U|U + V)$ is invariant if $U$ is replaced by $1 - U$ or $V$ replaced by $2 - V$. Therefore the optimal $U$ and $V$ are symmetric. In particular, $U$ is equiprobable Bernoulli. Let $P[V = 0] = q$. Maximizing $H(U|U + V)$ over $0 \leq q \leq \frac{1}{2}$, we obtain the optimal $q = \frac{\phi'}{\phi}$ and $H(U|U + V) = \log \phi$, which, in view of (238)-(240), gives (71). The upper bound in (71) follows from (29) (see Remark 11 in Appendix E).

For $\lambda \geq 2$, replacing $6$ by $\lambda^2$ in (235) and letting $U$ and $V$ be equiprobable on $\{0, \ldots, \lambda - 1\}$, we have
\[
d(X_1 + \lambda X_2) = \frac{H(U) + H(V)}{2 \log \lambda} = 1
\]
(242)
and
\[
d(X_1 + X_2) = \frac{H(U + V)}{2 \log \lambda} \tag{243}
\]
\[
= \frac{1}{2} + \frac{1}{\lambda \log \lambda} \sum_{i=1}^{\lambda-1} i \log \frac{\lambda}{i}. \tag{244}
\]
Therefore
\[
\text{DoF}(H_i) \geq \frac{3}{2} - \frac{1}{\lambda \log \lambda} \sum_{i=1}^{\lambda-1} i \log \frac{\lambda}{i} \tag{245}
\]
\[
= \frac{3}{2} - \frac{1 + o(1)}{4 \log \lambda}, \tag{246}
\]
as \( \lambda \to \infty \), where we have used \( \int_0^1 x \log \frac{1}{x} \, dx = \frac{\gamma}{2} \log e \).

On the other hand, using Theorem 3, we have
\[
\text{DoF}(H_i) \leq \frac{3}{2} - \frac{1}{56 \log \lambda + 16} \tag{247}
\]
which, together with (246) implies that
\[
\text{DoF}(H_i) = \frac{3}{2} + \Theta \left( \frac{1}{\log \lambda} \right), \tag{248}
\]
which approaches the upper bound \( \frac{3}{2} \) in (240) at the speed of \( \frac{1}{\log \lambda} \) as \( \lambda \to \infty \). \( \square \)

VI. CONCLUSIONS

Under very general conditions on the cross-channel gains, we have given a formula for the degrees of freedom of the \( K \)-user interference channel in terms of a single-letter optimization (over \( K \) independent scalar inputs) of a linear combination of information dimensions. Moreover, this optimization always yields a lower bound on the degrees of freedom for any channel matrix. Although information dimension is defined as a limit (c.f. three equivalent definitions in Section II-A), we have given a profusion of examples of its computation. This is in sharp contrast to the multi-letter limiting characterization in (39) whose computation poses formidable challenges. Other benefits of the information dimension approach include:

- It shows that degrees of freedom and degrees of freedom region, i.e., the respective limit in (3) and (62), always exist for any channel \( H \) satisfying the conditions of Theorem 4, thus excluding the possibility of any oscillatory behavior in \( \log \text{SNR} \), as was the case in the high-SNR asymptotics of minimum mean-square error for certain input distributions (see [11, Sec. IV-E]).

- A quantitative connection between the Gaussian interference channel and a deterministic interference channel is established. In particular, in Remark 6 it is shown that the capacity regions of (1) and (160) differ by at most universal constants.

- It recovers and improves known results with unified proofs, many of which are consequences of the calculus of information dimension. For instance, (27) and (69) are pure additive-combinatorial results, whereas the counterpart in [5] relies on additional techniques of deterministic channel approximation.

- It explains the sensitivity of the degrees of freedom to the rationality of \( H \), because the behavior of information dimension under projection depends crucially on whether the coefficients are rational (see Section II-D).

- The power constraint can be significantly weakened. In fact the same degrees of freedom holds even if \( \mathbb{E}[X^2] = \infty \), as long as (7) is satisfied.

- It reveals a connection between interference channels and the sumset theory for Shannon entropy [45]. In particular, the converse results, based on the new entropy inequalities given in Theorem 3 and Appendix E, yield tighter upper bounds than those obtained in [5] using additive combinatorial inequalities for set cardinality. For instance, (71) improves the upper bound 1.488 in [5, p. 4945]. A more prominent example is the prototypical three-user interference channel with lower-triangular channel matrices considered in Section III-G. Applying the upper bound in [5, Lemmas 11 and 7] yields \( \frac{2}{3} - \text{DoF}(H_i) = \Omega \left( \frac{1}{\sqrt{n}} \right) \) as \( \lambda \to \infty \), while Theorem 11 reveals that the gap \( \frac{2}{3} - \text{DoF}(H_i) \) in fact behaves as \( \Theta \left( \frac{1}{\log \lambda} \right) \).

The degrees-of-freedom formula in this paper is obtained by single-letterizing the limiting expression and analyzing the high-SNR behavior of the mutual information. Therefore this method does not directly provide explicit interference alignment schemes. However, for those channel matrices whose DoF is unknown (e.g., rationals), by appropriately choosing the input distribution the single-letter formula often yields much better achievable rates than those that have been reported with explicit constructive schemes, e.g., (71) improves the lower bound \( \frac{2}{3} + \log_3 \lambda \approx 1.19 \) in [5, p. 4945]. This, in turn, incentivizes the search for better constructive schemes.

In this paper we have focused on real-valued single-antenna interference channels. In a follow-up to this work, the degrees of freedom of the MIMO interference channel has been investigated in [26]. In particular, it has been shown that the same single-letter formula in Theorem 4 also applies to the MIMO case. In the SISO case, we have shown that the optimal scalar input distribution must be singular with half dimension. In contrast, the singularity of the input distribution in the multi-dimensional case turns out to be much more friendly. A simple example, closely related to the subspace based interference alignment results of Cadambe-Jafar [4], is a distribution concentrated on a linear subspace.

Although the right side of (39) can be single-letterized, solving the maximization problem in (32) is a non-trivial task in general. In fact, computing the information dimension of linear combinations of independent random variables is, in general, challenging. However, as we have illustrated with several examples, this does not necessarily imply that computing the supremum is necessarily hard. For instance, Theorem 11 succeeds in solving for the supremum even though determining the dimension is a long-standing open problem if the coefficients of the linear combination are irrational (see the Furstenberg conjecture [46, p. 194]).

APPENDIX A

PROOF OF LEMMA 1

Proof: We only prove (17). The rest of the results can be found in [8] or [10, Sec. II]. Let \( U, V, W \) be integer valued.
By the data processing theorem, we have $I(V; U + V + W) \leq I(V; V + W)$, i.e.,

$$H(U + V + W) - H(V + W) \leq H(U + W) - H(W). \quad (249)$$

Replacing $U, V, W$ by $[X]_m, [Y]_m, [Z]_m$ respectively and applying Lemma 9, we have

$$H([X + Y + Z]_m) - H([Y + Z]_m) \leq H([X + Z]_m) - H([Z]_m) + \log 3 + 2. \quad (250)$$

Dividing both sides by $m$ and letting $m \to \infty$ yields (17). □

**APPENDIX B**

**PROOF OF LEMMA 4**

**Proof:** Let $U$ be an open interval $(a, b)$. Then

$$F_j(U) = (w_j + ra, w_j + rb). \quad (251)$$

Therefore,

$$\bigcup_j F_j(U) \subset U \iff \min_j w_j + ra \geq a \text{ and } \max_j w_j + rb \leq b \quad (252)$$

and for $i \neq j$,

$$F_i(U) \cap F_j(U) = \emptyset \iff |w_i - w_j| \geq r(b - a). \quad (253)$$

Therefore, there exist $a < b$ that satisfy (252) and (253) if and only if (77) holds. □

**APPENDIX C**

**PROOF OF THEOREM 13**

**Proof:** Let $X$ be of the form in (73). By Theorem 12, $d_a(X)$ exists. Let $M = \max_{1 \leq j \leq m} |w_j|$ and $m_k = \lfloor r^{-k} \rfloor$ where $k \in \mathbb{N}$. Then $|X| \leq \frac{M}{1 - r}$. By Lemma 2,

$$d_a(X) = \lim_{k \to \infty} \frac{H_a(\lfloor m_k X \rfloor)}{k \log \frac{1}{r}}. \quad (254)$$

Since

$$|m_k X| - \sum_{i=0}^{k-1} r^{-i} W_{k-i} | \leq \frac{2rM}{1 - r}, \quad (255)$$

applying Lemma 8 to (255) and substituting into (254) yield

$$d_a(X) \leq \lim_{k \to \infty} \frac{H_a(\sum_{i=0}^{k-1} r^{-i} W_{k-i})}{k \log \frac{1}{r}} \leq \frac{H_a(P)}{\log \frac{1}{r}}, \quad (256)$$

where we have used

$$H_a\left(\sum_{i=0}^{k-1} r^{-i} W_{k-i}\right) \leq H_a(W^k) = kH_a(P). \quad (257)$$

If the open set condition is satisfied, (81) holds, see [47], [16, Ch. 17].

**APPENDIX D**

**PROOF OF LEMMA 6**

**Proof:** First consider the scalar case of $n = 1$. Recall that $I(X, \text{snr}) = I(X; Y)$, where $Y = X + \epsilon Z$, $Z \sim \mathcal{N}(0, 1)$ is independent of $X$, and $\epsilon = \frac{1}{\sqrt{\text{snr}}}$. Denote the integer closest to $Z$ by

$$[Z] \triangleq [Z + 1/2]. \quad (258)$$

Note that

$$I(X; Y | [Z]) = I(X; Y, [Z]) \quad (259)$$

$$= I(X; Y) + I(X; [Z] | Y). \quad (260)$$

Then

$$I(X; Y | [Z]) - H([Z]) \leq I(X, \text{snr}) \leq I(X; Y | [Z]) \quad (261)$$

Next we bound the conditional mutual information in (261). Denote the standard normal density $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. For $t \in \mathbb{Z}$, let

$$f_t(w) = \frac{1}{\frac{1}{2} \varphi(t + \frac{1}{2})} \mathbf{1}[|w| \leq 1/2] \quad (262)$$

$$p_t(y) = \frac{1}{e} \mathbb{E}\left[ f_t \left( \frac{y - X}{\epsilon} \right) \right] \quad (263)$$

denote the density of $W \triangleq Z - [Z]$ and $X + \epsilon W$ conditioned on the event $\{[Z] = t\}$, respectively. Note that the density $f_t$ is related to the uniform distribution through

$$e^{-t|t|} \mathbf{1}[|w| \leq 1/2] \leq f_t(w) \leq e^{t|t|} \mathbf{1}[|w| \leq 1/2] \quad (264)$$

and, in view of (263),

$$e^{-|t|} \leq \frac{\epsilon}{P_X(B(\gamma, \epsilon/2))} \leq e^{|t|}. \quad (265)$$

Then for any $t \in \mathbb{Z}$,

$$I(X; Y | [Z] = t) \quad (266)$$

$$= I(X; X + \epsilon W | [Z] = t) \quad (267)$$

$$\geq \mathbb{E}\left[ \log \frac{f_t(W)}{\epsilon p_t(X + \epsilon W)} \right] - 2|t| \log e, \quad (268)$$

$$\geq \mathbb{E}\left[ \log \frac{1}{P_X(B(X + \epsilon W, \epsilon/2))} \right] - 2|t| \log e, \quad (269)$$

where (268) follows from (264), (265), and (269) follows from $B(X + \epsilon W, \epsilon/2) \subset B(X, \epsilon)$ since $|W| \leq 1/2$.

For the upper bound, let $U$ be uniformly distributed on $[-1/2, 1/2]$ and $Q_y$ denote the distribution of $X + 3\epsilon U$, whose density is given by the convolution:

$$q_y(y) = \frac{P_X(B(y, 3\epsilon/2))}{3\epsilon}. \quad (270)$$
By the characterization of mutual information $I(A;B) = \min_{Q_B} D(P_{B|A} \| Q_B | P_A)$, we have

$$I(X; Y | [Z] = t)$$

$$= I(X; X + \epsilon W | [Z] = t)$$

$$\leq D(P_{X + \epsilon W | X} \| Q_Y | P_X, [Z] = t)$$

$$= \mathbb{E} \left[ \log \frac{1}{P_X(B(X + \epsilon W, 3\epsilon/2))} \right]$$

$$\leq \mathbb{E} \left[ \log \frac{1}{P_X(B(X, \epsilon))} \right] + |t| \log e + \log 3,$$  \hspace{1cm} (273)

where (273) is due to (264) and $B(X + \epsilon W, 3\epsilon/2) \supset B(X, \epsilon)$. Assembling (261), (269) and (273), we have

$$-2 \log e \mathbb{E}[|Z|] - H(|Z|) - \log 3 \leq I(X; \text{snr}) - \mathbb{E} \left[ \log \frac{1}{P_X(B(X, \epsilon))} \right]$$

$$\leq \log e \mathbb{E}[|Z|].$$ \hspace{1cm} (274) \hspace{1cm} (275) \hspace{1cm} (276)

In view of the facts that $|Z| \leq |Z| + 1/2$, $\mathbb{E}[|Z|] = \sqrt{\frac{2}{\pi}}$, and $H(|Z|) \leq 1 + \mathbb{E}[|Z|]$, (274) and (276) can be further bounded by

$$-2 \log e \left( \frac{1}{2} + \sqrt{\frac{2}{\pi}} \right) + \frac{3}{2} + \frac{\pi}{2} + \log 3$$

and

$$\log e \left( \frac{1}{2} + \sqrt{\frac{2}{\pi}} \right)$$

from below and above, respectively, leading to the desired (82) and (83). The proof for the vector case follows in analogous fashion. \hspace{1cm} \Box

**Remark 10:** Using the same arguments that lead to (269) and (273), we have

$$0 \leq I(X; X + \epsilon U) - \mathbb{E} \left[ \log \frac{1}{P_X(B(X, \epsilon))} \right] \leq \log 3.$$ \hspace{1cm} (277)

where $U$ is uniform over $[-1/2, 1/2]$. Similarly, inspecting the above proof we note that (277), and hence Lemma 6, continue to hold for well-behaved noise densities (with different absolute constants), such as exponential and Laplace distributions, which, when restricted on a finite interval, do not deviate too much from the uniform distribution.

**APPENDIX E**

**SUMSET ENTROPY INEQUALITIES**

In this appendix we develop new entropy inequalities for linear combinations of group-valued random variables with integer coefficients. We then prove the counterpart for information dimension announced earlier in Theorem 3. To this end, first we present several auxiliary results from sumset theory for Shannon entropy.

**Lemma 14 (Submodularity [45, Lemma A.2]):** If $X_0, X_1, X_2, X_{12}$ are discrete random variables such that $X_0 = f(X_1) = g(X_2)$ and $X_{12} = h(X_1, X_2)$, where $f, g, h$ are deterministic functions. Then

$$H(X_0) + H(X_{12}) \leq H(X_1) + H(X_2).$$ \hspace{1cm} (278)

In the remainder of this appendix, $G$ denotes an arbitrary abelian group.

**Lemma 15 (Ruzsa's Triangle Inequality [45, Th. 1.7]):** Let $X$ and $Y$ be $G$-valued random variables. Define the Ruzsa distance$^{12}$ as

$$\Delta(X, Y) \triangleq H(X' - Y') - \frac{1}{2}H(X') - \frac{1}{2}H(Y'),$$ \hspace{1cm} (279)

where $X'$ and $Y'$ are independent copies of $X$ and $Y$. Then

$$\Delta(X, Z) \leq \Delta(X, Y) + \Delta(Y, Z),$$ \hspace{1cm} (280)

that is, for independent $(X, Y, Z)$,

$$H(X - Z) \leq H(X - Y) + H(Y - Z) - H(Y).$$ \hspace{1cm} (281)

The following entropic analog of Plünnecke-Ruzsa’s inequality [48] is pointed out by Tao [49, p. 203], which is also implied by [50, Proposition 1.3].

**Lemma 16:** Let $X, Y_1, \ldots, Y_m$ be independent $G$-valued random variables. Then

$$H(X + Y_1 + \ldots + Y_m) \leq H(X) + \sum_{i=1}^m (H(X + Y_i) - H(X)).$$ \hspace{1cm} (282)

The following entropy inequality for group-valued random variables is based on results from [51]–[53], whose proof can be found in [54, Sec. II-E].

**Lemma 17:**

$$\frac{1}{2} \leq \frac{\Delta(X, X)}{\Delta(X, -X)} \leq 2.$$ \hspace{1cm} (283)

The foregoing entropy inequalities deal with sums of independent random variables with coefficients equal to $\pm 1$. In Theorem 14 we present a new upper bound on the entropy of the linear combination of independent random variables with general integer coefficients. The following lemma lies at the heart of the proof.

**Lemma 18:** Let $X, X'$ and $Z$ be independent $G$-valued random variables where $X'$ has the same distribution as $X$. Let $p, r \in \mathbb{Z} \setminus \{0\}$. Then

$$H(pX + Z) \leq H((p - r)X + rX' + Z) + \Delta(X, X).$$ \hspace{1cm} (284)

Furthermore, if $p$ is even, we have

$$H(pX + Z) \leq H\left(\frac{p}{2}X + Z\right) + H(2X - X') - H(X).$$ \hspace{1cm} (285)

**Proof:** Applying Lemma 14 with

$$X_1 = ((p - r)X + rX' + Z, X - X'),$$ \hspace{1cm} (286)

$$X_2 = (X, Z), X_0 = pX + Z$$ and $X_{12} = (X, Z, X')$ yields (284), while (285) follows from an application of (281) (with the independent triple $(X, Y, Z)$ replaced by $(Z, -\frac{p}{2}X', pX)$), yielding

$$H(pX + Z) \leq H\left(\frac{p}{2}X' + Z\right) + H\left(pX - \frac{p}{2}X'\right) - H\left(\frac{p}{2}X'\right) \leq \frac{1}{2}H(X + Z) + H(2X - X') - H(X).$$ \hspace{1cm} (287)

\hspace{1cm} (288)

$^{12}$ $\Delta(\cdot, \cdot)$ is not a metric since $\Delta(X, X) > 0$ unless $X$ is deterministic.
Theorem 14: Let $X$ and $Y$ be independent $G$-valued random variables. Let $p, q \in \mathbb{Z}\backslash\{0\}$. Then
\[
H(pX + qY) - H(X + Y) \leq \tau_{p,q}(2H(X + Y) - H(X) - H(Y)),
\]
where
\[
\tau_{p,q} = 7[\log |p|] + 7[\log |q|] + 2. \tag{290}
\]
It is interesting to compare Theorem 14 and its proof techniques to its additive-combinatorial counterpart [55, Th. 1.3], which gives an upper bound on the cardinality of the sum of dilated subsets.

Proof of Theorem 14: Let $\{X_k\}$ and $\{Y_k\}$ be sequences of independent copies of $X$ and $Y$ respectively. By (281), we have
\[
H(pX + qY) \leq H(pX - X_1) + H(X + qY) - H(X)
\]
\[
\leq H(pX - X_1) + H(qY - Y_1) + H(X + Y) - H(X) - H(Y). \tag{292}
\]
Next we upper bound $H(pX - X_1)$. First we assume that $p$ is positive. If $p$ is even, applying (285) yields
\[
H(pX - X_1) \leq H\left(\frac{p}{2}X - X_1\right) + H(2X - X_1) - H(X). \tag{293}
\]
If $p$ is odd, applying (284) with $r = 1$ then applying (285), we have
\[
H(pX - X_1) \leq H\left((p - 1)X + X_2 - X_1\right) + \Delta(X, X) \tag{294}
\]
\[
\leq H\left(\frac{p - 1}{2}X + X_2 - X_1\right) + \Delta(X, X)
\]
\[
+ H(2X - X_1) - H(X) \tag{295}
\]
\[
\leq H\left(\frac{p - 1}{2}X + X_2 - X_1\right) + 5\Delta(X, X), \tag{296}
\]
where (296) holds because
\[
H(2X - X_1) \leq H(X + X_2 - X_1) + \Delta(X, X) \tag{297}
\]
\[
\leq H(X) + \Delta(X, -X) + 2\Delta(X, X) \tag{298}
\]
\[
\leq H(X) + 4\Delta(X, X). \tag{299}
\]
where (297)–(299) are due to (284), Lemma 16 and Lemma 17, respectively.

Combining (293) and (296) gives the following upper bound
\[
H(pX - X_1) \leq H\left(\frac{p}{2}X + X_2 - X_1\right) + 5\Delta(X, X). \tag{300}
\]

Iterating the above procedure with $p$ replaced by $\left\lfloor \frac{p}{2} \right\rfloor$, we have
\[
H(pX - X_1) \leq H\left(X - X_1 + \sum_{i=2}^{[\log p] + 1} X_i\right) + 5[\log p]\Delta(X, X) \tag{301}
\]
\[
\leq H(X) + [\log p]\Delta(X, -X) + (5[\log p] + 1)\Delta(X, X) \tag{302}
\]
\[
\leq H(X) + (7[\log p] + 1)\Delta(X, X), \tag{303}
\]
where (302) are (303) are consequences of Lemma 16 and Lemma 17, respectively. Similarly, for negative $p$, the above proof remains valid if we apply (284) with $r = -1$ instead of $1$. Then (303) holds with $p$ replaced by $-p$. Assembling (303) and (292), we have
\[
H(pX + qY) - H(X + Y) \leq (7[\log |p|] + 1)\Delta(X, X) + (7[\log |q|] + 1)\Delta(Y, Y) \tag{304}
\]
\[
\leq (14[\log |p|] + 2)\Delta(X, -Y) + (14[\log |q|] + 2)\Delta(X, -Y), \tag{305}
\]
\[
= (7[\log |p|] + 7[\log |q|] + 2)(2H(X + Y) - H(X) - H(Y)), \tag{306}
\]
where (305) is due to the triangle inequality in Lemma 15: $\max(\Delta(X, X), \Delta(Y, Y)) \leq 2\Delta(X, -Y)$. The proof of (289) is now completed. □

Finally, we prove Theorem 3 by applying Theorem 14 (with the abelian group $G$ being $\mathbb{R}$ endowed with the conventional addition).

Proof: First we assume that $p' = q' = 1$. In view of Lemma 9, applying (289) to $[X_m]$ and $[Y_m]$ yields
\[
H([pX + qY]_m) - H([X + Y]_m)
\]
\[
= H([pX]_m + [qY]_m) - H([X]_m + [Y]_m) + O(1) \tag{307}
\]
\[
\leq \tau_{p,q}(2H([X]_m) + [Y]_m) - H([X]_m) - H([Y]_m) + O(1) \tag{308}
\]
\[
\leq \tau_{p,q}(2H([X]_m) + [Y]_m) - H([X]_m) - H([Y]_m)) + O(1), \tag{309}
\]
Dividing both sides by $m$ and sending $m \to \infty$ yields
\[
d(pX + qY) - d(X + Y) \leq \tau_{p,q}(2d(X) - d(X) - d(Y)). \tag{310}
\]

Then
\[
4d(pX + qY) - d(X + Y) \tag{311}
\]
\[
= [d(X) + d(Y) - 2d(X + Y)]
\]
\[
+ [d(pX + qY) - d(X) - d(X + Y)]
\]
\[
+ [d(pX + qY) - d(Y) - d(X + Y)] + 2d(pX + qY) \tag{312}
\]
\[
\leq - \frac{1}{\tau_{p,q}}[d(pX + qY) - d(X + Y)] + 2,
\]
where (312) is due to (310), (16) and (13). Therefore
\[
d(pX + qY) - d(X + Y) \leq \frac{2\tau_{p,q}}{4\tau_{p,q} + 1} = \frac{1}{2} - \frac{1}{8\tau_{p,q} + 2}, \tag{313}
\]
which is the desired (27).

Finally, in view of the scale-invariance of information dimension and (313), we have
\[
d(pX + qY) - d(p'X') + (p'X' + q'y') \tag{314}
\]
\[
d(pq'p'X' + p'qq'y') - d(p'X' + q'y') \tag{315}
\]
\[
\leq \frac{1}{2} - \epsilon(p', q'). \tag{315}
\]
Remark 11: For the special case of $p = 2$ and $q = 1$, we proceed as follows:

\[
H(2X + Y) \leq \Delta(X, X) + H(X + X' + Y) \leq \Delta(X, X) + 2H(X + Y) - H(Y) \leq 4H(X + Y) - 2H(Y) - H(X).
\]

where

- (316): by (284);
- (317): by Lemma 16;
- (318): by triangle inequality $\Delta(X,Y) \leq 2\Delta(X, -Y)$.

By (318), we have $d(2X + Y) - d(X + Y) \leq 3d(X + Y) - d(X) - 2d(Y)$. Consequently, starting with $6d(2X + Y) - d(X + Y)$ and using the same manipulations as in (311)–(313), we obtain (29). Analogously, the case of $p = 1$ and $q = -1$ follows from the inequality $\Delta(X, Y) \leq 3\Delta(X, -Y)$ [45, (2.2)], i.e., $H(X - Y) \leq 3H(X + Y) - H(X) - H(Y)$, which yields (30).

APPENDIX F

PROOF OF (61)

Proof: The converse follows from Theorem 5 by choosing $\pi$ as the cyclic shifts on $\{1, 2, 3\}$. To achieve \( \frac{3}{2} \) degrees of freedom, for each \( i \), we choose $\delta X_i = \mu_2$ defined in Section IV-C, with $d(X_i) = \frac{1}{2}$. Note that the odd bits in the binary expansion of $X_i$ are independent equiprobable Bernoulli distributed, while the even bits are zero. Therefore $X_1 + 2X_2$ are uniformly distributed on the unit interval and $d(X_1 + 2X_2) = 1$. Similarly, we have $d\delta X^3, H = \frac{3}{2}$. Note that (61) can also be achieved by letting each user transmit one half degrees of freedom, and decoding both messages at each receiver, since the rates fall within the MAC capacity region.

ACKNOWLEDGMENT

The paper has benefited from thorough suggestions by the anonymous reviewers. Stimulating discussions with Yann Bugeaud, Mike Hochman and Pablo Shmerkin are gratefully acknowledged. The paper also benefited from detailed comments from Helmut Bölcskei, David Stotz, and Chinmay Vaze.

REFERENCES


Yihong Wu (S’10–M’13) received the B.E. degree from Tsinghua University, Beijing, China, in 2006 and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ, in 2008 and 2011, all in electrical engineering. Before joining the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign as an assistant professor in 2013, he was a postdoctoral fellow with the Statistics Department, The Wharton School, University of Pennsylvania, Philadelphia, PA. His research interests are in information theory and high-dimensional statistics.

Dr. Wu was a recipient of the Princeton University Wallace Memorial Honors Fellowship in 2010. He was a recipient of 2011 Marconi Society Paul Baran Young Scholar Award and the Best Student Paper Award at the 2011 IEEE International Symposium on Information Theory (ISIT).

Shlomo Shamai (Shitz) (S’80–M’82–SM’88–F’94) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, in 1975, 1981 and 1986 respectively.

During 1975-1985 he was with the Communications Research Labs, in the capacity of a Senior Research Engineer. Since 1986 he is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, where he is now a Technion Distinguished Professor, and holds the William Fondiller Chair of Telecommunications. His research interests encompass a wide spectrum of topics in information theory and statistical communications.

Dr. Shamai (Shitz) is a member of the Israeli Academy of Sciences and Humanities and a foreign member of the US National Academy of Engineering. He is the recipient of the 2011 Claude E. Shannon Award and the 2014 Rothschild Prize in Mathematics/Computer Sciences and Engineering.

He has been awarded the 1999 van der Pol Gold Medal of the Union Radio Scientifique Internationale (URSI), and is a co-recipient of the 2000 IEEE Donald G. Fink Prize Paper Award, the 2003, and the 2004 Joint IT/COM societies paper award, the 2007 IEEE Information Theory Society Paper Award, the 2009 European Commission FP7, Network of Excellence in Wireless COMMunications (NEWCOM++) Best Paper Award, the 2010 Thomson Reuters Award for International Excellence in Scientific Research, and the 2014 EURASIP Best Paper Award (for the EURASIP Journal on Wireless Communications and Networking). He is also the recipient of 1985 Alon Grant for distinguished young scientists and the 2000 Technion Henry Taub Prize for Excellence in Research. He has served as Associate Editor for the Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has also served twice on the Board of Governors of the Information Theory Society. He has served on the Executive Editorial Board of the IEEE TRANSACTIONS ON INFORMATION THEORY.

Sergio Verdú (S’80–M’84–SM’88–F’93) received the Telecommunications Engineering degree from the Universitat Politècnica de Barcelona in 1980, and the Ph.D. degree in Electrical Engineering from the University of Illinois at Urbana-Champaign in 1984. Since 1984 he has been a member of the faculty of Princeton University, where he is the Eugene Higgins Professor of Electrical Engineering, and is a member of the Program in Applied and Computational Mathematics.

Sergio Verdú is the recipient of the 2007 Claude E. Shannon Award, and the 2008 IEEE Richard W. Hamming Medal. He is a member of the National Academy of Engineering and the National Academy of Sciences, and was awarded a Doctore Honoris Causa from the Universitat Politècnica de Catalunya in 2005.

Verdú is a recipient of several paper awards from the IEEE: the 1992 Donald Fink Paper Award, the 1998 and 2012 Information Theory Paper Awards, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Prize Award, the 2006 Joint Communications/Information Theory Paper Award, and the 2009 Stephen O. Rice Prize from IEEE Communications Society. In 1998, Cambridge University Press published his book Multisuser Detection, for which he received the 2000 Frederick E. Terman Award from the American Society for Engineering Education.


WU et al.: INFORMATION DIMENSION AND THE DoFs OF THE INTERFERENCE CHANNEL 279