Title
Degrees of Freedom Region of the (M, N1, N2) MIMO Broadcast Channel with Partial CSIT: An Application of Sum-set Inequalities

Authors
Davoodi, Arash Gholami
Jafar, Syed A

Publication Date
2019-07-12

DOI
10.1109/isit.2019.8849793

Degrees of Freedom Region of the (M, N_1, N_2) MIMO Broadcast Channel with Partial CSIT: An Application of Sum-set Inequalities Based on Aligned Image Sets

Arash Gholami Davoodi and Syed A. Jafar
Center for Pervasive Communications and Computing (CPCC)
University of California Irvine, Irvine, CA 92697
Email: {gholamid, syed}@uci.edu

Abstract

The degrees of freedom (DoF) region is characterized for the 2-user multiple input multiple output (MIMO) broadcast channel (BC), where the transmitter is equipped with M antennas, the two receivers are equipped with N_1 and N_2 antennas, and the levels of channel state information at the transmitter (CSIT) for the two users are parameterized by β_1, β_2, respectively. The achievability of the DoF region was established by Hao, Rassouli and Clerckx, but no proof of optimality was heretofore available. The proof of optimality is provided in this work with the aid of sum-set inequalities based on the aligned image sets (AIS) approach.

1 Introduction

The availability of channel state information at the transmitter(s) (CSIT) greatly affects the capacity of wireless networks, so much so that even the coarse degrees of freedom (DoF) metric is significantly impacted. Under perfect CSIT a K-user interference channel has $K/2$ DoF [1] and the corresponding K-user MISO BC has K DoF almost surely [2]. However, if CSIT is limited to finite precision then the DoF collapse to unity in both cases. The large gap between the two extremes underscores the importance of studying partial CSIT settings. A key obstacle for these studies tends to be the proof of optimality once an achievable DoF region has been established based on the best known achievable schemes. For instance, the conjecture by Lapidoth, Shamai and Wigger [3], that the DoF collapse under finite precision CSIT, remained open for nearly a decade, until it was finally settled using an unconventional (combinatorial) argument, called the aligned image sets (AIS) approach in [4]. The AIS approach seeks to directly bound the number of codewords that can be resolved at one receiver while aligning at another receiver, under arbitrary levels of CSIT. Since its introduction in [4], the AIS approach has been successfully applied to construct proofs of optimality for a number of basic broadcast and interference channel settings. With each application the AIS approach has been further generalized, broadening its utility and scope. The unconventional nature of the AIS approach, in particular its reliance on combinatorial reasoning from first principles to bound the sizes of the aligned image sets, makes these generalizations quite challenging. Particularly relevant to this work is the recent effort in [5] to derive a new class of sumset inequalities based on the AIS approach, to serve as a toolkit for future DoF studies. In this work we demonstrate the utility of these sumset inequalities by providing the proof of optimality

for a DoF region for the 2-user MIMO BC under partial CSIT, that was shown to be achievable by by Hao, Rassouli and Clerckx in [6], but whose optimality was heretofore open.

The setting of interest is a 2-user MIMO BC where the transmitter is equipped with M antennas, the two receivers are equipped with N_1 and N_2 antennas, and the levels of CSIT for the two users are parameterized by $\beta_1, \beta_2 \in [0, 1]$, respectively, such that $\beta_i = 0$ represents no CSIT, $\beta_i = 1$ represents perfect CSIT, and the intermediate values represent corresponding levels of partial CSIT. Existing results for this channel focus primarily on the two extremes of perfect CSIT and no CSIT. Exact capacity is known for the MIMO BC if the CSIT is perfect [7]. The collapse of DoF under no CSIT has been shown for this channel under restrictive assumptions such as isotropic fading that essentially appeal to the degraded BC perspective [2, 8, 9, 10]. The particular setting of the MISO BC, where each user is equipped with only one antenna, i.e., $N_1 = N_2 = 1$, has recently seen much progress based on the AIS approach in [4, 11, 12], leading ultimately to its full GDoF characterization in [12] with arbitrary channel strengths and arbitrary channel uncertainty levels. For arbitrary antenna configurations and arbitrary levels of partial CSIT, an achievable DoF region is established by Hao, Rassouli and Clerckx in [6]. The optimality of this achievable region has been shown in [6] for certain parameter regimes (mainly $N_1 \leq N_2, M \leq N_2$), based on existing bounds, as well as AIS arguments. However, the general DoF region characterization remains open. Our main goal in this work is to provide a complete DoF region characterization by providing the proof of optimality that was heretofore missing for the remaining parameter regime. Remarkably, the proof makes use of the sumset inequalities recently developed in [5].

2 Notation and Definitions

For $n \in \mathbb{N}$, define the notation $[n] = \{1, 2, \cdots, n\}$. The cardinality of a set A is denoted as $|A|$. The notation $X^{[n]}$ stands for $\{X(1), X(2), \cdots, X(n)\}$. Moreover, $X_i^{[n]}$ also stands for $\{X_i(t) : \forall t \in [n]\}$. The support of a random variable X is denoted as $\text{supp}(X)$. The sets $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{R}^n$ and $\mathbb{Q}^n$ stand for the sets of real numbers, rational numbers, all n-tuples of real numbers and all n-tuples of rational numbers, respectively. Moreover, the set $\mathbb{R}^{2+}$ is defined as the set of all pairs of non-negative numbers. For any set S, we define the set S^c as the complement of the set S. If A is a set of random variables, then $H(A)$ refers to the joint entropy of the random variables in A. Conditional entropies, mutual information and joint and conditional probability densities of sets of random variables are similarly interpreted. Moreover, we use the Landau $O(\cdot)$ and $o(\cdot)$ notations as follows. For functions $f(x), g(x)$ from $\mathbb{R}$ to $\mathbb{R}$, $f(x) = O(g(x))$ denotes that $\limsup_{x\to\infty} \frac{|f(x)|}{|g(x)|} < \infty$. $f(x) = o(g(x))$ denotes that $\limsup_{x\to\infty} \frac{|f(x)|}{|g(x)|} = 0$. We use the notation $A \doteq B$ to indicate that the difference $|A - B|$ is negligible in the DoF sense. We use $\mathbb{P}(\cdot)$ to denote the probability function $\text{Prob}(\cdot)$. For any real number x we define $\lfloor x \rfloor$ as the largest integer that is smaller than or equal to x when $x > 0$, the smallest integer that is larger than or equal to x when $x < 0$, and x itself when x is an integer. The number $X_{r,s}$ may be represented as X_{rs} if there is no cause of ambiguity. For any vector $V = \begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix}^T$ and non-negative integer numbers m and n less than k, let us define the notation $V_{m\to n}$ as follows,

$$V_{m\to n} \triangleq \begin{cases} \begin{bmatrix} v_{m+1} & \cdots & v_{m+n} \end{bmatrix}^T, & m+n \leq k \\ \begin{bmatrix} v_{m+1} & \cdots & v_k & v_1 & \cdots & v_{m+n-k} \end{bmatrix}^T, & k < m+n \end{cases} \tag{1}$$

For any two vectors $V = \begin{bmatrix} v_1 & \cdots & v_{k_1} \end{bmatrix}^T$ and $W = \begin{bmatrix} w_1 & \cdots & w_{k_2} \end{bmatrix}^T$ define their concatenation as

$$V \bigtriangledown W \triangleq \begin{bmatrix} v_1 & \cdots & v_{k_1} & w_1 & \cdots & w_{k_2} \end{bmatrix}^T$$

Finally, for any $m \times n$ matrix $V = \begin{bmatrix} v_{11} & \cdots & v_{1n} \\ \ddots & \vdots & \ddots \\ v_{m1} & \cdots & v_{mn} \end{bmatrix}$ and $a, b, c, d \in \mathbb{N}$ where $a + b \leq m$ and $c + d \leq n$, define

$$V_{(a \to b):(c \to d)} \triangleq \begin{bmatrix} v_{a+1,c+1} & \cdots & v_{a+1,c+d} \\ \ddots & \vdots & \ddots \\ v_{a+b,c+1} & \cdots & v_{a+b,c+d} \end{bmatrix} \tag{2}$$

The following definitions, inherited from [5], are replicated here for the sake of completeness.

Definition 1 (Power Levels) *Consider integer valued variables X_i over alphabet $\mathcal{X}_{\lambda_i}$,*

$$\mathcal{X}_{\lambda_i} \triangleq \{0, 1, 2, \cdots, \bar{P}^{\lambda_i} - 1\} \tag{3}$$

where $\bar{P}^{\lambda_i}$ is a compact notation for $\left\lfloor \sqrt{P^{\lambda_i}} \right\rfloor$. We refer to $P \in \mathbb{R}_+$ as power*, and are primarily interested in limits as $P \to \infty$. Quantities that do not depend on P will be referred to as constants. The constant $\lambda_i \in \mathbb{R}_+$ denotes the* power level *of X_i.*

Definition 2 *For non-negative real numbers X, λ_1 and λ_2, define $(X)_{\lambda_1}$ and $(X)_{\lambda_1}^{\lambda_2}$ as,*

$$(X)_{\lambda_1} \triangleq X - \bar{P}^{\lambda_1} \left\lfloor \frac{X}{\bar{P}^{\lambda_1}} \right\rfloor \tag{4}$$

$$(X)_{\lambda_1}^{\lambda_2} \triangleq \left\lfloor \frac{X - \bar{P}^{\lambda_2} \left\lfloor \frac{X}{\bar{P}^{\lambda_2}} \right\rfloor}{\bar{P}^{\lambda_1}} \right\rfloor \tag{5}$$

In words, for any $X \in \mathcal{X}_{\lambda_1+\lambda_2}$, $(X)_{\lambda_1}^{\lambda_1+\lambda_2}$ retrieves the top λ_2 power levels of X, while $(X)_{\lambda_1}$ retrieves the bottom λ_1 levels of X. $(X)_{\lambda_1}^{\lambda_3}$ retrieves only the part of X that lies between power levels λ_1 and λ_3. Note that $X \in \mathcal{X}_\lambda$ can be expressed as $X = \bar{P}^{\lambda_1}(X)_{\lambda_1}^{\lambda} + (X)_{\lambda_1}$ for $0 \leq \lambda_1 < \lambda$. Equivalently, suppose $X_1 \in \mathcal{X}_{\lambda_1}$, $X_2 \in \mathcal{X}_{\lambda_2}$, $0 < \lambda_2$ and $X = X_1 + X_2\bar{P}^{\lambda_1}$. Then $X_1 = (X)_{\lambda_1}$, $X_2 = (X)_{\lambda_1}^{\lambda_1+\lambda_2}$. A conceptual illustration of power level partitions is shown in Figure 1.

Since expressions of the form $(X)_{1-\lambda}^{1}$ appear frequently in this particular work, let us define a compact notation for this as follows.

$$(X)^{\lambda} \triangleq (X)_{1-\lambda}^{1}. \tag{6}$$

Definition 3 *For the vector $\mathbf{V} = \begin{bmatrix} v_1 & v_2 & \cdots & v_k \end{bmatrix}^T$, we define $(\mathbf{V})_{\lambda_1}$ and $(\mathbf{V})_{\lambda_1}^{\lambda_2}$ as,*

$$(\mathbf{V})_{\lambda_1} \triangleq \begin{bmatrix} (v_1)_{\lambda_1} & (v_2)_{\lambda_1} & \cdots & (v_k)_{\lambda_1} \end{bmatrix}^T \tag{7}$$

$$(\mathbf{V})_{\lambda_1}^{\lambda_2} \triangleq \begin{bmatrix} (v_1)_{\lambda_1}^{\lambda_2} & (v_2)_{\lambda_1}^{\lambda_2} & \cdots & (v_k)_{\lambda_1}^{\lambda_2} \end{bmatrix}^T \tag{8}$$

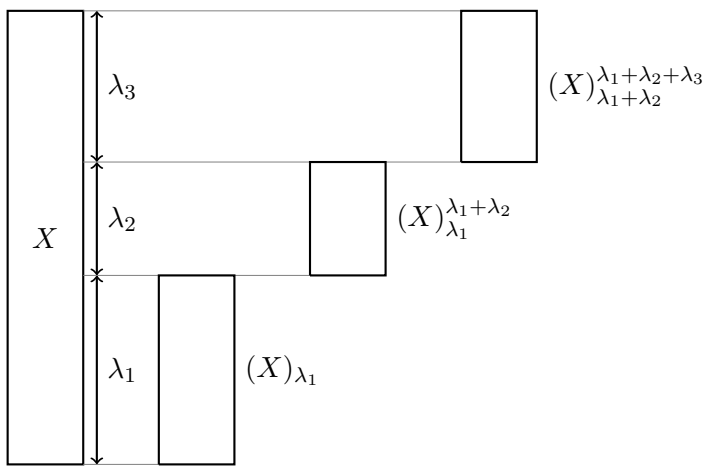

Figure 1: Conceptual depiction of an arbitrary variable $X \in \mathcal{X}_{\lambda_1+\lambda_2+\lambda_3}$, and its power-level partitions $(X)_{\lambda_1}$, $(X)_{\lambda_1}^{\lambda_1+\lambda_2}$ and $(X)_{\lambda_1+\lambda_2}^{\lambda_1+\lambda_2+\lambda_3}$.

Definition 4 (Bounded Density Channel Set $\mathcal{G}$) *Let $\mathcal{G}$ be a set of real-valued random variables, which satisfies both of the following conditions.*

1. *The magnitudes of all the random variables in $\mathcal{G}$ are bounded away from infinity, i.e., there exists a constant $\Delta < \infty$ such that for all $g \in \mathcal{G}$ we have $|g| \leq \Delta$.*
2. *There exists a finite positive constant $f_{\max}$, such that for all finite cardinality disjoint subsets $\mathcal{G}_1, \mathcal{G}_2$ of $\mathcal{G}$, the joint probability density function of all random variables in $\mathcal{G}_1$, conditioned on all random variables in $\mathcal{G}_2$, exists and is bounded above by $f_{\max}^{|\mathcal{G}_1|}$.*

Without loss of generality we will assume that $f_{\max} \geq 1, \Delta \geq 1$.

Definition 5 (Arbitrary Channel Set $\mathcal{H}$) *Let $\mathcal{H}$ be a set of real-valued constants with magnitudes bounded away from infinity, i.e., for all $h \in \mathcal{H}$ we have $|h| \leq \Delta$.*

Definition 6 *For real numbers $x_1 \in \mathcal{X}_{\eta_1}, x_2 \in \mathcal{X}_{\eta_2}, \cdots, x_k \in \mathcal{X}_{\eta_k}$ and the vectors $\vec{\gamma} = (\gamma_1, \gamma_2, \cdots, \gamma_k)$ and $\vec{\delta} = (\delta_1, \delta_2, \cdots, \delta_k)$ define the notations $L_j^g(x_i, 1 \leq i \leq k)$, $L_j^h(x_i, 1 \leq i \leq k)$, $L_j^{g\vec{\gamma}\vec{\delta}}(x_i, 1 \leq i \leq k)$ and $L_j^{h\vec{\gamma}\vec{\delta}}(x_i, 1 \leq i \leq k)$ to represent,*

$$L_j^g(x_1, x_2, \cdots, x_k) = \sum_{1 \leq i \leq k} \lfloor g_{j_i} x_i \rfloor \tag{9}$$

$$L_j^h(x_1, x_2, \cdots, x_k) = \sum_{1 \leq i \leq k} \lfloor h_{j_i} x_i \rfloor \tag{10}$$

$$L_j^{g\vec{\gamma}\vec{\delta}}(x_1, x_2, \cdots, x_k) = \sum_{1 \leq i \leq k} \lfloor g_{j_i} (x_i)_{\delta_i}^{\gamma_i} \rfloor \tag{11}$$

$$L_j^{h\vec{\gamma}\vec{\delta}}(x_1, x_2, \cdots, x_k) = \sum_{1 \leq i \leq k} \lfloor h_{j_i} (x_i)_{\delta_i}^{\gamma_i} \rfloor \tag{12}$$

for distinct random variables $g_{j_i} \in \mathcal{G}$, arbitrary real valued constants $h_{j_i} \in \mathcal{H}$, and arbitrary real valued constants $0 \leq \delta_i \leq \gamma_i \leq \eta_i$. For the vector $V = \begin{bmatrix} v_1 & v_2 & \cdots & v_k \end{bmatrix}^T$ we similarly define the notations $L_j^g(V)$ and $L_j^h(V)$ to represent,

$$L_j^g(V) = \sum_{1 \leq i \leq k} \lfloor g_{j_i} v_i \rfloor \tag{13}$$

$$L_j^h(V) = \sum_{1 \leq i \leq k} \lfloor h_{j_i} v_i \rfloor \tag{14}$$

Noting that these functions are *approximately* (because of the $\lfloor \cdot \rfloor$ operations) linear, for simplicity we refer to them as linear combinations. In particular, we will refer to L^g functions as *random* linear combinations and to L^h functions as *arbitrary* linear combinations. The variables x_i, v_i will generally be used to represent different parts of transmitted signals. Note that the subscripts, such as L_j, will be used to distinguish among different linear combinations, and may be dropped if there is no potential for ambiguity.

Definition 7 *For the linear combinations $A = L^{g\vec{\gamma}\vec{\delta}}(x_i, 1 \leq i \leq k)$ and $B = L^{h\vec{\gamma}\vec{\delta}}(x_i, 1 \leq i \leq k)$ where $x_1 \in \mathcal{X}_{\eta_1}, x_2 \in \mathcal{X}_{\eta_2}, \cdots, x_k \in \mathcal{X}_{\eta_k}$ we define $\mathcal{T}(A)$, $\mathcal{T}(B)$, $\mathcal{T}((A)_\mu^\lambda)$ and $\mathcal{T}((B)_\mu^\lambda)$ as,*

$$\mathcal{T}(A) = \mathcal{T}(B) = \max_{j \in [k]} (\gamma_j - \delta_j)^+ \tag{15}$$

$$\mathcal{T}((A)_\mu^\lambda) = \mathcal{T}((B)_\mu^\lambda) = (\min(\lambda, \max_{j \in [k]} (\gamma_j - \delta_j)^+) - \mu)^+ \tag{16}$$

$$\tag{17}$$

Note that the terminology from Definition (6) is invoked in Definition (7). Figure 2 provides a visual illustration of $L^{h\vec{\gamma}\vec{\delta}}$ and $\mathcal{T}(A)$. From the definition of $\mathcal{T}(A)$ and $\mathcal{T}(B)$ in (16), it follows that,

$$A \in \{a : a \in \mathbb{Z}, |a| \leq k\Delta \bar{P}^{\mathcal{T}(A)}\} \tag{18}$$

$$B \in \{a : a \in \mathbb{Z}, |a| \leq k\Delta \bar{P}^{\mathcal{T}(B)}\} \tag{19}$$

$$(A)_\mu^\lambda \in \{a : a \in \mathbb{Z}, |a| \leq \bar{P}^{\mathcal{T}((A)_\mu^\lambda)}\} \tag{20}$$

$$(B)_\mu^\lambda \in \{a : a \in \mathbb{Z}, |a| \leq \bar{P}^{\mathcal{T}((B)_\mu^\lambda)}\} \tag{21}$$

This is because the magnitudes of all[1] elements of $\mathcal{G}, \mathcal{H}$ are bounded from above by Δ.

3 Sum-set Inequalities: Previous Results in [5]

Let us recall Theorem 4 of [5].

[1] Consider the terms $A = L^{g\vec{\gamma}\vec{\delta}}(x_i, 1 \leq i \leq k)$ and $(B)_\mu^\lambda = (L^{h\vec{\gamma}\vec{\delta}}(x_i, 1 \leq i \leq k))_\mu^\lambda$ and let us bound them as follows.

$$|A| = |\sum_{1 \leq i \leq k} \lfloor g_i (x_i)_{\delta_i}^{\gamma_i} \rfloor| \leq k\Delta \max_{1 \leq i \leq k} (x_i)_{\delta_i}^{\gamma_i} \leq k\Delta \bar{P}^{\max_{1 \leq i \leq k} (\gamma_i - \delta_i)^+} \leq k\Delta \bar{P}^{\mathcal{T}(A)}. \tag{22}$$

$$|(B)_\mu^\lambda| = \left| \left\lfloor \frac{B - \bar{P}^\lambda \left\lfloor \frac{B}{\bar{P}^\lambda} \right\rfloor}{\bar{P}^\mu} \right\rfloor \right| \leq \left| \frac{B - \bar{P}^\lambda \left\lfloor \frac{B}{\bar{P}^\lambda} \right\rfloor}{\bar{P}^\mu} \right| \leq |\frac{\bar{P}^{\min(\lambda, \max_{j \in [k]} (\gamma_j - \delta_j)^+)}}{\bar{P}^\mu}| \leq \bar{P}^{\mathcal{T}((B)_\mu^\lambda)} \tag{23}$$

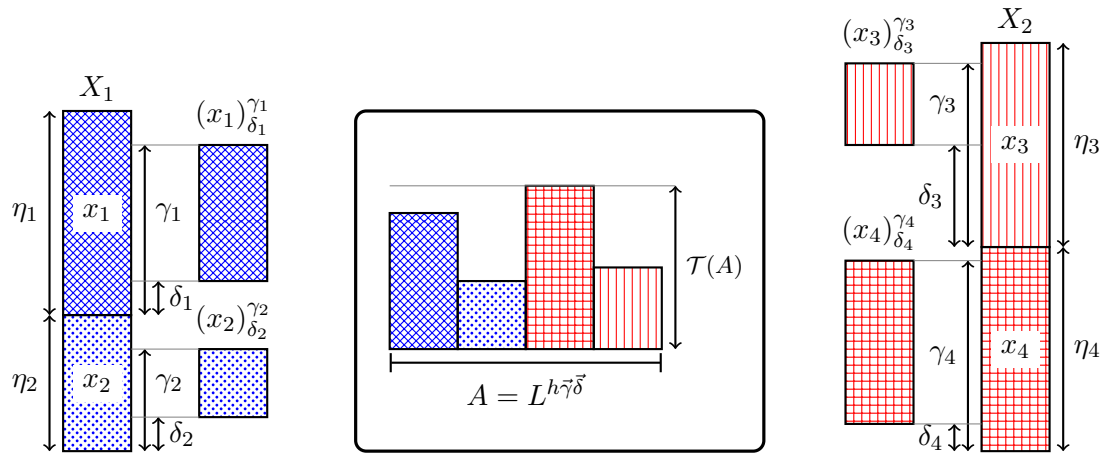

Figure 2: Visual illustration of $L^{\vec{\gamma}\vec{\delta}}$ and $\mathcal{T}(A)$. In this example, $x_1 \in \mathcal{X}_{\eta_1}$ and $x_2 \in \mathcal{X}_{\eta_2}$ are obtained as partitions of $X_1 \in \mathcal{X}_{\eta_1+\eta_2}$. Similarly, $x_3 \in \mathcal{X}_{\eta_3}$ and $x_4 \in \mathcal{X}_{\eta_4}$ are obtained as partitions of $X_2 \in \mathcal{X}_{\eta_3+\eta_4}$. Note that (γ_i, δ_i) are only used to further trim the size of x_i, yielding $(x_i)_{\delta_i}^{\gamma_i}$ as the trimmed versions. These trimmed variables are then combined with arbitrary coefficients to produce $A = L^{\vec{\gamma}\vec{\delta}}$. Finally, note that $\mathcal{T}(A)$ represents the size (power level) of the largest trimmed variable involved in $L^{\vec{\gamma}\vec{\delta}}$.

Theorem 1 (Theorem 4 in [5]) *Consider KM non-negative numbers $\{\lambda_{km} : k \in [K], m \in [M]\}$ and random variables $X_j(t) \in \mathcal{X}_{\max_{k\in[K]}\{\lambda_{k,1}+\lambda_{k,2}+\cdots+\lambda_{k,M}\}}$, $j \in [N]$, $t \in \mathbb{N}$, independent of $\mathcal{G}$, and $\forall k \in [K], K \leq N$, define*

$$
\begin{aligned}
Z_k(t) &= L_k^g(t)(X_1(t), X_2(t), \cdots, X_N(t)) && (24)\\
Z_{k,1}(t) &= L_{k1}^{h\vec{\gamma}_{k1}\vec{\delta}_{k1}}(t)((X_j(t))_{\sum_{r=1}^{i-1}\lambda_{kr}}^{\sum_{r=1}^{i}\lambda_{kr}}, i \in I_{k,1}, j \in [N]) && (25)\\
Z_{k,2}(t) &= L_{k2}^{h\vec{\gamma}_{k2}\vec{\delta}_{k2}}(t)((X_j(t))_{\sum_{r=1}^{i-1}\lambda_{kr}}^{\sum_{r=1}^{i}\lambda_{kr}}, i \in I_{k,2}, j \in [N]) && (26)\\
&\vdots \\
Z_{k,l_k}(t) &= L_{kl_k}^{h\vec{\gamma}_{kl_k}\vec{\delta}_{kl_k}}(t)((X_j(t))_{\sum_{r=1}^{i-1}\lambda_{kr}}^{\sum_{r=1}^{i}\lambda_{kr}}, i \in I_{k,l_k}, j \in [N]) && (27)
\end{aligned}
$$

The channel uses are indexed by $t \in \mathbb{N}$. $I_{kk'} \subset [M], k \in [K], k' \in [l_k]$, such that $i < j \Rightarrow m(k,i) \geq m(k,j)$, where

$$m(a,b) \triangleq \min\{m : m \in I_{a,b}\}.$$

If for all $k \in [K]$ and for each $s \in \{1, 2, \cdots, l_k - 1\}$,

$$\mathcal{T}(Z_{k,s+1}) + \mathcal{T}(Z_{k,s+2}) + \cdots + \mathcal{T}(Z_{k,l_k}) \quad \leq \quad \lambda_{k,1} + \lambda_{k,2} + \cdots + \lambda_{k,(m(k,s)-1)} \qquad (28)$$

then for any acceptable random variable W [2]

$$H(Z_1^{[n]}, \cdots, Z_K^{[n]} \mid W, \mathcal{G}) \quad \geq \quad H(Z_{1,1}^{[n]}, \cdots, Z_{K,l_K}^{[n]} \mid W) + Kn\ o(\log \bar{P}) \qquad (29)$$

[2] Let $\mathcal{G}(Z) \subset \mathcal{G}$ denote the set of all bounded density channel coefficients that appear in $Z_1^{[n]}, \cdots, Z_K^{[n]}$, and let W be a random variable such that conditioned on any $\mathcal{G}_o \subset (\mathcal{G}/\mathcal{G}(Z)) \cup \{W\}$, the channel coefficients $\mathcal{G}(Z)$ satisfy the bounded density assumption.

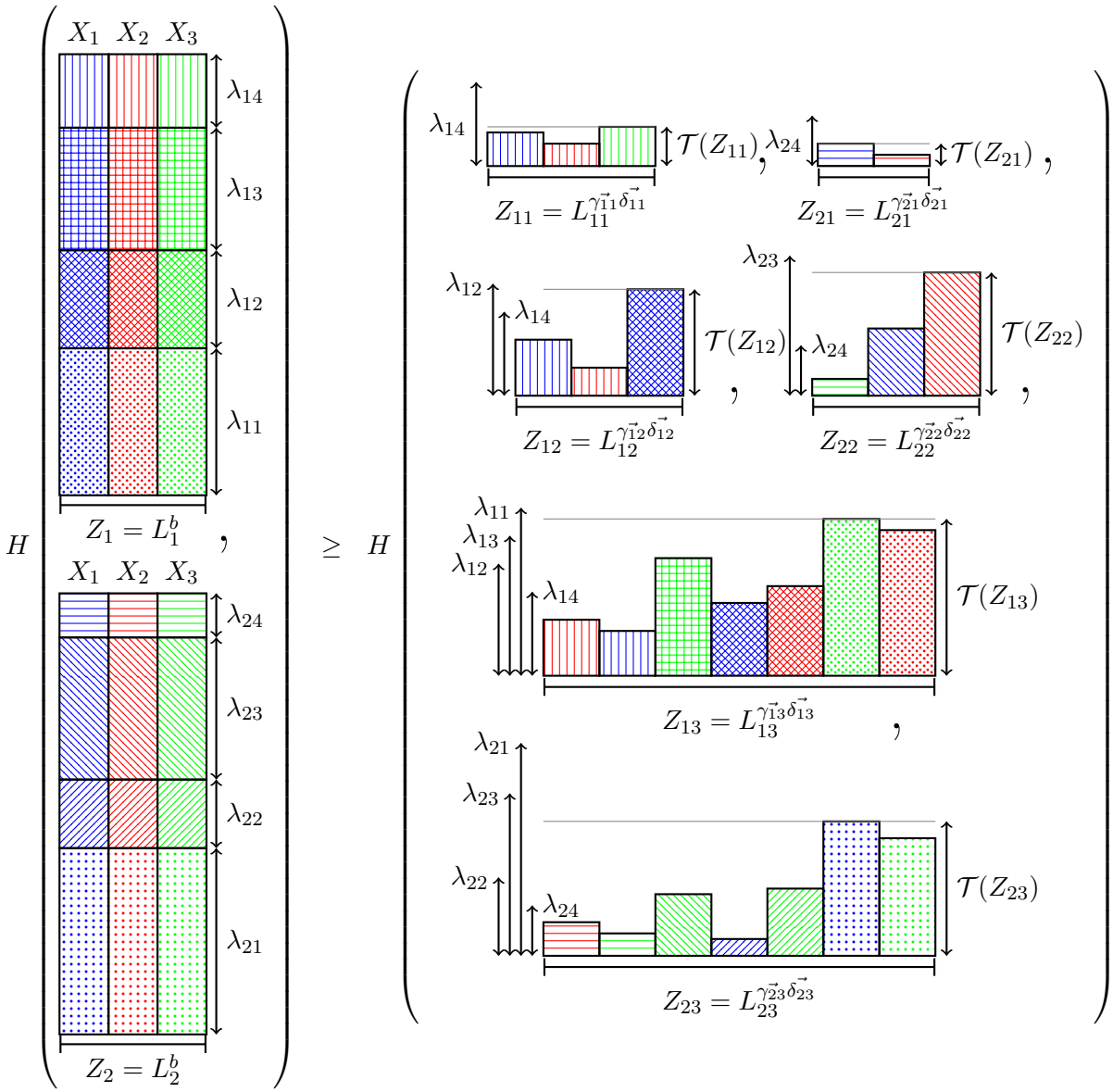

Figure 3: Illustration of an application of Theorem 1 of [5]. Note that in this figure we dropped the time index (t) for convenience. On the left is the joint entropy of the sum (bounded density linear combination) of $N = 3$ dependent random variables, $X_1(t), X_2(t), X_3(t) \in \mathcal{X}_{\max_{k\in[2]}\{\lambda_{k1}+\lambda_{k2}+\lambda_{k3}+\lambda_{k4}\}}$, $(M = 4)$, which is bounded below by the joint entropy of $l_1 + l_2 = 6$ arbitrary linear combinations, $Z_{11}, Z_{12}, Z_{13}, Z_{21}, Z_{22}, Z_{23}$, comprised of power level partitions of the two random variables. In this example, $I_{11} = I_{21} = \{4\}, I_{12} = \{2, 4\}, I_{22} = \{3, 4\}, I_{13} = I_{23} = \{1, 2, 3, 4\}$. Condition (28) is verified as $\lambda_{11} + \lambda_{12} + \lambda_{13} \geq \mathcal{T}(Z_{12}) + \mathcal{T}(Z_{13})$, $\lambda_{21} + \lambda_{22} + \lambda_{23} \geq \mathcal{T}(Z_{22}) + \mathcal{T}(Z_{23})$, $\lambda_{21} \geq \mathcal{T}(Z_{23})$ and $\lambda_{11} \geq \mathcal{T}(Z_{13})$.

4 System Model

In this work we will focus on the setting where all variables take only real values. Extensions to complex valued settings may be cumbersome but are expected to be conceptually straightforward as shown in [4]. We will focus on the two-user MIMO BC equipped with M antennas at the transmitter and N_1, N_2 antennas at the two receivers, with the assumption throughout that

$$N_1 \leq N_2 \leq M \leq N_1 + N_2, \tag{30}$$

since this is the only non-trivial setting where the DoF remain open. For all cases where the condition $N_1 \leq N_2 \leq M$ is not true, the DoF are already established in [6]. For all cases where $M > N_1 + N_2$, it is easy to see that the DoF region is not affected if the number of transmit antennas is reduced to $N_1 + N_2$, as follows. First, from the achievability side, note that the DoF innerbound shown in [6] remains unaffected if the number of transmit antennas is reduced to $N_1 + N_2$. Then from the outer bound perspective we note that the capacity cannot be reduced if a genie informs the transmitter of the $M - N_1 - N_2$ dimensional transmit signal space that is not heard by either user, which allows the transmitter to discard these $M - N_1 - N_2$ transmit dimensions (antennas) without loss of generality, thereby reducing the effective number of transmit dimensions to $N_1 + N_2$. Therefore in order to establish the DoF region for all cases where $M > N_1 + N_2$, it suffices to show that the DoF outer bound matches the DoF inner bound for the setting $M = N_1 + N_2$.

4.1 The Channel

If perfect CSIT was available, then for generic channel realizations in the two user (M, N_1, N_2) MIMO broadcast channel with $N_1 \leq N_2 \leq M \leq N_1 + N_2$, there are $M - N_1$ transmit directions available to the transmitter that are in the null-space of the channel matrix between the transmitter and the first receiver. Similarly, there are $M - N_2$ transmit directions available to the transmitter that are in the null-space of the channel matrix between the transmitter and the second receiver. A canonical representation of the channel (obtained by applying a change of basis operation at the transmitter) makes these directions explicit by mapping them to transmit antennas, so an M dimensional input vector $\mathbf{X}$ is partitioned as follows.

$$\mathbf{X}_a(t) = [\mathbf{X}(t)]_{0 \to (M-N_2)} \tag{31}$$

$$\mathbf{X}_b(t) = [\mathbf{X}(t)]_{(M-N_2) \to (N_1+N_2-M)} \tag{32}$$

$$\mathbf{X}_c(t) = [\mathbf{X}(t)]_{N_1 \to (M-N_1)} \tag{33}$$

Recall that the notation $\mathbf{X}_{m \to n}$ as used here stands for $(\mathbf{X}_{m+1}, \mathbf{X}_{m+2}, \cdots, \mathbf{X}_{m+n})$. Thus, the partition $\mathbf{X}_a$ contains transmit directions that are in the null-space of User 2 but not User 1, the partition $\mathbf{X}_c$ contains transmit directions that are in the null space of User 1 but not User 2, and $\mathbf{X}_b$ contains transmit directions that are not in the null space of either user. Further, note that if $M = N_1 + N_2$, then $\mathbf{X}_b$ disappears, and the partition is simply $\mathbf{X} = \mathbf{X}_a \bigtriangledown \mathbf{X}_c$. If $M < N_1 + N_2$ then the partition is $\mathbf{X} = \mathbf{X}_a \bigtriangledown \mathbf{X}_b \bigtriangledown \mathbf{X}_c$. Evidently, for the first user, zero forcing is possible only along the $M - N_1$ dimensional space corresponding to $\mathbf{X}_c$, and for the second user, zero forcing is possible only along the $M - N_2$ dimensional space corresponding to $\mathbf{X}_a$.

With partial CSIT, only partial zero-forcing is possible based on channel estimates available to the transmitter. Therefore, the channel model for the two user (M, N_1, N_2) MIMO BC with partial CSIT, is represented in its canonical form by the following input output equations. See Fig 4.

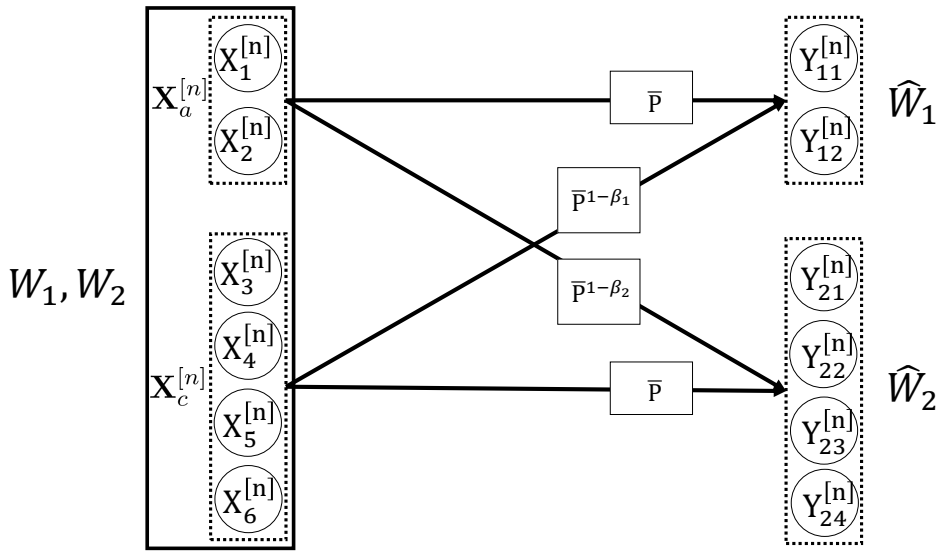

Figure 4: Equivalent channel for two user $(M, N_1, N_2) = (6, 2, 4)$ MIMO BC.

$$\mathbf{Y}_1(t) = \sqrt{P}\mathbf{G}_{1ab}(t)\left[\mathbf{X}_a(t) \bigtriangledown \mathbf{X}_b(t)\right] + \sqrt{P^{1-\beta_1}}\mathbf{G}_{1c}(t)\mathbf{X}_c(t) + \mathbf{\Gamma}_1(t) \tag{34}$$

$$\mathbf{Y}_2(t) = \sqrt{P}\mathbf{G}_{2bc}(t)\left[\mathbf{X}_b(t) \bigtriangledown \mathbf{X}_c(t)\right] + \sqrt{P^{1-\beta_2}}\mathbf{G}_{2a}(t)\mathbf{X}_a(t) + \mathbf{\Gamma}_2(t) \tag{35}$$

The dimensions of these symbols are listed as follows.

$$\mathbf{Y}_1(t), \mathbf{\Gamma}_1(t) : N_1 \times 1, \qquad \mathbf{Y}_2(t), \mathbf{\Gamma}_2(t) : N_2 \times 1, \tag{36}$$

$$\mathbf{G}_{1ab}(t) : N_1 \times N_1, \qquad \mathbf{G}_{1c}(t) : N_1 \times (M - N_1), \tag{37}$$

$$\mathbf{G}_{2bc}(t) : N_2 \times N_2, \qquad \mathbf{G}_{2a}(t) : N_2 \times (M - N_2). \tag{38}$$

Here, over channel use $t \in \mathbb{N}$, the vector of symbols seen at Receiver i, $i \in \{1, 2\}$, is $\mathbf{Y}_i(t)$, and the vector of symbols sent from the transmitter is $\mathbf{X}(t) = \mathbf{X}_a(t) \bigtriangledown \mathbf{X}_b(t) \bigtriangledown \mathbf{X}_c(t)$. Channel matrices $\mathbf{G}_{1ab}(t), \mathbf{G}_{2bc}(t)$ correspond to directions along which no zero-forcing is possible, while $\mathbf{G}_{1c}, \mathbf{G}_{2a}$ correspond to directions that can be partially zero-forced based on channel estimates available to the transmitter. Note that due to partial CSIT, the dimensions that can be partially zero-forced have channel strength diminished by the negative power exponents β_1, β_2 for users 1 and 2 respectively, relative to those directions along which no zero-forcing is possible. The quality of CSIT is captured by $\beta_1, \beta_2 \in [0, 1]$. As the CSIT parameters $\beta_j, j \in \{1, 2\}$, take values in the interval from 0 to 1 they cover the full range from no CSIT (i.e., no zero-forcing ability) to perfect CSIT (perfect zero-forcing ability) in the DoF sense. $\mathbf{\Gamma}_i(t)$ are the zero-mean unit variance additive white Gaussian noise terms seen at outputs $\mathbf{Y}_i(t)$, independent of all inputs and channel realizations. The input vector $\mathbf{X}(t)$ is subject to unit power constraint.

All channel coefficients are distinct random variables drawn from the bounded density channel set $\mathcal{G}$, (see Definition 4), therefore all channel coefficient magnitudes are bounded above by $\Delta < \infty$. Further, in order to avoid degenerate conditions, we assume that all channel matrices have determinants bounded away from zero, i.e., the absolute value of the determinant of each square channel submatrix is greater than a positive constant ϵ.

Perfect channel state information at the receivers (CSIR) is assumed to be available for all

channels. In terms of CSIT, we assume that the transmitter is aware of the bounded density probability density functions of all channels, but not the actual channel realizations.

4.2 DoF

The definitions of achievable rates $R_i(P)$ and capacity region $\mathcal{C}(P)$ are standard. The DoF region is defined as

$$\mathcal{D} = \{(d_1, d_2) : \exists (R_1(P), R_2(P)) \in \mathcal{C}(P), \text{ s.t. } d_k = \lim_{P\to\infty} \frac{R_k(P)}{\frac{1}{2}\log(P)}, \forall k \in \{1,2\}\} \tag{39}$$

5 Main Result

The following theorem characterizes the complete DoF region of the two-user MIMO BC with arbitrary levels of partial CSIT $\beta_1, \beta_2 \in [0,1]$, for arbitrary (unconstrained) choice of parameters M, N_1, N_2.

Theorem 2 *Without loss of generality, assume $N_1 \leq N_2$. The DoF region is expressed as follows.*

1. *For $N_2 \leq M$:*

$$\mathcal{D} = \Big\{(d_1, d_2) \in \mathbb{R}^{2+} \text{ such that}$$

$$d_1 \leq N_1, \tag{40}$$

$$d_2 \leq N_2, \tag{41}$$

$$\frac{d_1}{N_1} + \frac{d_2}{N_2} \leq 1 + \frac{M - N_1}{N_2}\beta_1, \tag{42}$$

$$d_1 + d_2 \leq N_2 + (M - N_2)\beta_2, \tag{43}$$

$$d_1 + d_2 \leq N_2 + (M - N_2)\beta_o\Big\} \tag{44}$$

where β_o is defined as,

$$\beta_o = \begin{cases} \frac{\beta_1\beta_2(M-N_2)}{(N_2-N_1)(1-\beta_1)+(M-N_2)\beta_2}, & \beta_1 + \beta_2 < 1 \\ \frac{N_1-N_2+(N_2-N_1)\beta_2+(M-N_1)\beta_1}{M-N_1}, & \beta_1 + \beta_2 \geq 1 \end{cases} \tag{45}$$

2. *For $N_2 > M$:*

$$\mathcal{D} = \Big\{(d_1, d_2) \in \mathbb{R}^{2+} \text{ such that}$$

$$d_1 \leq N_1 \tag{46}$$

$$d_1 + d_2 \leq M \tag{47}$$

$$\frac{d_1}{N_1} + \frac{d_2}{M} \leq 1 + \frac{M - N_1}{M}\beta_1\Big\} \tag{48}$$

Remark 1 *Wang and Varanasi [13] studied the DoF of the two-user MIMO broadcast channel with general message set (a common message and two private messages) under hybrid CSIT models where for each user the CSIT is either perfect (P), delayed (D), or not available (N). While the 'PP', 'PD', 'DP', 'DD', 'NN' settings are fully settled, for the 'PN', 'NP', 'DN' and 'ND' settings, only the linear DoF regions (i.e., the DoF region restricted to linear achievable schemes) are found and it is conjectured that the same regions are optimal even without restriction to linear achievable schemes. It is further explained in [13] that the 'NP' setting (where no CSIT is available for the first user and perfect CSIT is available for the second user)*[3] *is the key, i.e., if the 'NP' case can be solved then the other three cases can be easily resolved. Furthermore, tight outer bounds that include a common message are found directly from the setting with only private messages by reducing the decoding requirement for the common message to only one of the receivers. Since the 'NP' setting corresponds to $\beta_1 = 0, \beta_2 = 1$, evidently Theorem 2 settles the conjectures of Wang and Varanasi [13] in the affirmative.*

Remark 2 *Achievability of the DoF region of Theorem 2 is established by Hao et al. in [6] based on a rate-splitting scheme that includes interesting 'space-time' scheduling aspects. Partial converse results are also presented in [6] based on relatively straightforward applications of the aligned image sets (AIS) argument [4]. The problem that remains open is the proof of the outer bound* (44) *for $N_1 \leq N_2 \leq M \leq N_1 + N_2$, which is the main contribution of this work. Our proof exemplifies the utility of the 'sum-set inequalities' that were recently developed from AIS arguments in [5].*

The proof of Theorem 2 (i.e., the proof of (44) for $N_1 \leq N_2 \leq M \leq N_1 + N_2$) appears in Section 8 and is partitioned into two cases, corresponding to $\beta_1 + \beta_2 \geq 1$ and $\beta_1 + \beta_2 < 1$, that are covered in Section 8.3.1 and Section 8.3.2, respectively. For ease of exposition let us first illustrate the main ideas of the proof with two examples — the $(M, N_1, N_2) = (5, 2, 3)$ MIMO BC with $(\beta_1, \beta_2) = (1/2, 2/3)$ as representative of the case $\beta_1 + \beta_2 \geq 1$, and the $(M, N_1, N_2) = (4, 1, 3)$ MIMO BC with $(\beta_1, \beta_2) = (1/4, 1/2)$ for the case $\beta_1 + \beta_2 < 1$.

6 Example 1. $(M, N_1, N_2) = (5, 2, 3)$ with $(\beta_1, \beta_2) = (\frac{1}{2}, \frac{2}{3})$

For the two-user $(5, 2, 3)$ MIMO BC with $(\frac{1}{2}, \frac{2}{3})$ levels of partial CSIT, from Theorem 2 the DoF region is computed as,

$$\mathcal{D}_1 = \left\{ (d_1, d_2) \in \mathbb{R}^{2+} : \quad d_1 \leq 2, \quad d_2 \leq 3, \quad \frac{d_1}{2} + \frac{d_2}{3} \leq \frac{3}{2}, \quad d_1 + d_2 \leq 3 + \frac{7}{9} \right\} \tag{49}$$

and $\beta_o = \frac{7}{18}$ from (45). The challenge is to prove the bound (44), i.e., $d_1 + d_2 \leq 3 + 7/9$.

6.1 Deterministic Model

The first step of the AIS approach is to transform the channel model into the deterministic setting, such that a DoF outer bound for the deterministic setting is also a DoF outer bound for the original channel. This deterministic transformation produces a BC with input $\bar{\mathbf{X}}(t) = \bar{\mathbf{X}}_a(t) \bigtriangledown \bar{\mathbf{X}}_c(t)$, and outputs $\bar{\mathbf{Y}}_1(t), \bar{\mathbf{Y}}_2(t)$.

$$\bar{\mathbf{Y}}_1(t) \quad = \quad \left[\bar{Y}_{11}(t) \quad \bar{Y}_{12}(t) \right]^T \tag{50}$$

[3]Note that without loss of generality we assume that $N_1 \leq N_2$ whereas [13] assumes that $N_2 \leq N_1$. Thus our user indices are switched relative to [13]. As a consequence, what is referred to as the 'PN' setting in [13] corresponds to the 'NP' setting in this paper.

$$\bar{Y}_{1r}(t) = L^g_{1r(t)}\left(\bar{\mathbf{X}}_a(t) \bigtriangledown (\bar{\mathbf{X}}_c(t))^{1/2}\right), \forall r \in [2] \quad (51)$$

$$\bar{\mathbf{Y}}_2(t) = \begin{bmatrix}\bar{Y}_{21}(t) & \bar{Y}_{22}(t) & \bar{Y}_{23}(t)\end{bmatrix}^T \quad (52)$$

$$\bar{Y}_{2r}(t) = L^g_{2r(t)}\left((\bar{\mathbf{X}}_a(t))^{1/3} \bigtriangledown \bar{\mathbf{X}}_c(t)\right), \forall r \in [3] \quad (53)$$

where $\bar{\mathbf{X}}_a(t)$ and $\bar{\mathbf{X}}_c(t)$ are defined as,

$$\bar{\mathbf{X}}_a(t) = \begin{bmatrix}\bar{X}_1(t) & \bar{X}_2(t)\end{bmatrix}^T \quad (54)$$

$$\bar{\mathbf{X}}_c(t) = \begin{bmatrix}\bar{X}_3(t) & \bar{X}_4(t) & \bar{X}_5(t)\end{bmatrix}^T \quad (55)$$

and $\bar{X}_m(t) \in \{0, 1, \cdots, \bar{P}\}$, $\forall m \in [5]$.

6.2 A Key Lemma

The key to the proof of the bound, $d_1 + d_2 \leq 3 + \frac{7}{9}$, is the following lemma, which makes use of sumset inequalities from [5].

Lemma 1 *For the two-user MIMO BC with* $(M, N_1, N_2) = (5, 2, 3)$ *and* $(\beta_1, \beta_2) = (\frac{1}{2}, \frac{2}{3})$,

$$2H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) \leq 3H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) - H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) + 3n \log \bar{P} + n\ o\ (\log \bar{P}) \quad (56)$$

See Figure 5 for an accompanying illustration for Lemma 1. The proof of Lemma 1 is presented in

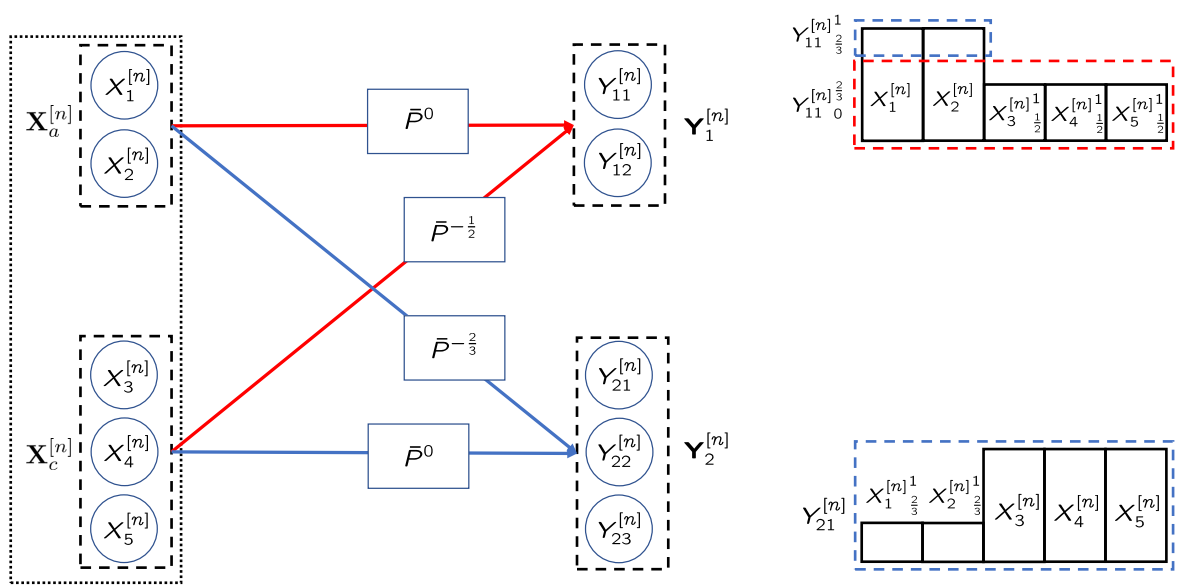

Figure 5: Illustration corresponding to Lemma 1.

Appendix A .

6.3 Proof of the Bound $d_1 + d_2 \leq 3 + \frac{7}{9}$

1. Starting from Fano's Inequality for the first receiver and suppressing $no(\log(P))$ terms that are inconsequential for DoF, we have,

$$
\begin{aligned}
3nR_1 &\leq 3I(\bar{\mathbf{Y}}_1^{[n]}; W_1 \mid \mathcal{G}) \\
&= 2H(\bar{\mathbf{Y}}_1^{[n]} \mid \mathcal{G}) - 3H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) + H(\bar{\mathbf{Y}}_1^{[n]} \mid \mathcal{G}) && (57)\\
&= 2H(\bar{\mathbf{Y}}_1^{[n]} \mid \mathcal{G}) - 3H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})_{2/3}, (\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid \mathcal{G}) && (58)\\
&= 2H(\bar{\mathbf{Y}}_1^{[n]} \mid \mathcal{G}) - 3H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})_{2/3} \mid (\bar{\mathbf{Y}}_1^{[n]})^{1/3}, \mathcal{G}) && (59)\\
&\leq \left(4 + \frac{4}{3}\right) n \log \bar{P} - 3H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid \mathcal{G}) && (60)\\
&\leq \left(7 + \frac{4}{3}\right) n \log \bar{P} - 2H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid \mathcal{G}) - H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) && (61)\\
&\leq \left(7 + \frac{4}{3}\right) n \log \bar{P} - 2H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_2, \mathcal{G}) && (62)
\end{aligned}
$$

where (58) follows from Definition 2 and (59) from the chain rule. (60) is implied by the fact that the entropy of a random variable is bounded by logarithm of the cardinality of its support, i.e., $2H(\bar{\mathbf{Y}}_1^{[n]} \mid \mathcal{G}) \leq 4n \log \bar{P}$ and $H((\bar{\mathbf{Y}}_1^{[n]})_{2/3} \mid (\bar{\mathbf{Y}}_1^{[n]})^{1/3}, \mathcal{G}) \leq \frac{4}{3} n \log \bar{P}$. (61) is obtained from Lemma 1. (62) follows from the property that for independent random variables B and C, $I(A; B) \leq I(A; B \mid C)$. As a result, we have $I((\bar{\mathbf{Y}}_1^{[n]})^{1/3}; W_1 \mid \mathcal{G}) \leq I((\bar{\mathbf{Y}}_1^{[n]})^{1/3}; W_1 \mid W_2, \mathcal{G}) \leq H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_2, \mathcal{G})$.

2. Similarly, starting from Fano's Inequality for the second receiver we have,

$$
\begin{aligned}
3nR_2 &\leq I(\bar{\mathbf{Y}}_2^{[n]}; W_2 \mid \mathcal{G}) + 2I(\bar{\mathbf{Y}}_2^{[n]}; W_2 \mid W_1, \mathcal{G}) && (63)\\
&\leq H(\bar{\mathbf{Y}}_2^{[n]} \mid \mathcal{G}) - H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2, \mathcal{G}) + 2H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) \\
&\leq 3n \log \bar{P} - H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2, \mathcal{G}) + 2H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}). && (64)
\end{aligned}
$$

(64) is true for the same reason as (60), i.e., because the entropy of a discrete random variable is bounded by logarithm of the cardinality of its support, i.e., $H(\bar{\mathbf{Y}}_2^{[n]} \mid \mathcal{G}) \leq 3n \log \bar{P}$.

3. Summing the inequalities (62) and (64) we obtain,

$$
\begin{aligned}
& 3nR_1 + 3nR_2 \\
\leq\ & \left(10 + \frac{4}{3}\right) n \log \bar{P} - H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_2, \mathcal{G}) && (65)\\
\leq\ & \left(10 + \frac{4}{3}\right) n \log \bar{P} && (66)
\end{aligned}
$$

where (66) is obtained by a direct application of the sumset inequalities of [5], as explained below. From (66), the sum DoF bound $d_1 + d_2 \leq 3 + \frac{7}{9}$ follows immediately.

4. Finally, we explain how (66) is implied by the sumset inequalities of [5]. Specifically, we need Theorem 4 of [5] to prove that $H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_2, \mathcal{G}) \leq H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2, \mathcal{G})$. In order to apply the result of Theorem 1 (i.e., Theorem 4 of [5]) to our setting, let us set $M = 1$, $K = 2$, $\lambda_{1,1} = \lambda_{2,1} = 1$, $l_1 = l_2 = 1$, and $I_{11} = I_{21} = \{1\}$. Then, the inequality (29) reduces to,

$$H(Z_1^{[n]}, Z_2^{[n]} \mid W, \mathcal{G}) \quad \geq \quad H(Z_{11}^{[n]}, Z_{21}^{[n]} \mid W) + n\ o(\log \bar{P}). \tag{67}$$

Let us specialize $W = W_2$ and define $Z_1(t)$, $Z_2(t)$, $Z_{11}(t)$, $Z_{21}(t)$, $t \in [n]$, as,

$$\bar{Y}_{21}(t) = L^g_{21(t)}((\bar{X}_1(t))^{1/3}, (\bar{X}_2(t))^{1/3}, \bar{X}_3(t), \bar{X}_4(t), \bar{X}_5(t)) \qquad = Z_1(t) \tag{68}$$

$$\bar{Y}_{22}(t) = L^g_{22(t)}((\bar{X}_1(t))^{1/3}, (\bar{X}_2(t))^{1/3}, \bar{X}_3(t), \bar{X}_4(t), \bar{X}_5(t)) \qquad = Z_2(t) \tag{69}$$

$$(\bar{Y}_{11}(t))^{1/3} \doteq L^g_{z11(t)}((\bar{X}_1(t))^{1/3}, (\bar{X}_2(t))^{1/3}) \qquad = Z_{11}(t) \tag{70}$$

$$(\bar{Y}_{12}(t))^{1/3} \doteq L^g_{z21(t)}(t)((\bar{X}_1(t))^{1/3}, (\bar{X}_2(t))^{1/3}) \qquad = Z_{21}(t) \tag{71}$$

Note that L^g functions can be used instead of L^h functions in Z_{11}, Z_{21}, because it only weakens the result of Theorem 1. In other words, Theorem 1 makes the stronger claim that (181) holds even if channel coefficients are chosen as arbitrary constants in Z_{11}, Z_{21}. Since the claim is true for arbitrary constants, it is also true for randomly chosen coefficients, i.e., L^g functions may be used instead of L^h functions in Z_{11}, Z_{21}. Next, consider the '$\doteq$' in (70) and (71). This is justified as follows.

Let us prove (70), and (71) is similarly implied. In order to prove (70) we will show that $Z_{11}(t) = (\bar{Y}_{11}(t))^{1/3} - \delta_{11}(t)$ where $H(\delta_{11}(t))$ is bounded by a constant which does not scale with P. Since adding or subtracting bounded entropy noise terms can only make a difference of the order of $no(\log(P))$ which is inconsequential in the DoF sense, the $\doteq$ in (70) and (71) is justified.

$$\begin{aligned}
& (\bar{Y}_{11}(t))^{1/3} \\
= \quad & \left\lfloor \frac{\bar{Y}_{11}(t) - \bar{P}\left\lfloor \frac{\bar{Y}_{11}(t)}{\bar{P}} \right\rfloor}{\bar{P}^{2/3}} \right\rfloor && (72)\\
= \quad & \left\lfloor \frac{\bar{Y}_{11}(t)}{\bar{P}^{2/3}} \right\rfloor - \left\lfloor \frac{\bar{P}\left\lfloor \frac{\bar{Y}_{11}(t)}{\bar{P}} \right\rfloor}{\bar{P}^{2/3}} \right\rfloor + \delta_a(t) && (73)\\
= \quad & \left\lfloor \frac{\bar{Y}_{11}(t)}{\bar{P}^{2/3}} \right\rfloor - \delta_b(t) + \delta_a(t) && (74)\\
= \quad & \left\lfloor \frac{\lfloor g_1(t)X_1(t)\rfloor + \lfloor g_2(t)X_2(t)\rfloor + \sum_{k=3}^{5}\lfloor g_k(t)(X_k(t))^{1/2}\rfloor}{\bar{P}^{2/3}} \right\rfloor - \delta_b(t) + \delta_a(t) && (75)\\
= \quad & \lfloor g_1(t)(X_1(t))^{1/3}\rfloor + \lfloor g_2(t)(X_2(t))^{1/3}\rfloor + \delta_c(t) - \delta_b(t) + \delta_a(t) && (76)\\
= \quad & Z_{11}(t) + \delta_{11}(t) && (77)
\end{aligned}$$

where $\delta_{11}(t) = \delta_a(t) - \delta_b(t) + \delta_c(t)$. Here, $\delta_a(t)$ is a random variable which can only take values from the set $\{-1, 0, 1\}$ as $\lfloor A\rfloor + \lfloor B\rfloor - \lfloor A + B\rfloor \in \{-1, 0, 1\}$ for any real numbers A and B. Next, consider δ_b, whose entropy is bounded as follows.

$$H(\delta_b) \quad \leq \quad H\left(\left\lfloor \frac{\bar{Y}_{11}(t)}{\bar{P}} \right\rfloor\right) \leq O(1) \tag{78}$$

(78) is true as $|\bar{Y}_{11}(t)| \leq 5\Delta\bar{P}$. Similarly, $H(\delta_c) \leq O(1)$.

Thus, from (181) we have,

$$\begin{aligned} H(Z_{11}^{[n]}, Z_{21}^{[n]} \mid W_2, \mathcal{G}) &\leq H(Z_1^{[n]}, Z_2^{[n]} \mid W_2, \mathcal{G}) + n\ o(\log \bar{P}) && (79)\\ \Rightarrow H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_2, \mathcal{G}) &\leq H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]} \mid W_2, \mathcal{G}) + n\ o(\log \bar{P}) && (80)\\ &\leq H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2, \mathcal{G}) + n\ o(\log \bar{P}). && (81) \end{aligned}$$

7 Example 2. $(M, N_1, N_2) = (4, 1, 3)$ with $(\beta_1, \beta_2) = (\frac{1}{4}, \frac{1}{2})$

For the two-user$(M, N_1, N_2) = (4, 1, 3)$ with $(\beta_1, \beta_2) = (\frac{1}{4}, \frac{1}{2})$ levels of partial CSIT, from Theorem 2 the DoF region is computed as,

$$\mathcal{D}_1 = \Big\{(d_1, d_2) \in \mathbb{R}^{2+}: \quad d_1 \leq 1, \quad d_2 \leq 3, \quad d_1 + \frac{d_2}{3} \leq \frac{5}{4}, \quad d_1 + d_2 \leq 3 + \frac{1}{16}\Big\} \tag{82}$$

and $\beta_o = \frac{1}{16}$ from (45). The challenge is to prove the bound (44), i.e., $d_1 + d_2 \leq 3 + \frac{1}{16}$.

7.1 Deterministic Model

Similar to Section 6.1, the deterministic transformation produces a BC with input $\bar{\mathbf{X}}(t) = \bar{\mathbf{X}}_a(t) \bigtriangledown \bar{\mathbf{X}}_c(t)$, and outputs $\bar{\mathbf{Y}}_1(t), \bar{\mathbf{Y}}_2(t)$.

$$\begin{aligned} \bar{Y}_1(t) &= L_{1(t)}^g \Big(\bar{\mathbf{X}}_a(t) \bigtriangledown (\bar{\mathbf{X}}_c(t))^{3/4}\Big) && (83)\\ \bar{\mathbf{Y}}_2(t) &= \big[\bar{Y}_{21}(t) \quad \bar{Y}_{22}(t)\big]^T && (84)\\ \bar{Y}_{2r}(t) &= L_{2r(t)}^g \Big((\bar{\mathbf{X}}_a(t))^{1/2} \bigtriangledown \bar{\mathbf{X}}_c(t)\Big), \forall r \in [2] && (85) \end{aligned}$$

where $\bar{\mathbf{X}}_a(t)$ and $\bar{\mathbf{X}}_c(t)$ are defined as,

$$\begin{aligned} \bar{\mathbf{X}}_a(t) &= \bar{X}_1(t) && (86)\\ \bar{\mathbf{X}}_c(t) &= \big[\bar{X}_2(t) \quad \bar{X}_2(t) \quad \bar{X}_3(t)\big]^T && (87) \end{aligned}$$

and $\bar{X}_m(t) \in \{0, 1, \cdots, \bar{P}\}$, $\forall m \in [4]$.

7.2 A Key Lemma

To prove the bound $d_1 + d_2 \leq 3 + \frac{1}{16}$ we need the following lemma.

Lemma 2 *For the two-user MIMO BC with* $(M, N_1, N_2) = (4, 1, 3)$ *and* $(\beta_1, \beta_2) = (\frac{1}{4}, \frac{1}{2})$,

$$H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) \leq 4H(\bar{Y}_1^{[n]} \mid W_1, \mathcal{G}) - 3H((\bar{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G}) + \frac{3}{4} n \log \bar{P} + n\ o\ (\log \bar{P}) \tag{88}$$

See Figure 6 for an accompanying illustration for Lemma 2. The proof of Lemma 2 is presented in Appendix B.

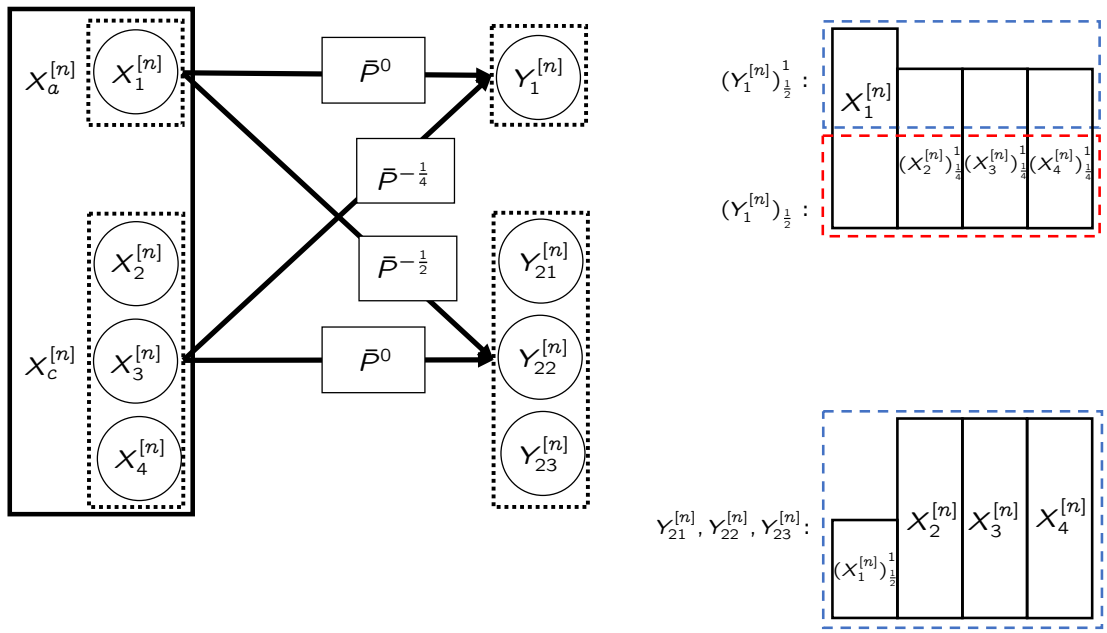

Figure 6: Illustration corresponding to Lemma 2.

7.3 Proof of the Bound $d_1 + d_2 \leq 3 + \frac{1}{16}$

1. Starting with Fano's Inequality for the first receiver, we have,

$$
\begin{aligned}
& 4nR_1 \\
\leq\ & H(\bar{Y}_1^{[n]} \mid \mathcal{G}) - H(\bar{Y}_1^{[n]} \mid W_1, \mathcal{G}) + 3H(\bar{Y}_1^{[n]} \mid \mathcal{G}) - 3H(\bar{Y}_1^{[n]} \mid W_1, \mathcal{G}) && (89)\\
=\ & H(\bar{Y}_1^{[n]} \mid \mathcal{G}) - H(\bar{Y}_1^{[n]} \mid W_1, \mathcal{G}) \\
& +3H((\bar{Y}_1^{[n]})_{1/2}, (\bar{Y}_1^{[n]})^{1/2} \mid \mathcal{G}) - 3H((\bar{Y}_1^{[n]})_{1/2}, (\bar{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G}) && (90)\\
=\ & H(\bar{Y}_1^{[n]} \mid \mathcal{G}) - H(\bar{Y}_1^{[n]} \mid W_1, \mathcal{G}) \\
& +3H((\bar{Y}_1^{[n]})^{1/2} \mid \mathcal{G}) - 3H((\bar{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G}) \\
& +3H((\bar{Y}_1^{[n]})_{1/2} \mid (\bar{Y}_1^{[n]})^{1/2}, \mathcal{G}) - 3H((\bar{Y}_1^{[n]})_{1/2} \mid (\bar{Y}_1^{[n]})^{1/2}, W_1, \mathcal{G}) && (91)\\
\leq\ & \frac{5}{2} n \log \bar{P} - H(\bar{Y}_1^{[n]} \mid W_1, \mathcal{G}) - 3H((\bar{Y}_1^{[n]})_{1/2} \mid (\bar{Y}_1^{[n]})^{1/2}, W_1, \mathcal{G}) \\
& +3H((\bar{Y}_1^{[n]})^{1/2} \mid \mathcal{G}) - 3H((\bar{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G}) && (92)\\
\leq\ & \frac{13}{4} n \log \bar{P} - H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_1, \mathcal{G}) + 3H((\bar{Y}_1^{[n]})^{1/2} \mid \mathcal{G}) - 3H((\bar{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G}) \\
& +n\ o(\log \bar{P}) && (93)\\
\leq\ & \frac{13}{4} n \log \bar{P} - H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_1, \mathcal{G}) + 3H((\bar{Y}_1^{[n]})^{1/2} \mid W_2, \mathcal{G}) + n\ o(\log \bar{P}) && (94)
\end{aligned}
$$

where (129) follows from Definition 2 and (91) is true from the chain rule. (92) is concluded as the entropy of a random variable is bounded by the logarithm of the cardinality of its

support, i.e., $3H((\bar{Y}_1^{[n]})_{1/2} \mid (\bar{Y}_1^{[n]})^{1/2}, \mathcal{G}) \leq \frac{3}{2} n \log \bar{P}$, $H(\bar{Y}_1^{[n]} \mid \mathcal{G}) \leq n \log \bar{P}$. (93) is obtained as from Lemma 2 and the chain rule we have

$$\begin{aligned} & H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_1, \mathcal{G}) \\ \leq \; & 4H(\bar{Y}_1^{[n]} \mid W_1, \mathcal{G}) - 3H((\bar{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G}) + \frac{3}{4} n \log \bar{P} + n \; o\,(\log \bar{P}) \\ = \; & H(\bar{Y}_1^{[n]} \mid W_1, \mathcal{G}) + 3H((\bar{Y}_1^{[n]})_{1/2} \mid (\bar{Y}_1^{[n]})^{1/2}, W_1, \mathcal{G}) + \frac{3}{4} n \log \bar{P} + n \; o\,(\log \bar{P}) \quad (95) \end{aligned}$$

(94) is true because $I(A; B) \leq I(A; B \mid C)$ when B is independent of C. As a result, we have $I((\bar{Y}_1^{[n]})^{1/2}; W_1 \mid \mathcal{G}) \leq H((\bar{Y}_1^{[n]})^{1/2} \mid W_2, \mathcal{G})$.

2. Similarly, writing Fano's Inequality for the second receiver we have,

$$\begin{aligned} nR_2 &\leq I(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]}; W_2 \mid \mathcal{G}) \\ &= H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid \mathcal{G}) - H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_2, \mathcal{G}) \quad (96) \\ nR_2 &\leq I(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]}; W_2 \mid W_1, \mathcal{G}) \\ &= H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_1, \mathcal{G}) \quad (97) \end{aligned}$$

Scaling (96) and (97) by 3 and 1 respectively, and summing them together we have,

$$\begin{aligned} & 4nR_2 \\ \leq \; & 3H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid \mathcal{G}) - 3H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_2, \mathcal{G}) + H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_1, \mathcal{G}) \\ \leq \; & 9n \log \bar{P} - 3H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_2, \mathcal{G}) + H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_1, \mathcal{G}) \quad (98) \end{aligned}$$

(98) is concluded similar to (92) as the entropy of a random variable is bounded by logarithm of the cardinality of its support, i.e., $H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid \mathcal{G}) \leq 2n \log \bar{P}$.

3. Summing the inequalities (94) and (98) results in,

$$\begin{aligned} & 4nR_1 + 4nR_2 \\ \leq \; & \frac{49}{4} n \log \bar{P} - 3H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_2, \mathcal{G}) + 3H((\bar{Y}_1^{[n]})^{1/2} \mid W_2, \mathcal{G}) + n \; o(\log \bar{P}) \quad (99) \\ \leq \; & \frac{49}{4} n \log \bar{P} + n \; o(\log \bar{P}) \quad (100) \end{aligned}$$

Dividing (100) by $4 \log \bar{P}$, $d_1 + d_2 \leq 3 + \frac{1}{16}$ is obtained.

4. The justification for (100) is similar to (66) as $(\bar{Y}_1(t))^{1/2}$ is a bounded density linear combination of random variables $(\bar{X}_1(t))^{1/2}, (\bar{X}_2(t))^1_{\frac{3}{4}}, (\bar{X}_3(t))^1_{\frac{3}{4}}, (\bar{X}_4(t))^1_{\frac{3}{4}}$ while $\bar{Y}_{21}(t)$ is a bounded density linear combination of random variables $(\bar{X}_1(t))^{1/2}, \bar{X}_2(t), \bar{X}_3(t), \bar{X}_4(t)$. Thus, we have

$$\begin{aligned} & H((\bar{Y}_1^{[n]})^{1/2} \mid W_2, \mathcal{G}) \\ \leq \; & H(\bar{Y}_{21}^{[n]} \mid W_2, \mathcal{G}) + n \; o(\log \bar{P}) \quad (101) \\ \leq \; & H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]}, \bar{Y}_{23}^{[n]} \mid W_2, \mathcal{G}) + n \; o(\log \bar{P}). \quad (102) \end{aligned}$$

8 Proof of Theorem 2

To prove Theorem 2 we only need to prove the outer bound (44). The proof for the general setting follows closely along the lines of the examples presented above. We start, as before, with the corresponding deterministic model.

8.1 Deterministic Channel Model

For all $t \in [n]$, the channel outputs in the deterministic model are $\bar{\mathbf{Y}}_1(t)$ and $\bar{\mathbf{Y}}_2(t)$, which are defined as follows.

$$\begin{aligned}
\bar{\mathbf{Y}}_1(t) &= \begin{bmatrix}\bar{Y}_{11}(t) & \bar{Y}_{12}(t) & \cdots & \bar{Y}_{1N_1}(t)\end{bmatrix}^T && (103)\\
\bar{Y}_{1r}(t) &= L^g_{1r}(t)\left(\bar{\mathbf{X}}_a(t) \bigtriangledown \bar{\mathbf{X}}_b(t) \bigtriangledown (\bar{\mathbf{X}}_c(t))^1_{\beta_1}\right), \forall r \in [N_1] && (104)\\
\bar{\mathbf{Y}}_2(t) &= \begin{bmatrix}\bar{Y}_{21}(t) & \bar{Y}_{22}(t) & \cdots & \bar{Y}_{2N_2}(t)\end{bmatrix}^T && (105)\\
\bar{Y}_{2r}(t) &= L^g_{2r}(t)\left((\bar{\mathbf{X}}_a(t))^{1-\beta_2} \bigtriangledown \bar{\mathbf{X}}_b(t) \bigtriangledown \bar{\mathbf{X}}_c(t)\right), \forall r \in [N_2] && (106)
\end{aligned}$$

and $\bar{\mathbf{X}}_a(t), \bar{\mathbf{X}}_b(t), \bar{\mathbf{X}}_c(t)$ are defined as

$$\begin{aligned}
\bar{\mathbf{X}}(t) &= [\bar{X}_1(t), \bar{X}_2(t), \cdots, \bar{X}_M(t)] && (107)\\
\bar{\mathbf{X}}_a(t) &= [\bar{\mathbf{X}}(t)]_{0\to M-N_2} && (108)\\
\bar{\mathbf{X}}_b(t) &= [\bar{\mathbf{X}}(t)]_{M-N_2\to N_1+N_2-M} && (109)\\
\bar{\mathbf{X}}_c(t) &= [\bar{\mathbf{X}}(t)]_{N_1\to M-N_1} && (110)
\end{aligned}$$

and the random variables $\bar{X}_m(t)$ take values from the set $\{0, 1, \cdots, \bar{P}\}$, i.e.,

$$\bar{X}_m(t) \in \{0, 1, \cdots, \bar{P}\}, \forall m \in [M], t \in [n]. \quad (111)$$

8.2 Useful Lemma

The following lemma from [14] will be useful, and is reproduced here for the sake of completeness.

Lemma 3 *[$N_1 \le N_2$ [14]] Define the two random variables $\bar{\mathbf{U}}_1$ and $\bar{\mathbf{U}}_2$ as,*

$$\begin{aligned}
\bar{\mathbf{U}}_1 &= \left(U^{[n]}_{11}, U^{[n]}_{12}, \cdots, U^{[n]}_{1N_1}\right) && (112)\\
\bar{\mathbf{U}}_2 &= \left(U^{[n]}_{21}, U^{[n]}_{22}, \cdots, U^{[n]}_{2N_2}\right) && (113)
\end{aligned}$$

where for any $t \in [n]$, $U_{1j}(t)$ and $U_{2j}(t)$ are defined as,

$$\begin{aligned}
U_{1j}(t) &= L^g_{1j}(t)\left((\bar{\mathbf{V}}_1(t))^\eta_{\eta-\lambda_{11}} \bigtriangledown (\bar{\mathbf{V}}_2(t))^\eta_{\eta-\lambda_{12}} \bigtriangledown \cdots \bigtriangledown (\bar{\mathbf{V}}_l(t))^\eta_{\eta-\lambda_{1l}}\right), \forall j \in [N_1] && (114)\\
U_{2j}(t) &= L^g_{2j}(t)\left((\bar{\mathbf{V}}_1(t))^\eta_{\eta-\lambda_{21}} \bigtriangledown (\bar{\mathbf{V}}_2(t))^\eta_{\eta-\lambda_{22}} \bigtriangledown \cdots \bigtriangledown (\bar{\mathbf{V}}_l(t))^\eta_{\eta-\lambda_{2l}}\right), \forall j \in [N_2] && (115)
\end{aligned}$$

where $\bar{\mathbf{V}}_i(t) = \begin{bmatrix}\bar{V}_{i1}(t) & \cdots & \bar{V}_{iM_i}(t)\end{bmatrix}^T$, $\bar{V}_{im}(t) \in \mathcal{X}_\eta$ are all independent of $\mathcal{G}$, and $0 \le \lambda_{1i}, \lambda_{2i} \le \eta$ for all $i \in [l]$. Without loss of generality, $(\lambda_{1i} - \lambda_{2i})^+$ are sorted in descending order, i.e.,

$(\lambda_{1i} - \lambda_{2i})^+ \geq (\lambda_{1i'} - \lambda_{2i'})^+$ *if* $1 \leq i < i' \leq l$. *Then, for any acceptable*[4] *random variable* W, *if* $N_1 \leq \min(N_2, \sum_{i=1}^{l} M_i)$, *then we have,*

$$\begin{aligned} & H(\bar{\mathbf{U}}_1 \mid W, \mathcal{G}) - H(\bar{\mathbf{U}}_2 \mid W, \mathcal{G}) \\ \leq \; & n\big((N_1 - \sum_{i=1}^{s} M_i)(\lambda_{1,s+1} - \lambda_{2,s+1})^+ + \sum_{i=1}^{s} M_i(\lambda_{1i} - \lambda_{2i})^+\big) \log \bar{P} + n \; o\;(\log \bar{P}) \end{aligned} \quad (116)$$

where s *must satisfy the condition* $\sum_{i=1}^{s} M_i \leq N_1 < \sum_{i=1}^{s+1} M_i$.

8.3 Split $\bar{\mathbf{Y}}_1$ into $\bar{\mathbf{Y}}_{1a}, \bar{\mathbf{Y}}_{1\tilde{a}}$

Since $\bar{\mathbf{X}}_a$ is a vector random variable of size $M - N_2$, using a change of basis operation at the receiver, the N_1 dimensional vector $\bar{\mathbf{Y}}_1$ can be partitioned into $\bar{\mathbf{Y}}_{1\tilde{a}}$, which is its projection into the $N_1 + N_2 - M$ dimensional space that does not contain $\bar{\mathbf{X}}_a(t)$, and a projection $\bar{\mathbf{Y}}_{1a}$ into the $M - N_2$ dimensional space that contains $\bar{\mathbf{X}}_a(t)$.

$$\bar{\mathbf{Y}}_{1\tilde{a}}(t) = \begin{bmatrix} \bar{Y}_{11\tilde{a}}(t) & \bar{Y}_{12\tilde{a}}(t) & \cdots & \bar{Y}_{1N_1+N_2-M\tilde{a}}(t) \end{bmatrix}^T \quad (117)$$

$$\bar{Y}_{1r\tilde{a}}(t) = L^g_{1r\tilde{a}}(t)\left(\bar{\mathbf{X}}_b(t) \bigtriangledown (\bar{\mathbf{X}}_c(t))^1_{\beta_1}\right), \forall r \in [N_1 + N_2 - M] \quad (118)$$

$$\bar{\mathbf{Y}}_{1a}(t) = \begin{bmatrix} \bar{Y}_{11a}(t) & \bar{Y}_{12a}(t) & \cdots & \bar{Y}_{1M-N_2a}(t) \end{bmatrix}^T \quad (119)$$

$$\bar{Y}_{1ra}(t) = L^g_{1ra}(t)\left(\bar{\mathbf{X}}_a(t) \bigtriangledown \bar{\mathbf{X}}_b(t) \bigtriangledown (\bar{\mathbf{X}}_c(t))^1_{\beta_1}\right), \forall r \in [M - N_2]. \quad (120)$$

8.3.1 Proof of bound (44) when $\beta_1 + \beta_2 \geq 1$

When $\beta_1 + \beta_2 \geq 1$, the bound (44) reduces to,

$$d_1 + d_2 \leq N_2 + (M - N_2)\beta_o \quad (121)$$

where β_o is equal to,

$$\beta_o = \frac{N_1 - N_2 + (N_2 - N_1)\beta_2 + (M - N_1)\beta_1}{M - N_1} \quad (122)$$

Corresponding to Lemma 1, in this general setting we need the following lemma which is the key to the proof of the outer bound.

Lemma 4 *For the two-user* (M, N_1, N_2) *MIMO BC with partial CSIT where* $\beta_1 + \beta_2 \geq 1$, *we have,*

$$\begin{aligned} & \hat{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) \\ \leq \; & (\hat{N}_1 + \hat{N}_2) H(\bar{\mathbf{Y}}_1^n \mid W_1, \mathcal{G}) - \hat{N}_2 H(\bar{\mathbf{Y}}^n_{1\tilde{a}}, (\bar{\mathbf{Y}}^n_{1a})^{1-\beta_2} \mid W_1, \mathcal{G}) + n\hat{N}_0 \log \bar{P} + n \; o\;(\log \bar{P}) \end{aligned} \quad (123)$$

where the numbers $\hat{N}_0$, $\hat{N}_1$ *and* $\hat{N}_2$ *are defined as,*

$$\hat{N}_0 = \hat{N}_1(\hat{N}_2 + \hat{N}_1)\beta_1 \quad (124)$$

$$\hat{N}_1 = M - N_2 \quad (125)$$

$$\hat{N}_2 = N_2 - N_1 \quad (126)$$

[4] Let $\mathcal{G}(U) \subset \mathcal{G}$ denote the set of all bounded density channel coefficients that appear in $\bar{\mathbf{U}}_1, \bar{\mathbf{U}}_2$. W is acceptable if conditioned on any $\mathcal{G}_o \subset (\mathcal{G}/\mathcal{G}(U)) \cup \{W\}$, the channel coefficients $\mathcal{G}(U)$ satisfy the bounded density assumption. For instance, any random variable W independent of $\mathcal{G}$ can be utilized in Lemma 3.

See Appendix C for proof of Lemma 4. Now, let us prove the bound (121).

1. Starting with Fano's Inequality for the first receiver and suppressing $no(\log(P))$ terms that are inconsequential for DoF, we have,

$$
\begin{aligned}
& n(\hat{N}_1+\hat{N}_2)R_1 \\
& \leq (\hat{N}_1+\hat{N}_2)I(W_1;\bar{\mathbf{Y}}_1^{[n]} \mid \mathcal{G}) && (127)\\
& = \hat{N}_1 H(\bar{\mathbf{Y}}_1^{[n]} \mid \mathcal{G}) - (\hat{N}_1+\hat{N}_2)H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1,\mathcal{G}) + \hat{N}_2 H(\bar{\mathbf{Y}}_1^{[n]} \mid \mathcal{G}) && (128)\\
& = \hat{N}_1 H(\bar{\mathbf{Y}}_1^{[n]} \mid \mathcal{G}) - (\hat{N}_1+\hat{N}_2)H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1,\mathcal{G}) + \hat{N}_2 H((\bar{\mathbf{Y}}_{1a}^{[n]})_{\beta_2},(\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2},\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]} \mid \mathcal{G}) && (129)\\
& \leq n\hat{N}_1 N_1 \log(\bar{P}) - (\hat{N}_1+\hat{N}_2)H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1,\mathcal{G}) + \hat{N}_2 H(\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]},(\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2} \mid \mathcal{G}) \\
& \qquad + \hat{N}_2 H((\bar{\mathbf{Y}}_{1a}^{[n]})_{\beta_2} \mid \bar{\mathbf{Y}}_{1\tilde{a}}^{[n]},(\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2},\mathcal{G}) && (130)\\
& \leq n(\hat{N}_1 N_1 + \hat{N}_1\hat{N}_2\beta_2)\log(\bar{P}) - (\hat{N}_1+\hat{N}_2)H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1,\mathcal{G}) + \hat{N}_2 H(\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]},(\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2} \mid \mathcal{G}) && (131)\\
& \leq n(\hat{N}_1 N_1 + \hat{N}_1\hat{N}_2\beta_2)\log\bar{P} + n\hat{N}_o\log(\bar{P}) - \hat{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1,\mathcal{G}) - \hat{N}_2 H(\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]},(\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2} \mid W_1,\mathcal{G}) \\
& \qquad + \hat{N}_2 H(\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]},(\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2} \mid \mathcal{G}) && (132)\\
& = n(\hat{N}_0 + \hat{N}_1 N_1 + \hat{N}_1\hat{N}_2\beta_2)\log\bar{P} - \hat{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1,\mathcal{G}) + \hat{N}_2 I(W_1;\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]},(\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2} \mid \mathcal{G}) && (133)\\
& \leq n(\hat{N}_0 + \hat{N}_1 N_1 + \hat{N}_1\hat{N}_2\beta_2)\log\bar{P} - \hat{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1,\mathcal{G}) + \hat{N}_2 H(\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]},(\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2} \mid W_2,\mathcal{G}) && (134)
\end{aligned}
$$

Here (129) follows from Definition 2. The chain rule of entropy, and the fact that since User 1 has only N_1 antennas, the entropy of $\bar{\mathbf{Y}}_1^{[n]}$ cannot be more than $nN_1\log(\bar{P}) + no(\log(P))$, justifies (130). Similarly, (131) is obtained because the entropy of a random variable is bounded by logarithm of the cardinality of its support, i.e., $\hat{N}_2 H((\bar{\mathbf{Y}}_{1b}^{[n]})_{\beta_2} \mid \bar{\mathbf{Y}}_{1a}^{[n]},(\bar{\mathbf{Y}}_{1b}^{[n]})^{1-\beta_2},\mathcal{G}) \leq n\hat{N}_2\hat{N}_1\beta_2\log\bar{P} + no(\log(P))$. Applying Lemma 4 to (131) produces (132). Finally, (134) is true because $I(A;B) \leq I(A;B|C) \leq H(B \mid C)$ whenever A is independent of C.

2. Similarly, starting with Fano's Inequality for the second receiver we have,

$$
\begin{aligned}
& n(\hat{N}_1+\hat{N}_2)R_2 \\
\leq\ & \hat{N}_1 I(\bar{\mathbf{Y}}_2^{[n]};W_2 \mid W_1,\mathcal{G}) + \hat{N}_2 I(\bar{\mathbf{Y}}_2^{[n]};W_2 \mid \mathcal{G}) && (135)\\
\leq\ & \hat{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1,\mathcal{G}) + \hat{N}_2 H(\bar{\mathbf{Y}}_2^{[n]} \mid \mathcal{G}) - \hat{N}_2 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2,\mathcal{G}) \\
\leq\ & \hat{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1,\mathcal{G}) + n\hat{N}_2 N_2\log\bar{P} - \hat{N}_2 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2,\mathcal{G}) && (136)
\end{aligned}
$$

(136) is justified similarly as (131), as the entropy of a random variable is bounded by logarithm of the cardinality of its support, i.e., $H(\bar{\mathbf{Y}}_2^{[n]} \mid \mathcal{G}) \leq nN_2\log\bar{P} + no\log(P)$.

3. Summing the inequalities (134) and (136) results in,

$$
\begin{aligned}
& n(\hat{N}_1+\hat{N}_2)R_1 + n(\hat{N}_1+\hat{N}_2)R_2 \\
\leq\ & n(\hat{N}_0 + \hat{N}_1 N_1 + \hat{N}_1\hat{N}_2\beta_2 + \hat{N}_2 N_2)\log\bar{P} - \hat{N}_2 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2,\mathcal{G}) \\
& + \hat{N}_2 H(\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]},(\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2} \mid W_2,\mathcal{G}) && (137)
\end{aligned}
$$

$$\leq \quad n(\hat{N}_0 + \hat{N}_1 N_1 + \hat{N}_1 \hat{N}_2 \beta_2 + \hat{N}_2 N_2) \log \bar{P} \tag{138}$$

(138) is immediately obtained from Lemma 3 by defining $\bar{\mathbf{U}}_1 = (\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]}, (\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2})$ and $\bar{\mathbf{U}}_2 = \bar{\mathbf{Y}}_2^{[n]}$, we have,

$$H(\bar{\mathbf{Y}}_{1\tilde{a}}^{[n]}, (\bar{\mathbf{Y}}_{1a}^{[n]})^{1-\beta_2} \mid W_2, \mathcal{G}) - H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2, \mathcal{G}) \quad \leq \quad n\ o\ (\log \bar{P}) \tag{139}$$

where (139) is true as all the parameters $(\lambda_{1i} - \lambda_{2i})^+$ in Lemma 3 are equal to zero. Finally, applying the DoF limit in (138), the bound $d_1 + d_2 \leq N_2 + (M - N_2)\beta_o$ is concluded.

8.3.2 Proof of bound (44) when $\beta_1 + \beta_2 < 1$

In this section we prove the bound (44) for general (M, N_1, N_2) when $\beta_1 + \beta_2 < 1$. The bound (44) simplifies to,

$$d_1 + d_2 \quad \leq \quad N_2 + (M - N_2)\beta_o \tag{140}$$

where β_o is equal to,

$$\beta_o \quad = \quad \frac{\beta_1 \beta_2 (M - N_2)}{(N_2 - N_1)(1 - \beta_1) + (M - N_2)\beta_2} \tag{141}$$

The proof relies on the following lemma.

Lemma 5 *For the two-user (M, N_1, N_2) MIMO BC with partial CSIT where $\beta_1 + \beta_2 < 1$, we have,*

$$\begin{aligned} & \breve{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) \\ \leq \quad & (\breve{N}_1 + \breve{N}_2) H(\bar{\mathbf{Y}}_1^n \mid W_1, \mathcal{G}) - \breve{N}_2 H(\bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_2} \mid W_1, \mathcal{G}) + n\breve{N}_0 \log \bar{P} + n\ o\ (\log \bar{P}) \end{aligned} \tag{142}$$

where the numbers $\breve{N}_0$, $\breve{N}_1$ and $\breve{N}_2$ are defined as,

$$\breve{N}_0 \quad = \quad \breve{N}_1 (M - N_1)\beta_1 \tag{143}$$

$$\breve{N}_1 \quad = \quad (M - N_2)\beta_2 \tag{144}$$

$$\breve{N}_2 \quad = \quad (N_2 - N_1)(1 - \beta_1) \tag{145}$$

See Figure 7 for the comparison of the two sides of (142).

See Appendix D for proof of Lemma 5. With the aid of Lemma 5, the proof of the bound (140) proceeds along the lines of the corresponding proof for $\beta_1 + \beta_2 \geq 1$ that was presented in Section 8.3.1.

1. Starting from Fano's Inequality for the first receiver and proceeding through the same set of steps as ((129)-(134)) we have,

$$\begin{aligned} & n(\breve{N}_1 + \breve{N}_2) R_1 \\ \leq \quad & n(\breve{N}_0 + \breve{N}_1 N_1 + \breve{N}_1 \breve{N}_2) \log \bar{P} - \breve{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) + \breve{N}_2 H(\bar{\mathbf{Y}}_{1a}^{[n]}, (\bar{\mathbf{Y}}_{1b}^{[n]})^{1-\beta_2} \mid W_2, \mathcal{G}) \\ & + n\ o\ (\log \bar{P}) \end{aligned} \tag{146}$$

where the difference between (134) and (146) is due to the difference in the definitions of $(\hat{N}_1, \hat{N}_2)$ versus $(\breve{N}_1, \breve{N}_2)$.

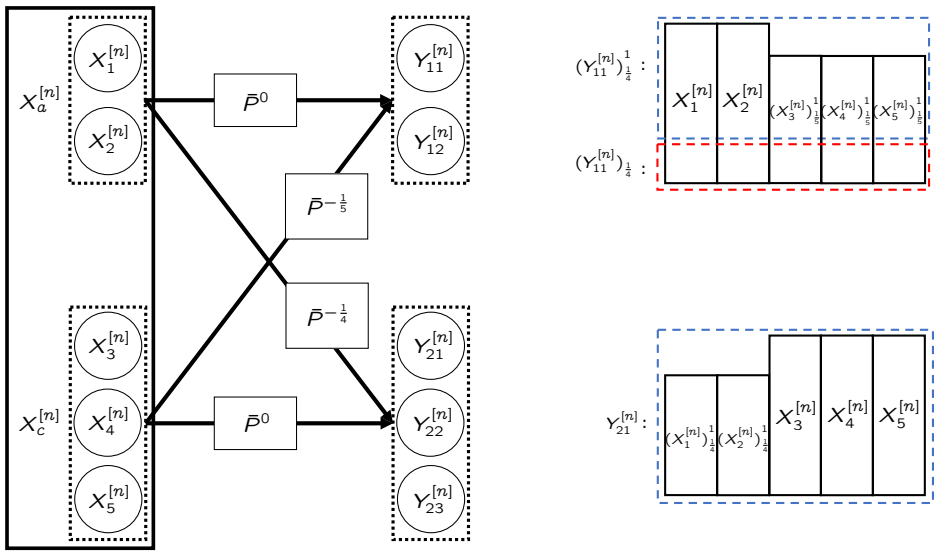

Figure 7: Illustration for Lemma 5

2. Similarly, for the second receiver we have,

$$\begin{aligned} & n(\breve{N}_1+\breve{N}_2)R_2 \\ \leq\ & \breve{N}_2 H(\bar{\mathbf{Y}}_2^{[n]} \mid \mathcal{G}) - \breve{N}_2 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2, \mathcal{G}) + \breve{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) \\ \leq\ & n\breve{N}_2 N_2 \log \bar{P} - \breve{N}_2 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2, \mathcal{G}) + \breve{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) \end{aligned} \tag{147}$$

3. Summing the inequalities (146) and (147) results in,

$$\begin{aligned} & n(\breve{N}_1+\breve{N}_2)R_1 + n(\breve{N}_1+\breve{N}_2)R_2 \\ \leq\ & n(\breve{N}_0 + \breve{N}_1 N_1 + \breve{N}_1\breve{N}_2 + \breve{N}_2 N_2)\log \bar{P} - \breve{N}_2 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_2, \mathcal{G}) \\ & + \breve{N}_2 H(\bar{\mathbf{Y}}_{1a}^{[n]}, (\bar{\mathbf{Y}}_{1b}^{[n]})^{1-\beta_2} \mid W_2, \mathcal{G}) + n\ o\ (\log \bar{P}) \qquad (148) \\ \leq\ & n(\breve{N}_0 + \breve{N}_1 N_1 + \breve{N}_1\breve{N}_2 + \breve{N}_2 N_2)\log \bar{P} + n\ o\ (\log \bar{P}) \qquad (149) \end{aligned}$$

(149) follows from Lemma 3 similar to (138). Finally, applying the DoF limit in (149) we obtain the bound, $d_1 + d_2 \leq N_2 + (M - N_2)\beta_o$.

9 Conclusion

The DoF region of the the two-user MIMO BC with arbitrary levels of partial CSIT was characterized as a function of the number of antennas and the levels of CSIT while perfect CSIR is assumed. The main challenge was deriving an outer bound that captures the difference of entropies caused by asymmetric number of antennas and asymmetric levels of partial CSIT which was accomplished with the aid of sum-set inequalities and AIS approach.

A Proof of Lemma 1

Suppressing $no(\log(P))$ that are inconsequential for DoF, we proceed as follows.

$$
\begin{aligned}
&2H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) \\
&= 2H((\bar{\mathbf{Y}}_2^{[n]}))^{1/2}, (\bar{\mathbf{Y}}_2^{[n]})_{1/2} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) && (150)\\
&\leq 2H((\bar{\mathbf{Y}}_2^{[n]}))^{1/2} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) + 3n \log \bar{P} && (151)\\
&\leq 2H((\bar{X}_3^{[n]}))^{1/2}, (\bar{X}_4^{[n]}))^{1/2}, (\bar{X}_5^{[n]}))^{1/2} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) + 3n \log \bar{P} && (152)\\
&\leq 2H((\bar{\mathbf{Y}}_1^{[n]})^{1/3}, (\bar{X}_3^{[n]}))^{1/2}, (\bar{X}_4^{[n]}))^{1/2}, (\bar{X}_5^{[n]}))^{1/2} \mid W_1, \mathcal{G}) + H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) + 3n \log \bar{P} && (153)\\
&= 2H((\bar{X}_3^{[n]}))^{1/2}, (\bar{X}_4^{[n]}))^{1/2}, (\bar{X}_5^{[n]}))^{1/2} \mid (\bar{\mathbf{Y}}_1^{[n]})^{1/3}, W_1, \mathcal{G}) + 3H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) + 3n \log \bar{P} && (154)\\
&\leq H((\bar{X}_3^{[n]}))^{1/2}, (\bar{X}_4^{[n]}))^{1/2} \mid (\bar{\mathbf{Y}}_1^{[n]})^{1/3}, W_1, \mathcal{G}) + H((\bar{X}_3^{[n]}))^{1/2}, (\bar{X}_5^{[n]}))^{1/2} \mid (\bar{\mathbf{Y}}_1^{[n]})^{1/3}, W_1, \mathcal{G}) \\
&\quad + H((\bar{X}_4^{[n]}))^{1/2}, (\bar{X}_5^{[n]}))^{1/2} \mid (\bar{\mathbf{Y}}_1^{[n]})^{1/3}, W_1, \mathcal{G}) + 3H((\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) + 3n \log \bar{P} && (155)\\
&= H((\bar{X}_3^{[n]}))^{1/2}, (\bar{X}_4^{[n]}))^{1/2}, (\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) + H((\bar{X}_3^{[n]}))^{1/2}, (\bar{X}_5^{[n]}))^{1/2}, (\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) \\
&\quad + H((\bar{X}_4^{[n]}))^{1/2}, (\bar{X}_5^{[n]}))^{1/2}, (\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) + 3n \log \bar{P} && (156)\\
&\leq 3H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) + 3n \log \bar{P} && (157)
\end{aligned}
$$

(150) follows from Definition 2, and (151) is true because $\mathbf{Y}_2$ has 3 antennas and the entropy of the bottom half of the signal power levels on each of them can altogether contribute at most $\frac{3}{2} n \log \bar{P}$. Next, (152) is true because $(\mathbf{Y}_2)^{1/2}$ is a function of $(\bar{X}_3^{[n]}))^{1/2}, (\bar{X}_4^{[n]}))^{1/2}, (\bar{X}_5^{[n]}))^{1/2}$, and (153) simply uses the property that $H(A|C) \leq H(A, B|C)$. The chain rule of entropy produces (154) and (156), and (155) follows from the sub-modularity property of the entropy function, i.e., for any three random variables A, B and C,

$$
2H(A, B, C) \quad \leq \quad H(A, B) + H(A, C) + H(B, C) \tag{158}
$$

Finally, (157) is where we use the sumset inequality from Theorem 1. For instance,

$$
H((\bar{X}_3^{[n]}))^{1/2}, (\bar{X}_5^{[n]}))^{1/2}, (\bar{\mathbf{Y}}_1^{[n]})^{1/3} \mid W_1, \mathcal{G}) \quad \leq \quad H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) + n\ o(\log \bar{P}) \tag{159}
$$

is obtained from Theorem 1 by setting $\lambda_1 = \lambda_2 = \frac{1}{2}$, $I_{11} = I_{21} = 2$, $I_{12} = I_{22} = 1$ and defining $Z_{11}(t)$, $Z_{12}(t)$, $Z_{21}(t)$, $Z_{22}(t)$, $Z_1(t)$, $Z_2(t)$ as,

$$
\begin{aligned}
Z_{11}(t) &= (\bar{Y}_{11}(t))^{1/3} \doteq L^h_{11(t)}((\bar{X}_1(t))^{1/3}, (\bar{X}_2(t))^{1/3}) && (160)\\
Z_{12}(t) &= L^h_{12(t)}((\bar{X}_1(t))_{1/2}, (\bar{X}_2(t))_{1/2}, (\bar{X}_3(t))^{1/2}, (\bar{X}_4(t))^{1/2}, (\bar{X}_5(t))^{1/2}) = (\bar{X}_3(t))^{1/2} && (161)\\
Z_{21}(t) &= (\bar{Y}_{12}(t))^{1/3} \doteq L^h_{21(t)}((\bar{X}_1(t))^{1/3}, (\bar{X}_2(t))^{1/3}) && (162)\\
Z_{22}(t) &= L^h_{12(t)}((\bar{X}_1(t))_{1/2}, (\bar{X}_2(t))_{1/2}, (\bar{X}_3(t))^{1/2}, (\bar{X}_4(t))^{1/2}, (\bar{X}_5(t))^{1/2}) = (\bar{X}_5(t))^{1/2} && (163)\\
Z_1(t) &= \bar{Y}_{11}(t) = L^g_{y11(t)}(\bar{X}_1(t), \bar{X}_2(t), (\bar{X}_3(t))^{1/2}, (\bar{X}_4(t))^{1/2}, (\bar{X}_5(t))^{1/2}) && (164)\\
Z_2(t) &= \bar{Y}_{12}(t) = L^g_{y12(t)}(\bar{X}_1(t), \bar{X}_2(t), (\bar{X}_3(t))^{1/2}, (\bar{X}_4(t))^{1/2}, (\bar{X}_5(t))^{1/2}) && (165)
\end{aligned}
$$

Note that for any i, j the linear combination $L^h_{ij}(t)$ is arbitrary linear combination of random variables satisfying the condition in Definition 6. Thus, some of the coefficients in $L^h_{ij}(t)$ can be chosen to be zero, e.g., we choose $L^h_{12}(t)((\bar{X}_1(t))_{1/2}, (\bar{X}_2(t))_{1/2}, (\bar{X}_3(t))^{1/2}, (\bar{X}_4(t))^{1/2}, (\bar{X}_5(t))^{1/2})$ to be $(\bar{X}_3(t))^{1/2}$. Thus, from Theorem 1 we have,

$$\begin{aligned} H(Z^{[n]}_{11}, Z^{[n]}_{12}, Z^{[n]}_{21}, Z^{[n]}_{22} \mid W_1, \mathcal{G}) &\leq H(Z^{[n]}_1, Z^{[n]}_2 \mid W_1, \mathcal{G}) + n\ o(\log \bar{P}) && (166) \\ \Rightarrow H((\bar{X}^{[n]}_3))^{1/2}, (\bar{X}^{[n]}_5))^{1/2}, (\bar{\mathbf{Y}}^{[n]}_1)^{1/3} \mid W_1, \mathcal{G}) &\leq H(\bar{\mathbf{Y}}^{[n]}_1 \mid W_1, \mathcal{G}) + n\ o(\log \bar{P}) && (167) \end{aligned}$$

See Figure 8 and 9.

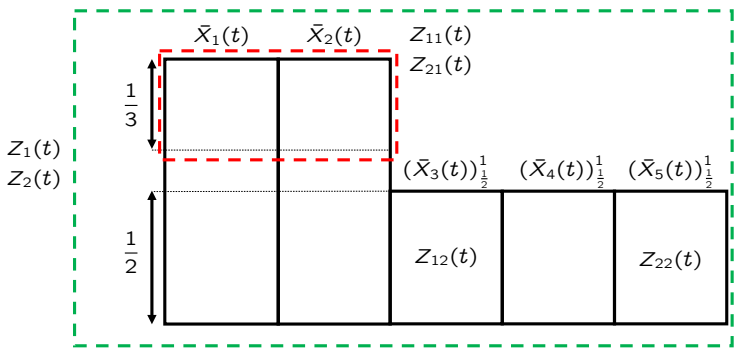

Figure 8: The random variables in ((160)-(165)) are illustrated.

B Proof of Lemma 2

Note that $2H((\bar{Y}^{[n]}_{21})_{1/4}, (\bar{Y}^{[n]}_{22})_{1/4} \mid W_1, \mathcal{G}) \leq \log \bar{P}$. Therefore, it is sufficient to prove,

$$\begin{aligned} & 2H((\bar{Y}^{[n]}_{21})^{3/4}, (\bar{Y}^{[n]}_{22})^{3/4} \mid W_1, \mathcal{G}) + 3H((\bar{Y}^n_1)^{1/2} \mid W_1, \mathcal{G}) \\ \leq\ & 5H(\bar{Y}^n_1 \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) \end{aligned} \tag{168}$$

For any $i \in [7]$, define $C_i(t)$ as

$$C_i(t) = \begin{cases} (\bar{X}_5(t))^{1}_{\frac{3}{4}}, i = 1 \\ (\bar{X}_5(t))^{\frac{3}{4}}_{1/2}, i = 2 \\ L^g_3((\bar{X}_1(t))^{1}_{\frac{3}{4}}, (\bar{X}_2(t))^{1}_{\frac{3}{4}}, (\bar{X}_3(t))^{1}_{\frac{3}{4}}, (\bar{X}_4(t))^{1/2}_{1/4}, (\bar{X}_5(t))^{1/2}_{1/4}), i = 3 \\ (\bar{X}_4(t))^{1}_{\frac{3}{4}}, i = 4 \\ (\bar{X}_4(t))^{\frac{3}{4}}_{1/2}, i = 5 \\ L^g_6((\bar{X}_1(t))^{1}_{\frac{3}{4}}, (\bar{X}_2(t))^{1}_{\frac{3}{4}}, (\bar{X}_3(t))^{1}_{\frac{3}{4}}, (\bar{X}_4(t))^{1/2}_{1/4}, (\bar{X}_5(t))^{1/2}_{1/4}), i = 6 \\ C_1(t), i = 7 \end{cases} \tag{169}$$

Starting from the left side of (168), we have

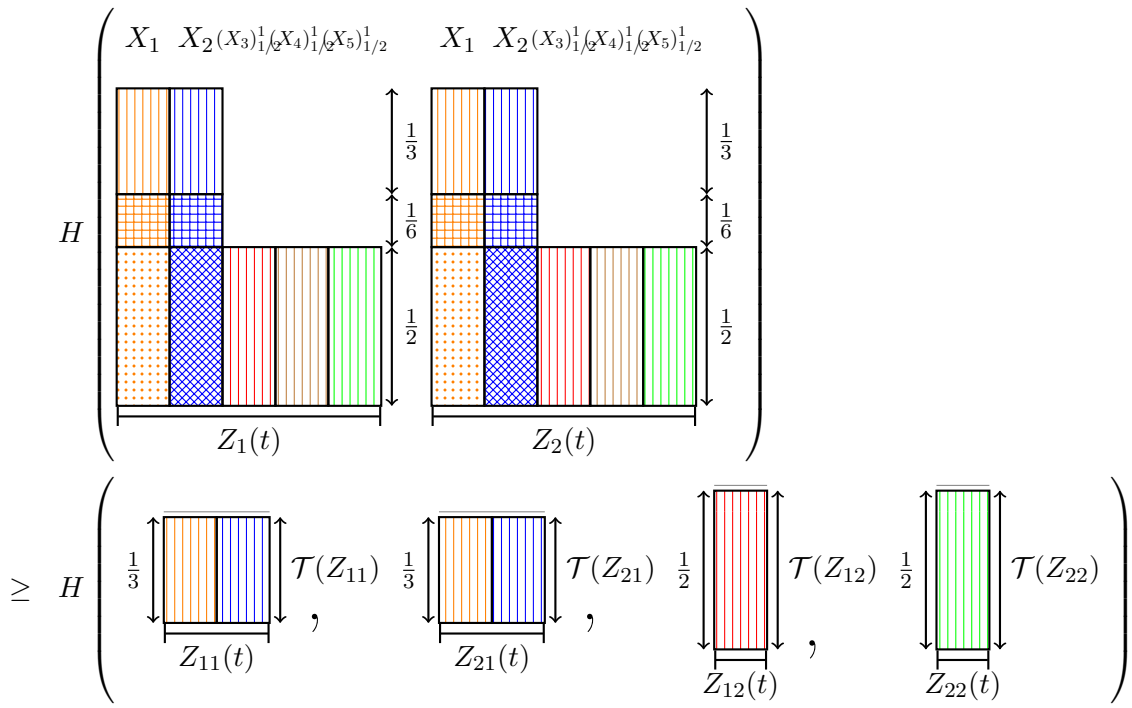

Figure 9: Illustration of (167).

$$
\begin{aligned}
& 2H((\bar{Y}_{21}^{[n]})^{3/4}, (\bar{Y}_{22}^{[n]})^{3/4} \mid W_1, \mathcal{G}) + 3H((\bar{Y}_1^n)^{1/2} \mid W_1, \mathcal{G}) \\
= \;& 2H(C_1^{[n]}, C_2^{[n]}, C_3^{[n]}, C_4^{[n]}, C_5^{[n]}, C_6^{[n]} \mid W_1, \mathcal{G}) + 3H((\bar{Y}_1^n)^{1/2} \mid W_1, \mathcal{G}) && (170) \\
\leq \;& 2H((\bar{Y}_1^{[n]})^{1/2}, C_2^{[n]}, C_3^{[n]}, C_4^{[n]}, C_5^{[n]}, C_6^{[n]} \mid W_1, \mathcal{G}) + 3H((\bar{Y}_1^n)^{1/2} \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (171) \\
= \;& 2H(C_2^{[n]}, C_3^{[n]}, C_4^{[n]}, C_5^{[n]}, C_6^{[n]} \mid (\bar{Y}_1^{[n]})^{1/2}, W_1, \mathcal{G}) + 5H((\bar{Y}_1^n)^{1/2} \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (172) \\
\leq \;& \sum_{i=2}^{6} H(C_i^{[n]}, C_{i+1}^{[n]} \mid (\bar{Y}_1^{[n]})^{1/2}, W_1, \mathcal{G}) + 5H((\bar{Y}_1^n)^{1/2} \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (173) \\
= \;& \sum_{i=2}^{6} H((\bar{Y}_1^{[n]})^{1/2}, C_i^{[n]}, C_{i+1}^{[n]} \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (174) \\
\leq \;& 5H(\bar{Y}_1^n \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (175)
\end{aligned}
$$

(170) follows from Definition 2 and definition of $C_i(t)$ in (169). (172) is true from the chain rule and (173) is concluded from sub-modularity properties of entropy function, i.e., for any m random variables $\{X_1, X_2, \cdots, X_m\}$ where we define X_{k+m} as X_k for positive numbers k we have,

$$
nH(X_1, \cdots, X_m \mid W) \leq \sum_{i=1}^{m} H(X_i, X_{i+1}, \cdots, X_{i+n} \mid W) \tag{176}
$$

for any random variable W if $n \leq m$. (171) follows from Theorem 1 to our setting. Let us set $M = 1$, $N = 5$, $K = 1$, $\lambda_{1,1} = 1$, $l_1 = 1$, and $I_{1,1} = \{1\}$. Then, the inequality (29) reduces to,

$$
H(Z_1^{[n]} \mid W, \mathcal{G}) \quad \geq \quad H(Z_{11}^{[n]} \mid W) + n\ o(\log \bar{P}). \tag{177}
$$

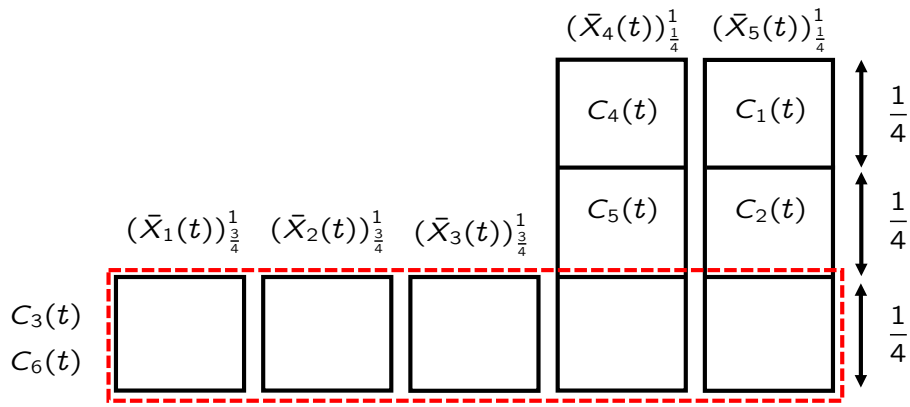

Figure 10: The random variables $(\bar{X}_1(t))^1_{\frac{3}{4}}, (\bar{X}_2(t))^1_{\frac{3}{4}}, (\bar{X}_3(t))^1_{\frac{3}{4}}, (\bar{X}_4(t))^{3/4}, (\bar{X}_5(t))^{3/4}$ and their partitions $C_i(t)$ are specified.

where

$$Z_1(t) = (\bar{Y}_1(t))^{1/2} \tag{178}$$
$$Z_{11}(t) = C_1(t) \tag{179}$$
$$W = C_2^{[n]}, C_3^{[n]}, C_4^{[n]}, C_5^{[n]}, C_6^{[n]}, W_1 \tag{180}$$

(175) follows from Theorem 1 to our setting similar to (171). Let us set $N = 5$, $M = 3$, $K = 1$, $\lambda_{1,3} = \frac{1}{2}, \lambda_{1,1} = \lambda_{1,2} = \frac{1}{4}$, $I_{1,1} = 3, I_{1,2} = \{2, 3\}$, and $I_{1,2,3} = \{1, 2, 3\}$. Then, the inequality (29) reduces to,

$$H(Z_1^{[n]} \mid W, \mathcal{G}) \geq H(Z_{11}^{[n]}, Z_{12}^{[n]}, Z_{13}^{[n]} \mid W) + n\ o(\log \bar{P}). \tag{181}$$

where

$$Z_1(t) = \bar{Y}_1(t) \tag{182}$$
$$Z_{11}(t) = (\bar{Y}_1(t))^{1/2}$$
$$\doteq L^g((\bar{X}_1(t))^{1/2}, (\bar{X}_2(t))^{1/4}, (\bar{X}_3(t))^{1/4}, (\bar{X}_4(t))^{1/4}, (\bar{X}_5(t))^{1/4}) \tag{183}$$
$$Z_{12}(t) = C_i(t) \tag{184}$$
$$Z_{13}(t) = C_{i+1}(t) \tag{185}$$
$$W = W_1 \tag{186}$$

(183) follows similar to (70). Note that the condition (28) is satisfied as

$$\mathcal{T}(Z_{1,2}) + \mathcal{T}(Z_{1,3}) \leq \lambda_{1,1} + \lambda_{1,2} \tag{187}$$
$$\mathcal{T}(Z_{1,3}) \leq \lambda_{1,1} \tag{188}$$

Note that $\mathcal{T}(C_i(t)) = \frac{1}{4}$ from Figure 10. For instance, illustration of

$$H((\bar{Y}_1^{[n]})^{1/2}, C_2^{[n]}, C_3^{[n]} \mid W_1, \mathcal{G}) \leq H(\bar{Y}_1^n \mid W_1, \mathcal{G}) \tag{189}$$

is shown in Figure 11.

C Proof of Lemma 4

1. Define the random variables $\bar{\mathbf{Y}}_{1c}(t)$ and $\bar{\mathbf{Y}}_{1d}(t)$ from the random variables $\bar{\mathbf{Y}}_{1a}(t)$ and $\bar{\mathbf{Y}}_{1b}(t)$ in ((117)-(120)) as,

$$\begin{aligned}
\bar{\mathbf{Y}}_{1c}(t) &= \begin{bmatrix}\bar{y}_{11c}(t) & \bar{y}_{12c}(t) & \cdots & \bar{y}_{1,N_1+N_2-M,c}(t)\end{bmatrix}^T && (190)\\
\bar{y}_{1rc}(t) &= L_{3r}^g(t)\left((\bar{\mathbf{X}}_a(t))^{1-\beta_2} \bigtriangledown \bar{\mathbf{X}}_b(t) \bigtriangledown (\bar{\mathbf{X}}_c(t))^{1-\beta_1}\right), \forall r \in [N_1+N_2-M] && (191)\\
\bar{\mathbf{Y}}_{1d}(t) &= \begin{bmatrix}\bar{y}_{11d}(t) & \bar{y}_{12d}(t) & \cdots & \bar{y}_{1,M-N_2,d}(t)\end{bmatrix}^T && (192)\\
\bar{y}_{1rd}(t) &= L_{3r}^g(t)\left(\bar{\mathbf{X}}_a(t) \bigtriangledown (\bar{\mathbf{X}}_c(t))_{\beta_1}^1\right), \forall r \in [M-N_2] && (193)
\end{aligned}$$

As $[\bar{\mathbf{Y}}_{1c}(t) \bigtriangledown (\bar{\mathbf{Y}}_{1d}(t))^{1-\beta_2}] = \mathbf{A}'(t)[\bar{\mathbf{Y}}_{1a}(t) \bigtriangledown (\bar{\mathbf{Y}}_{1b}(t))^{1-\beta_2}]$ for some invertible matrix $\mathbf{A}'(t)$, the terms $\bar{\mathbf{Y}}_{1a}(t)$ and $\bar{\mathbf{Y}}_{1b}(t)$ are replaced with $\bar{\mathbf{Y}}_{1c}(t)$ and $\bar{\mathbf{Y}}_{1d}(t)$, respectively.[5]

2. As $\bar{\mathbf{X}}_b(t)$ and $\bar{\mathbf{X}}_c(t)$ are vector random variables of size N_1+N_2-M and $M-N_1$, the random variables $\bar{\mathbf{Y}}_{2a}(t)$ and $\bar{\mathbf{Y}}_{2b}(t)$ are defined from the random variable $\bar{\mathbf{Y}}_2(t)$ in (105) and (106) as,

$$\begin{aligned}
\bar{\mathbf{Y}}_{2a}(t) &= \begin{bmatrix}\bar{y}_{21a}(t) & \bar{y}_{22a}(t) & \cdots & \bar{y}_{2,N_1+N_2-M,a}(t)\end{bmatrix}^T && (194)\\
\bar{y}_{2ra}(t) &= L_{4r}^g(t)\left((\bar{\mathbf{X}}_a(t))^{1-\beta_2} \bigtriangledown \bar{\mathbf{X}}_b(t)\right), \forall r \in [N_1+N_2-M] && (195)\\
\bar{\mathbf{Y}}_{2b}(t) &= \begin{bmatrix}\bar{y}_{21b}(t) & \bar{y}_{22b}(t) & \cdots & \bar{y}_{2,M-N_2,b}(t)\end{bmatrix}^T && (196)\\
\bar{y}_{2rb}(t) &= L_{4r}^g(t)\left((\bar{\mathbf{X}}_a(t))^{1-\beta_2} \bigtriangledown \bar{\mathbf{X}}_c(t)\right), \forall r \in [M-N_1] && (197)
\end{aligned}$$

Similarly, as $[\bar{\mathbf{Y}}_{2a}(t) \bigtriangledown (\bar{\mathbf{Y}}_{2b}(t))^{1-\beta_2}] = \mathbf{A}(t)[\bar{\mathbf{Y}}_2(t)]$ for some invertible matrix $\mathbf{A}(t)$, the term $\bar{\mathbf{Y}}_2(t)$ is replaced with $\bar{\mathbf{Y}}_{2a}(t) \bigtriangledown \bar{\mathbf{Y}}_{2b}(t)$.

3. The entropy of a discrete random variable is bounded by logarithm of the cardinality of it, i.e., $H((\bar{\mathbf{Y}}_{2b}^{[n]})_{\beta_1} \mid W_1, \mathcal{G}) \leq n\hat{N}_2\beta_1 \log \bar{P}$. Thus, from the chain rule we have,

$$\begin{aligned}
& \hat{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, \bar{\mathbf{Y}}_{2b}^{[n]} \mid W_1, \mathcal{G}) - n\hat{N}_0 \log \bar{P} \\
= \; & \hat{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, \bar{\mathbf{Y}}_{2b}^{[n]} \mid W_1, \mathcal{G}) - n\hat{N}_1\hat{N}_2\beta_1 \log \bar{P} && (198)\\
\leq \; & \hat{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, \bar{\mathbf{Y}}_{2b}^{[n]} \mid W_1, \mathcal{G}) - \hat{N}_1 H((\bar{\mathbf{Y}}_{2b}^{[n]})_{\beta_1} \mid W_1, \mathcal{G}) && (199)\\
\leq \; & \hat{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, (\bar{\mathbf{Y}}_{2b}^{[n]})_{\beta_1}^1 \mid W_1, \mathcal{G}) && (200)
\end{aligned}$$

where (200) is true from the chain rule.

[5] Note that $\bar{\mathbf{Y}}_{1c}(t)$ and $\bar{\mathbf{X}}_c(t)$ do not exist simultaneously as $\bar{\mathbf{Y}}_{1c}(t)$ and $\bar{\mathbf{X}}_c(t)$ are vector random variables of size $(N_1+N_2-M) \times 1$ and $(M-N_1-N_2) \times 1$, respectively.

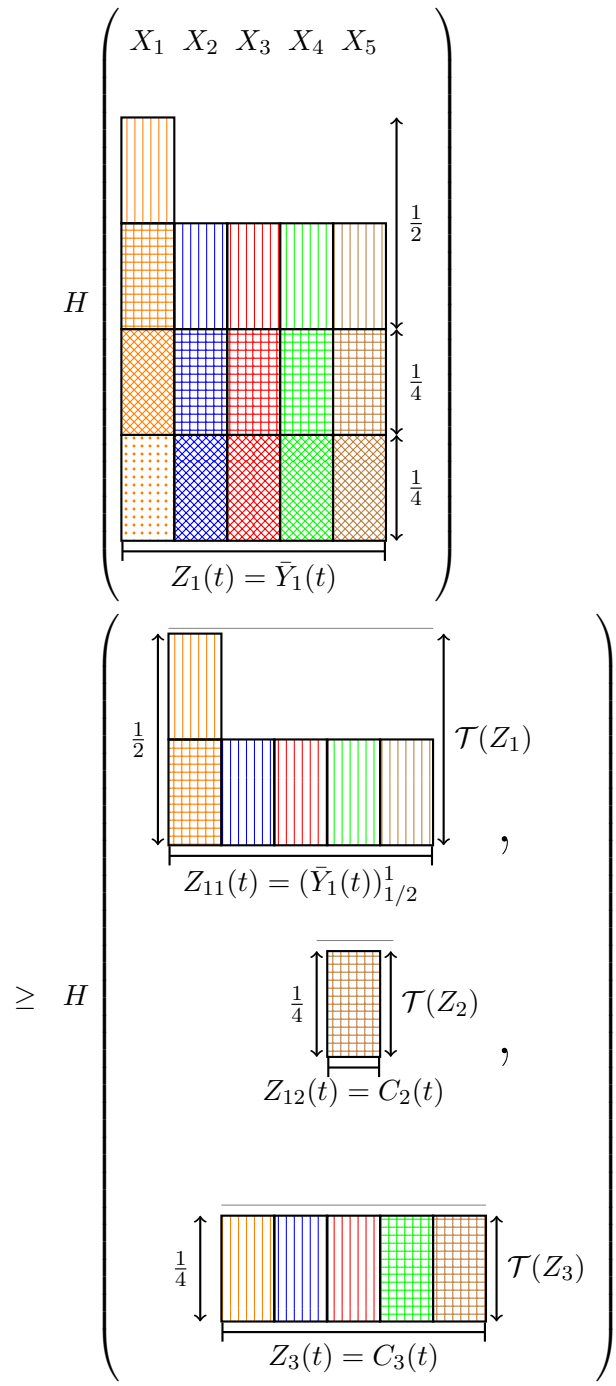

Figure 11: Illustration of (189).

4. Note that,

$$H((\bar{\mathbf{Y}}^n_{1d})_{\beta_2} \mid \bar{\mathbf{Y}}^n_{1c}, (\bar{\mathbf{Y}}^n_{1d})^{1-\beta_2}, W_1, \mathcal{G}) \quad = \quad H(\bar{\mathbf{Y}}^n_1 \mid W_1, \mathcal{G}) - H(\bar{\mathbf{Y}}^n_{1c}, (\bar{\mathbf{Y}}^n_{1d})^{1-\beta_2} \mid W_1, \mathcal{G}) \tag{201}$$

where (201) follows from the chain rule.

From the above observations in order to prove (121), it is sufficient to demonstrate the following inequality,

$$\begin{aligned} & \hat{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, (\bar{\mathbf{Y}}_{2b}^{[n]})^{1-\beta_1} \mid W_1, \mathcal{G}) + \hat{N}_2 H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})^{1-\beta_2} \mid W_1, \mathcal{G}) \\ \leq \; & (\hat{N}_1 + \hat{N}_2) H(\bar{\mathbf{Y}}_1^n \mid W_1, \mathcal{G}) + n \; o\,(\log \bar{P}) \end{aligned} \tag{202}$$

Before proceeding to proof of (202) let us define the random variables $\bar{C}_i(t)$ as the components of $\bar{\mathbf{X}}_c(t)$, i.e.,

$$\bar{C}_i(t) \;=\; \begin{cases} \bar{X}_{i-\hat{N}_1-\hat{N}_2+M}(t) & 0 < i \leq \hat{N}_1 + \hat{N}_2 \\ \bar{C}_{i-\hat{N}_1-\hat{N}_2}(t) & i > \hat{N}_1 + \hat{N}_2 \end{cases} \tag{203}$$

Starting from the left side of (202), we have

$$\begin{aligned} & \hat{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, (\bar{\mathbf{Y}}_{2b}^{[n]})_{\beta_1}^{1} \mid W_1, \mathcal{G}) + \hat{N}_2 H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})^{1-\beta_2} \mid W_1, \mathcal{G}) \\ \leq \; & \hat{N}_1 H(\bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{2b}^{[n]})_{\beta_1}^{1}, (\bar{\mathbf{Y}}_{1d}^{[n]})^{1-\beta_2} \mid W_1, \mathcal{G}) + \hat{N}_2 H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})^{1-\beta_2} \mid W_1, \mathcal{G}) \\ & + n \; o\,(\log \bar{P}) \qquad (204) \\ = \; & \hat{N}_1 H((\bar{\mathbf{Y}}_{2b}^{[n]})_{\beta_1}^{1} \mid \bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})^{1-\beta_2}, W_1, \mathcal{G}) + (\hat{N}_1 + \hat{N}_2) H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})^{1-\beta_2} \mid W_1, \mathcal{G}) \\ & + n \; o\,(\log \bar{P}) \qquad (205) \\ = \; & \hat{N}_1 H((\bar{\mathbf{X}}_{c}^{[n]})_{\beta_1}^{1} \mid \bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})^{1-\beta_2}, W_1, \mathcal{G}) + (\hat{N}_1 + \hat{N}_2) H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})^{1-\beta_2} \mid W_1, \mathcal{G}) \\ & + n \; o\,(\log \bar{P}) \qquad (206) \\ = \; & \hat{N}_1 H(\bar{C}_1^{[n]}, \bar{C}_2^{[n]}, \cdots, \bar{C}_{\hat{N}_1+\hat{N}_2}^{[n]} \mid \bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})^{1-\beta_2}, W_1, \mathcal{G}) \\ & + (\hat{N}_1 + \hat{N}_2) H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})^{1-\beta_2} \mid W_1, \mathcal{G}) + n \; o\,(\log \bar{P}) \qquad (207) \\ \leq \; & \sum_{j=1}^{\hat{N}_1+\hat{N}_2} H(\bar{C}_{i_1}^{[n]}, \bar{C}_{i_2}^{[n]}, \cdots, \bar{C}_{i_{\hat{N}_1}}^{[n]} \mid \bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})^{1-\beta_2}, W_1, \mathcal{G}) \\ & + (\hat{N}_1 + \hat{N}_2) H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})^{1-\beta_2} \mid W_1, \mathcal{G}) + n \; o\,(\log \bar{P}) \qquad (208) \\ = \; & \sum_{j=1}^{\hat{N}_1+\hat{N}_2} H(\bar{C}_{i_1}^{[n]}, \bar{C}_{i_2}^{[n]}, \cdots, \bar{C}_{i_{\hat{N}_1}}^{[n]}, \bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})^{1-\beta_2} \mid W_1, \mathcal{G}) + n \; o\,(\log \bar{P}) \qquad (209) \\ \leq \; & (\hat{N}_1 + \hat{N}_2) H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) + n \; o\,(\log \bar{P}) \qquad (210) \end{aligned}$$

Let us explain how (204) follows from Lemma 3. Set $M_1 = M_2 = 2$, and define $\bar{\mathbf{U}}_1$ and $\bar{\mathbf{U}}_2$ as,

$$\bar{\mathbf{U}}_1 \;=\; \left(\bar{\mathbf{Y}}_{2a}^{[n]}\right) \tag{211}$$

$$\bar{\mathbf{U}}_2 \;=\; \left(\bar{\mathbf{Y}}_{1c}^{[n]}\right) \tag{212}$$

$$W \;=\; \left((\bar{\mathbf{Y}}_{2b}^{[n]})_{\beta_1}^{1}, (\bar{\mathbf{Y}}_{1d}^{[n]})^{1-\beta_2}\right) \tag{213}$$

From (116), (204) is concluded as all the $(\lambda_{1i} - \lambda_{2i})^+$ are zero in the right side of (116). (205) is true from the chain rule, (206) is concluded as $\beta_1 + \beta_2 \geq 1$, (207) is obtained from the definition

of $\bar{C}_i^{[n]}$ in (230) and (208) follows from sub-modularity properties of the entropy function, i.e., for any m random variables $\{X_1, X_2, \cdots, X_m\}$ where we define X_{k+m} as X_k for positive numbers k we have,

$$nH(X_1, \cdots, X_m) \leq \sum_{i=1}^{m} H(X_i, X_{i+1}, \cdots, X_{i+n}) \tag{214}$$

if $n \leq m$. (209) is true from the chain rule. (210) is concluded from Theorem 1. From (193), define the random variables $Z_r(t)$, $Z_{r1}(t)$ and $Z_{r2}(t)$ for all $r \in [\hat{N}_1]$ and $t \in [n]$ as

$$\begin{aligned} Z_r(t) &= \bar{y}_{1rd}(t) && (215)\\ Z_{r1}(t) &= (\bar{y}_{1rd}(t))^1_{\beta_2} \\ &\doteq L^g_{3r}(t)\left((\bar{\mathbf{X}}_a(t))^1_{\beta_2}\right) && (216)\\ Z_{r2}(t) &= \bar{C}_{i_r}(t) && (217)\end{aligned}$$

where λ_{ri} is derived for any $i \in \{1, 2\}$ as

$$\lambda_{ri} = \begin{cases} \beta_2 & i = 1 \\ 1 - \beta_2 & i = 2 \end{cases} \tag{218}$$

D Proof of Lemma 5

Define the set s_β as the set $\{(\beta_1, \beta_2) \in \mathbb{R}^{2+}; \beta_1, \beta_2 \leq 1\}$. We claim that for any positive numbers Q_1 and Q_2 the real-valued function $S(\beta_1, \beta_2)$ defined as,

$$\begin{aligned} & S(\beta_1, \beta_2) \\ \stackrel{\text{def}}{=} & \sum_{r=1}^{Q_1} f_r(\beta_1, \beta_2) \\ & \times H(L^g_{rs}((\mathbf{X}^n_a)^{g_{rsa}(\beta_1,\beta_2)} \bigtriangledown (\mathbf{X}^n_b)^{g_{rsb}(\beta_1,\beta_2)} \bigtriangledown (\mathbf{X}^n_c)^{g_{rsc}(\beta_1,\beta_2)}), \forall s \in [Q_2]) \end{aligned} \tag{219}$$

is a continuous function on the set s_β under the conditions specified in the following lemma .

Lemma 6 *$S(\beta_1, \beta_2)$ is a continuous function on the set s_β if $f_r(\beta_1, \beta_2)$, $g_{rsl}(\beta_1, \beta_2)$ are bounded continuous functions on s_β for any $r \in [Q_1], s \in [Q_2], l \in \{a, b, c, d\}$.*

Proof of Lemma 6 is relegated to Section F. Therefore, it is sufficient to prove Lemma 5 for the non-negative rational numbers $(\beta_1, \beta_2) = (\frac{m}{q}, \frac{e}{q})$ where $m + e \leq q$ [6]. From the definition of $\bar{\mathbf{Y}}_{1c}(t)$, $\bar{\mathbf{Y}}_{1d}(t)$, $\bar{\mathbf{Y}}_{2a}(t)$, $\bar{\mathbf{Y}}_{2c}(t)$ and $\bar{\mathbf{Y}}_{2d}(t)$ in ((191)-(197)) we have,

$$\breve{N}_1 H(\bar{\mathbf{Y}}^{[n]}_{2a}, \bar{\mathbf{Y}}^{[n]}_{2c}, \bar{\mathbf{Y}}^{[n]}_{2d} \mid W_1, \mathcal{G}) - n\breve{N}_0 \log \bar{P}$$

[6]Note that for any real-valued continuous function $f(x_1, x_2, \cdots, x_n)$ on $\mathbb{R}^n$

$$\begin{aligned} \text{If} \quad & f(x_1, x_2, \cdots, x_n) \leq A, \forall (x_1, x_2, \cdots, x_n) \in \mathbb{Q}^n \\ \text{then} \quad & f(x_1, x_2, \cdots, x_n) \leq A, \forall (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n \end{aligned} \tag{220}$$

where (220) follows from continuity of function $f(x_1, x_2, \cdots, x_n)$ on $\mathbb{R}^n$. Therefore, proving that a function $f(x_1, x_2, \cdots, x_n)$ on $\mathbb{R}^n$ is bounded by some number A is equivalent to proving that the function is bounded by A for $(x_1, x_2, \cdots, x_n) \in \mathbb{Q}^n$. Thus, proving (142) for the set $\{(\beta_1, \beta_2) \in \mathbb{R}^{2+}; \beta_1 + \beta_2 \leq 1\}$ is equivalent to proving

$$
\begin{aligned}
&= \breve{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, \bar{\mathbf{Y}}_{2c}^{[n]}, \bar{\mathbf{Y}}_{2d}^{[n]} \mid W_1, \mathcal{G}) - n\breve{N}_1(M-N_1)\beta_1 \log \bar{P} && (222)\\
&\leq \breve{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, \bar{\mathbf{Y}}_{2c}^{[n]}, \bar{\mathbf{Y}}_{2d}^{[n]} \mid W_1, \mathcal{G}) - \breve{N}_1 H((\bar{\mathbf{Y}}_{2c}^{[n]})_{\beta_1}, (\bar{\mathbf{Y}}_{2d}^{[n]})_{\beta_1} \mid W_1, \mathcal{G}) && (223)\\
&\leq \breve{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, (\bar{\mathbf{Y}}_{2c}^{[n]})_{\beta_1}^{1}, (\bar{\mathbf{Y}}_{2d}^{[n]})_{\beta_1}^{1} \mid W_1, \mathcal{G}) && (224)
\end{aligned}
$$

where $\bar{\mathbf{Y}}_{2c}(t)$ and $\bar{\mathbf{Y}}_{2d}(t)$ are defined from the random variable $\bar{\mathbf{Y}}_{2b}(t)$ in (196) and (197) as,

$$
\begin{aligned}
\bar{\mathbf{Y}}_{2c}(t) &= [\bar{\mathbf{Y}}_{2b}(t)]_{0\to(M-N_2)} && (225)\\
\bar{\mathbf{Y}}_{2d}(t) &= [\bar{\mathbf{Y}}_{2b}(t)]_{(M-N_2)\to N_2-N_1} && (226)
\end{aligned}
$$

(224) is true from the chain rule. Similar to (202), from (201) it is sufficient to demonstrate the following inequality,

$$
\begin{aligned}
& \breve{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, (\bar{\mathbf{Y}}_{2c}^{[n]})_{\frac{m}{q}}^{1}, (\bar{\mathbf{Y}}_{2d}^{[n]})_{\frac{m}{q}}^{1} \mid W_1, \mathcal{G}) + \breve{N}_2 H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})_{\frac{e}{q}}^{1} \mid W_1, \mathcal{G})\\
\leq\ & (\breve{N}_1 + \breve{N}_2) H(\bar{\mathbf{Y}}_1^n \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (227)
\end{aligned}
$$

where the numbers $\breve{N}_1$ and $\breve{N}_2$ can be rewritten as,

$$
\begin{aligned}
\breve{N}_1 &= (M-N_2)\frac{e}{q} && (228)\\
\breve{N}_2 &= (N_2-N_1)\frac{q-m}{q} && (229)
\end{aligned}
$$

Before preceding to proof of (227) let us define the random variables $\bar{C}_i^{[n]}$ as the distinct $\frac{1}{q}$ power levels of $(\bar{\mathbf{Y}}_{2c}^{[n]})_{\frac{m}{q}}^{\frac{m+e}{q}}$ and $(\bar{\mathbf{Y}}_{2d}^{[n]})_{\frac{m}{q}}^{1}$, i.e.,

$$
\bar{C}_i(t) = \begin{cases}
(\bar{Y}_{2c}^{\lfloor \frac{i-1}{e} \rfloor + 1}(t))_{\frac{m+i-1-e\lfloor \frac{i-1}{e} \rfloor}{q}}^{\frac{m+i-e\lfloor \frac{i-1}{e} \rfloor}{q}} & 1 \leq i \leq q\breve{N}_1\\
(\bar{Y}_{2d}^{\lfloor \frac{\bar{i}}{q-m} \rfloor + 1}(t))_{\frac{\bar{i}-(q-m)\lfloor \frac{\bar{i}}{q-m} \rfloor}{q}}^{\frac{\bar{i}+1-(q-m)\lfloor \frac{\bar{i}}{q-m} \rfloor}{q}} & q\breve{N}_1 < i \leq q(\breve{N}_1 + \breve{N}_2)\\
\bar{C}_{i-q(\breve{N}_1+\breve{N}_2)}(t) & q(\breve{N}_1 + \breve{N}_2) < i
\end{cases} \qquad (230)
$$

where $\bar{i} = i - e(M-N_2) - 1$. For instance, $C_1(t)$ is defined as $(\bar{Y}_{2c}^{1}(t))_{\frac{m}{q}}^{\frac{m+1}{q}}$, i.e., the bottom $\frac{1}{q}$ power level of $\bar{Y}_{2c}^{1}(t)$. Starting from the left side of (227), we have

$$
\begin{aligned}
& \breve{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, (\bar{\mathbf{Y}}_{2c}^{[n]})_{\frac{m}{q}}^{1}, (\bar{\mathbf{Y}}_{2d}^{[n]})_{\frac{m}{q}}^{1} \mid W_1, \mathcal{G}) + \breve{N}_2 H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})_{\frac{e}{q}}^{1} \mid W_1, \mathcal{G})\\
=\ & \breve{N}_1 H(\bar{\mathbf{Y}}_{2a}^{[n]}, (\bar{\mathbf{Y}}_{2c}^{[n]})_{\frac{m}{q}}^{\frac{m+e}{q}}, (\bar{\mathbf{Y}}_{2c}^{[n]})_{\frac{m+e}{q}}^{1}, (\bar{\mathbf{Y}}_{2d}^{[n]})_{\frac{m}{q}}^{1} \mid W_1, \mathcal{G}) + \breve{N}_2 H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})_{\frac{e}{q}}^{1} \mid W_1, \mathcal{G}) && (231)
\end{aligned}
$$

(142) for the set $\{(\beta_1, \beta_2) \in \mathbb{Q}^{2+}; \beta_1 + \beta_2 \leq 1\}$, i.e.,

$$
\begin{aligned}
& \breve{N}_1 H(\bar{\mathbf{Y}}_2^{[n]} \mid W_1, \mathcal{G})\\
\leq\ & \breve{N}_1 H(\bar{\mathbf{Y}}_1^n \mid W_1, \mathcal{G}) + \breve{N}_2 H((\bar{\mathbf{Y}}_{1b}^n)_{\beta_2} \mid \bar{\mathbf{Y}}_{1a}^n, (\bar{\mathbf{Y}}_{1b}^n)^{1-\beta_2}, W_1, \mathcal{G}) + n\breve{N}_0 \log \bar{P} + n\ o\ (\log \bar{P})\\
& , (\beta_1, \beta_2) \in \mathbb{Q}^{2+}; \beta_1 + \beta_2 \leq 1 && (221)
\end{aligned}
$$

$$
\begin{aligned}
&\leq \breve{N}_1 H(\bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{2c}^{[n]})_{\frac{m}{q}}^{\frac{m+e}{q}}, (\bar{\mathbf{Y}}_{1d}^{[n]})_{\frac{e}{q}}^{1}, (\bar{\mathbf{Y}}_{2d}^{[n]})_{\frac{m}{q}}^{1} \mid W_1, \mathcal{G}) + \breve{N}_2 H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})_{\frac{e}{q}}^{1} \mid W_1, \mathcal{G}) \\
&\quad + n\ o\ (\log \bar{P}) && (232)\\
&= \breve{N}_1 H((\bar{\mathbf{Y}}_{2c}^{[n]})_{\frac{m}{q}}^{\frac{m+e}{q}}, (\bar{\mathbf{Y}}_{2d}^{[n]})_{\frac{m}{q}}^{1} \mid \bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})_{\frac{e}{q}}^{1}, W_1, \mathcal{G}) + (\breve{N}_1 + \breve{N}_2) H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})_{\frac{e}{q}}^{1} \mid W_1, \mathcal{G}) \\
&\quad + n\ o\ (\log \bar{P}) && (233)\\
&= (M - N_2)\frac{e}{q} H(\bar{C}_1^{[n]}, \bar{C}_2^{[n]}, \cdots, \bar{C}_{(M-N_2)e+(N_2-N_1)(q-m)}^{[n]} \mid \bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})_{\frac{e}{q}}^{1}, W_1, \mathcal{G}) \\
&\quad + ((M - N_2)\frac{e}{q} + (N_2 - N_1)\frac{q-m}{q}) H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})_{\frac{e}{q}}^{1} \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (234)\\
&\leq \frac{1}{q} \sum_{j=1}^{(M-N_2)e+(N_2-N_1)(q-m)} H(\bar{C}_{i_1}^{[n]}, \bar{C}_{i_2}^{[n]}, \cdots, \bar{C}_{i_{(M-N_2)e}}^{[n]} \mid \bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})_{\frac{e}{q}}^{1}, W_1, \mathcal{G}) \\
&\quad + ((M - N_2)\frac{e}{q} + (N_2 - N_1)\frac{q-m}{q}) H(\bar{\mathbf{Y}}_{1c}^{n}, (\bar{\mathbf{Y}}_{1d}^{n})_{\frac{e}{q}}^{1} \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (235)\\
&= \frac{1}{q} \sum_{j=1}^{(M-N_2)e+(N_2-N_1)(q-m)} H(\bar{C}_{i_j}^{[n]}, \bar{C}_{i_{j+1}}^{[n]}, \cdots, \bar{C}_{i_{j-1+(M-N_2)e}}^{[n]}, \bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})_{\frac{e}{q}}^{1} \mid W_1, \mathcal{G}) \\
&\quad + n\ o\ (\log \bar{P}) && (236)\\
&\leq \frac{(M - N_2)e + (N_2 - N_1)(q-m)}{q} H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (237)\\
&= (\breve{N}_1 + \breve{N}_2) H(\bar{\mathbf{Y}}_1^{[n]} \mid W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (238)
\end{aligned}
$$

(231) follows from Definition 2. Let us explain how (232) follows from Lemma 3. Set $M_1 = M_2 = 2$, and define $\bar{\mathbf{U}}_1$ and $\bar{\mathbf{U}}_2$ as,

$$
\begin{aligned}
\bar{\mathbf{U}}_1 &= \left(\bar{\mathbf{Y}}_{2a}^{[n]}, (\bar{\mathbf{Y}}_{2c}^{[n]})_{\frac{m+e}{q}}^{1}\right) && (239)\\
\bar{\mathbf{U}}_2 &= \left(\bar{\mathbf{Y}}_{1c}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})_{\frac{e}{q}}^{1}\right) && (240)\\
W &= \left((\bar{\mathbf{Y}}_{2c}^{[n]})_{\frac{m}{q}}^{\frac{m+e}{q}}, (\bar{\mathbf{Y}}_{2d}^{[n]})_{\frac{m}{q}}^{1}\right) && (241)
\end{aligned}
$$

From (116), (232) is concluded as all the $(\lambda_{1i} - \lambda_{2i})^+$ are zero in the right side of (116). (233) is true from the chain rule, (234) is obtained from the definition of $\bar{C}_i^{[n]}$ in (230) and (235) follows from sub-modularity properties of the entropy function, see (214). (236) and (238) are true from the chain rule. Let us clarify how (237) is concluded from Theorem 1 [Theorem 4 in [5]], i.e.,

$$
\begin{aligned}
& H(\bar{C}_{i_j}^{[n]}, \bar{C}_{i_{j+1}}^{[n]}, \cdots, \bar{C}_{i_{j-1+(M-N_2)e}}^{[n]}, (\bar{\mathbf{Y}}_{1d}^{[n]})_{\frac{e}{q}}^{1} \mid \bar{\mathbf{Y}}_{1c}^{[n]}, W_1, \mathcal{G}) && (242)\\
\leq\ & H(\bar{\mathbf{Y}}_{1d}^{[n]} \mid \bar{\mathbf{Y}}_{1c}^{[n]}, W_1, \mathcal{G}) + n\ o\ (\log \bar{P}) && (243)\\
& && (244)
\end{aligned}
$$

From (193), define the random variables $Z_r(t)$, and $Z_{ri}(t)$ for all $r \in [(M - N_2)]$ and $t \in [n]$ as

$$
\begin{aligned}
Z_r(t) &= \bar{y}_{1rd}(t) && (245)\\
Z_{r1}(t) &= (\bar{y}_{1rd}(t))_{\frac{e}{q}}^{1} && (246)
\end{aligned}
$$

$$Z_{r,m+1}(t) \quad = \quad \bar{C}_{i_{j+(r-1)e+m-1}}(t), \forall m \in [e] \tag{247}$$

where λ_{ri} is derived for any $i \in [e+1]$ as

$$\lambda_{ri} \quad = \quad \begin{cases} \frac{1}{q} & 0 < i \leq e \\ \frac{q-e}{q} & i = e+1 \end{cases} \tag{248}$$

Let us prove (237) in detail.

E Proof of (237)

We wish to prove that for the random variables $Z_r(t)$, and $Z_{ri}(t)$ defined in ((245)-(247)),

$$\begin{aligned} & H(Z_{11}^{[n]}, \cdots, Z_{(M-N_2),(e+1)}^{[n]} \mid W) - H(Z_1^{[n]}, \cdots, Z_{(M-N_2)}^{[n]} \mid W, \mathcal{G}) \\ \leq \quad & +n \; o(\log \bar{P}) \end{aligned} \tag{249}$$

E.1 Main idea of the proof

First of all, let us go over a toy example in order to better grasp the main idea of the proof. Assume $n = 1, M - N_2 = 1$. Moreover, define

$$Z \quad = \quad L_1^g(X_1, X_2) \tag{250}$$

$$Z' \quad = \quad L_2^g(X_1, X_2) \tag{251}$$

$$Z_a \quad = \quad (Z)_{0.8}^1 \tag{252}$$

$$Z_b \quad = \quad (Z')_{0.8} \tag{253}$$

First of all, note that we can assume that $Z_a = (Z)_{0.8}^1 \doteq L_1^g((X_1)_{0.8}^1, (X_2)_{0.8}^1)$ which is concluded similar to (70). Let us prove that,

$$H(Z_a, Z_b) - H(Z \mid \mathcal{G}) \quad \leq \quad o(\log \bar{P}) \tag{254}$$

Intuitively, (254) is true as treating Z' the same as Z we have,

$$H((Z)_{0.8}, (Z)_{0.8}^1) - H(Z \mid \mathcal{G}) \quad \leq \quad o(\log \bar{P}) \tag{255}$$

which is true from Definition 2. Let us prove (254) as follows. Define U' and U as (Z_a, Z_b) and Z, respectively. Then, our goal is to prove that

$$H(U') - H(U \mid \mathcal{G}) \quad \leq \quad o\,(\log \bar{P}) \tag{256}$$

Similar to the AIS approach in [5], we can assume that U is a function of $U', \mathcal{G}$. See section E.2 for details. Moreover, we have

$$H(U') - H(U \mid \mathcal{G}) \quad \leq \quad \log \{\mathrm{E}_{\mathcal{G}} \, |\mathcal{S}_\nu(\mathcal{G})|\} \tag{257}$$

where $\mathcal{S}_\nu(\mathcal{G})$ is the set of all U' which result in the same U as ν.

$$\mathrm{E}_{\mathcal{G}}\{|S_\nu(\mathcal{G})|\} = \sum_\lambda \mathbb{P}(\lambda) \tag{258}$$

where $\mathbb{P}(\lambda)$ is defined as the probability that λ and ν correspond to the same U.

E.1.1 Bounding the Probability of Image Alignment

Given $\mathcal{G}$, consider two distinct instances of U' denoted as $\lambda = (\lambda_1, \lambda_2)$ and $\nu = (\nu_1, \nu_2)$ produced by corresponding realizations of codewords (X_1, X_2) denoted by (E_1, E_2) and (F_1, F_2), respectively.

$$\begin{aligned}
\lambda_1 &= L_1^g((E_1)_{0.8}^1, (E_2)_{0.8}^1) && (259)\\
\lambda_2 &= (L_2^g(E_1, E_2))_{0.8} && (260)\\
\nu_1 &= L_1^g((F_1)_{0.8}^1, (F_2)_{0.8}^1) && (261)\\
\nu_2 &= (L_2^g(F_1, F_2))_{0.8} && (262)
\end{aligned}$$

Now, let us bound $\mathbb{P}(\lambda \in \mathcal{S}_\nu)$ from above. We wish to bound the probability that the images of these two codewords align, or in other words $U(\lambda, \mathcal{G}) = U(\nu, \mathcal{G})$. Thus, we have

$$\begin{aligned}
L_1^g(E_1, E_2) &= L_1^g(F_1, F_2) && (263)\\
\rightarrow \lfloor g_1 E_1 \rfloor + \lfloor g_2 E_2 \rfloor &= \lfloor g_1 F_1 \rfloor + \lfloor g_2 F_2 \rfloor && (264)
\end{aligned}$$

For fixed value of g_1 the random variable $g_2(E_2 - F_2)$ must take values within an interval of length no more than 4. Thus, the probability of which is no more than $\frac{4f_{\max}}{|E_2 - F_2|}$ if $E_2 \neq F_2$. The probability of alignment is bounded by $\frac{4f_{\max}}{\max(|E_1 - F_1|, |E_2 - F_2|)}$ if either $E_1 - F_1 \neq 0$ or $E_2 - F_2 \neq 0$.

E.1.2 Bounding the Average Size of Aligned Image Sets

Our goal is to prove that,

$$\mathrm{E}_{\mathcal{G}} |\mathcal{S}_\nu(\mathcal{G})| = \sum_{|\lambda_1 - \nu_1| \in \mathcal{Z}_1, |\lambda_2 - \nu_2| \in \mathcal{Z}_2} \frac{4f_{\max}}{\max(|E_1 - F_1|, |E_2 - F_2|)} \leq c \log \bar{P} \qquad (265)$$

for some positive constant c not depending on P which results in $H(U') - H(U \mid \mathcal{G}) \leq \log(c \log P)$. $\mathcal{Z}_i$ is defined as the support of the random variable $|\lambda_i - \nu_i|$ [7]. Define Δ_j as follows

$$\Delta_j \triangleq (E_j)_{0.8}^1 - (F_j)_{0.8}^1, \forall j \in \{1, 2\} \qquad (266)$$

Consider the following two cases of $\max_{j \in \{1,2\}} |\Delta_j| \geq 2$ and $\max_{j \in \{1,2\}} |\Delta_j| \leq 1$.

1. $\max_{j \in \{1,2\}} |\Delta_j| \geq 2$
   Any number $X \in \mathcal{X}_\delta$ can be written as $(X)_a^\delta \bar{P}^a + X_a$ for any non-negative number a less than δ. Thus, we have

$$\begin{aligned}
& |E_j - F_j| \\
= \; & |\Delta_j \bar{P}^{0.8} + (E_j)_{0.8} - (F_j)_{0.8}| && (267)\\
\geq \; & |\Delta_j \bar{P}^{0.8}| - |(E_j)_{0.8} - (F_j)_{0.8}| && (268)\\
\geq \; & |\Delta_j \bar{P}^{0.8}| - \bar{P}^{0.8} && (269)\\
= \; & (|\Delta_j| - 1)\bar{P}^{0.8} && (270)
\end{aligned}$$

   Therefore, $\mathrm{E}_{\mathcal{G}} |\mathcal{S}_\nu(\mathcal{G})|$ is bounded as follows.

$$\mathrm{E}_{\mathcal{G}} |\mathcal{S}_\nu(\mathcal{G})|$$

[7] From Definition 7, we know that $\mathcal{Z}_1 = \{a : a \in \mathbb{Z}, |a| \leq 2\Delta \bar{P}^{0.2}\}$, $\mathcal{Z}_2 = \{a : a \in \mathbb{Z}, |a| \leq 2\Delta \bar{P}^{0.8}\}$.

$$
\begin{aligned}
&= \sum_{|\lambda_1-\nu_1|\in\mathcal{Z}_1,|\lambda_2-\nu_2|\in\mathcal{Z}_2} \frac{4f_{\max}}{\max(|E_1-F_1|,|E_2-F_2|)} && (271)\\
&\leq \sum_{|\lambda_1-\nu_1|\in\mathcal{Z}_1,|\lambda_2-\nu_2|\in\mathcal{Z}_2,2\leq\max_{j\in\{1,2\}}|\Delta_j|} \frac{4f_{\max}}{\bar{P}^{0.8}\max(|\Delta_1|-1,|\Delta_2|-1)} && (272)\\
&\leq \sum_{|\lambda_1-\nu_1|\in\mathcal{Z}_1,|\lambda_2-\nu_2|\in\mathcal{Z}_2,|\lambda_1-\nu_1|<2\Delta+5} \frac{4f_{\max}}{\bar{P}^{0.8}} && (273)\\
&\quad + \sum_{|\lambda_1-\nu_1|\in\mathcal{Z}_1,|\lambda_2-\nu_2|\in\mathcal{Z}_2,2\Delta+5\leq|\lambda_1-\nu_1|} \frac{4f_{\max}}{\bar{P}^{0.8}(\frac{|\lambda_1-\nu_1|-4}{2\Delta}-1)} && (274)\\
&\leq \frac{8f_{\max}|\mathcal{Z}_2|(2\Delta+5)}{\bar{P}^{0.8}} + \frac{4f_{\max}|\mathcal{Z}_2|}{\bar{P}^{0.8}} \sum_{|\lambda_1-\nu_1|\in\mathcal{Z}_1,2\Delta+5\leq|\lambda_1-\nu_1|} \frac{1}{(|\lambda_1-\nu_1|-2\Delta-5)} && (275)\\
&\leq 8f_{\max}\Delta(2\Delta+5) + 8f_{\max}\Delta\log(2\Delta\bar{P}^{0.2}) && (276)
\end{aligned}
$$

where (274) is true as,

$$
\begin{aligned}
|\lambda_1-\nu_1| &= |L_1^g((E_1)_{0.8}^1,(E_2)_{0.8}^1) - L_1^g((F_1)_{0.8}^1,(F_2)_{0.8}^1)| && (277)\\
&\leq 4+2\Delta\max(|(E_1)_{0.8}^1-(F_1)_{0.8}^1,(E_2)_{0.8}^1-(F_2)_{0.8}^1|) && (278)
\end{aligned}
$$

(275) is concluded as for any summation $\sum_{a\in S_a,b\in S_b} f(a,b)$ and the real-valued function $f(a,b)$ we have,

$$
\sum_{a\in S_a,b\in S_b} f(a,b) \leq |S_a|\sum_{b\in S_b}\max_{a\in S_a} f(a,b) \leq |S_a||S_b|\max_{a\in S_a,b\in S_b} f(a,b) \quad (279)
$$

(276)) is true as the partial sum of harmonic series can be bounded above by logarithmic function, i.e., $\sum_{i=1}^n \frac{1}{i} \leq \log n$.

2. $\max_{j\in\{1,2\}}|\Delta_j| \leq 1$.
   In this case, from (278) the random variable $|\lambda_1-\nu_1|$ can only takes values from the set $\{a : a\in\mathbb{Z}, |a|\leq 4+2\Delta\}$. Therefore, we bound the expected value of $\mathrm{E}_{\mathcal{G}}$ as follows

$$
\begin{aligned}
&\sum_{|\lambda_1-\nu_1|\in\mathcal{Z}_1,|\lambda_2-\nu_2|\in\mathcal{Z}_2} \frac{4f_{\max}}{\max(|E_1-F_1|,|E_2-F_2|)} && (280)\\
\leq &\sum_{|\lambda_2-\nu_2|\in\mathcal{Z}_2} \frac{8f_{\max}(2+\Delta)}{\max(|E_1-F_1|,|E_2-F_2|)} && (281)
\end{aligned}
$$

Let us first compute the term $|\lambda_2-\nu_2|$.

$$
\begin{aligned}
&|\lambda_2-\nu_2|\\
&= |(L_2^g(E_1,E_2))_{0.8} - (L_2^g(F_1,F_2))_{0.8}| && (282)\\
&= |L_2^g(E_1,E_2) - \bar{P}^{0.8}\lfloor\frac{L_2^g(E_1,E_2)}{\bar{P}^{0.8}}\rfloor - L_2^g(F_1,F_2) + \bar{P}^{0.8}\lfloor\frac{L_2^g(F_1,F_2)}{\bar{P}^{0.8}}\rfloor|\\
&= I_1|L_2^g(E_1,E_2)-L_2^g(F_1,F_2)| + I_2\bar{P}^{0.8}\left(\lfloor\frac{L_2^g(E_1,E_2)}{\bar{P}^{0.8}}\rfloor - \lfloor\frac{L_2^g(F_1,F_2)}{\bar{P}^{0.8}}\rfloor\right) && (283)\\
&= I_1|L_2^g(E_1,E_2)-L_2^g(F_1,F_2)| + I_2\bar{P}^{0.8}I_3 && (284)
\end{aligned}
$$

where $I_1, I_2 \in \{-1, 1\}$ and I_3 is an integer-valued random variable taking numbers from the set $\{a : a \in \mathbb{Z}, |a| \leq 4+2\Delta\}$. I_1, I_2 take numbers from the set $\{-1, 1\}$ as for any real numbers a and b, we have $|a+b| = I_1|a| + I_2 b$ for $I_1, I_2 \in \{-1, 1\}$. Moreover, $|I_3| \leq 4 + 2\Delta$ follows similar to (278) and the fact that $\max_{j\in\{1,2\}} |\Delta_j| \leq 1$. Therefore, from (281) we have

$$\begin{aligned}
& \sum_{|\lambda_2-\nu_2|\in\mathcal{Z}_2} \frac{8f_{\max}(2+\Delta)}{\max(|E_1-F_1|, |E_2-F_2|)} \\
\leq & \sum_{|\lambda_2-\nu_2|\in\mathcal{Z}_2} \frac{8f_{\max}(2+\Delta)}{\max\left(\frac{|L_2^g(E_1,E_2)-L_2^g(F_1,F_2)|-4}{2\Delta}, 1\right)} && (285)\\
= & \sum_{|\lambda_2-\nu_2|\in\mathcal{Z}_2} \frac{16\Delta f_{\max}(2+\Delta)}{\max\left(||\lambda_2-\nu_2| - I_2 I_3 \bar{P}^{0.8} - 4I_1|, 1\right)} && (286)\\
\leq & \ 16\Delta f_{\max}(2+\Delta)2\Delta(4\Delta+13)\log(2\Delta\bar{P}^{0.8}) && (287)\\
= & \ O(\log \bar{P}) && (288)
\end{aligned}$$

(285) is concluded similar to (274) as,

$$|L_2^g(E_1,E_2) - L_2^g(F_1,F_2)| \ \leq \ 4 + 2\Delta \max(|E_1-F_1|, |E_2-F_2|) \tag{289}$$

(286) is true from (284) and (287) follows from the following inequality. For any positive integer number a and any integer-valued functions $b(n)$ and $c(n)$ whose absolute values are bounded by M we have

$$\sum_{i=1}^{an} \frac{1}{\max(|i - b(n)n - c(n)|, 1)} \leq a(2M+5)\log((a+M)n+M) \tag{290}$$

(290) is true as we count each number at most $a(2M+5)$ times, e.g., consider the term $\frac{1}{2}$. i can be $b(n)n+c(n)+2$ or $b(n)n+c(n)-2$ for any $b(n) \in \{1, 2, \cdots, a\}, c(n) \in \{-M, \cdots, 0, \cdots, M\}$.

E.2 Aligned Image Sets

Define $\bar{\mathbf{U}}'$ and $\bar{\mathbf{U}}$ as $\left(Z_{11}^{[n]}, \cdots, Z_{(M-N_2),(e+1)}^{[n]}\right)$ and $\left(Z_1^{[n]}, \cdots, Z_{(M-N_2)}^{[n]}\right)$, respectively. Let us prove,

$$H(\bar{\mathbf{U}}' \mid W, \mathcal{G}) - H(\bar{\mathbf{U}} \mid W, \mathcal{G}) \ \leq \ n\ o\ (\log \bar{P}) \tag{291}$$

We are only interested in the difference of entropies of $\bar{\mathbf{U}}'$ and $\bar{\mathbf{U}}$ conditioned on W and $\mathcal{G}$. Similar to the AIS approach in [5], we first claim that from the functional dependence, $\bar{\mathbf{U}}$ can be made a function of $\bar{\mathbf{U}}', W, \mathcal{G}$. Consider some instance of $\bar{\mathbf{U}}'$, e.g., $\nu^{[n]}$. For given W and channel realization $\mathcal{G}$, define aligned image set $\mathcal{S}_{\nu^{[n]}}(W, \mathcal{G})$ as the set of all $\bar{\mathbf{U}}'$ which result in the same $\bar{\mathbf{U}}$ as $\nu^{[n]}$. Since uniform distribution maximizes the entropy,

$$\begin{aligned}
\mathcal{D}_\Delta \ \triangleq \ & H(\bar{\mathbf{U}}' \mid W, \mathcal{G}) - H(\bar{\mathbf{U}} \mid W, \mathcal{G}) \\
\leq \ & H(\bar{\mathbf{U}}' \mid \bar{\mathbf{U}}, W, \mathcal{G}) && (292)\\
\leq \ & \mathrm{E}_{\mathcal{G}}\left\{\log|\mathcal{S}_{\nu^{[n]}}(W, \mathcal{G})|\right\} && (293)\\
= \ & \mathrm{E}_W\left\{\mathrm{E}_{\mathcal{G}}\left\{\log|\mathcal{S}_{\nu^{[n]}}(W, \mathcal{G})| \mid W\right\}\right\} && (294)\\
\leq \ & \max_{w\in\mathcal{W}} \mathrm{E}_{\mathcal{G}}\left\{\log|\mathcal{S}_{\nu^{[n]}}(W, \mathcal{G})| \mid W = w\right\} && (295)
\end{aligned}$$

$$\leq \max_{w\in\mathcal{W}} \log\left\{\mathrm{E}_{\mathcal{G}}|\mathcal{S}_{\nu^{[n]}}(W,\mathcal{G})|| \, W=w\right\} \tag{296}$$

where $\mathcal{W}$ is defined as the support of the random variable W. We are only interested in the difference of entropies of $\bar{\mathbf{U}}'$ and $\bar{\mathbf{U}}$ conditioned on W and $\mathcal{G}$, i.e., $H(\bar{\mathbf{U}}' \mid W,\mathcal{G}) - H(\bar{\mathbf{U}} \mid W,\mathcal{G})$. Similar to the AIS approach in [5], we start with functional dependence. From the functional dependence argument, without loss of generality $\bar{\mathbf{U}}$ can be made a function of $\bar{\mathbf{U}}', W, \mathcal{G}$. So, from (296), $\mathrm{E}_{\mathcal{G}}\{|S_{\nu^n}(W,\mathcal{G})| \mid W=w\}$ is what needed to be calculated. Expected value of size of the cardinality of aligned image set is equal to the summation of probability of alignment over all λ^n, or in the other words,

$$\mathrm{E}_{\mathcal{G}}\{|S_{\nu^n}(W,\mathcal{G})|| \, W=w\} = \sum_{\lambda^n} \mathbb{P}(\lambda^n) \tag{297}$$

where $\mathbb{P}(\lambda^n)$ is defined as the probability that λ^n and ν^n correspond to the same $\bar{\mathbf{U}}$.

E.3 Bounding the Probability of Image Alignment

Given $\mathcal{G}$ and $W=w$, consider two distinct instances of $\bar{\mathbf{U}}'$ denoted as $\lambda^{[n]} = (\lambda_{11}^{[n]}, \cdots, \lambda_{(M-N_2),(e+1)}^{[n]})$ and $\nu^{[n]} = (\nu_{11}^{[n]}, \cdots, \nu_{(M-N_2),(e+1)}^{[n]})$ produced by corresponding realizations of codewords $(X_1^n, X_2^n, \cdots, X_M^n)$ denoted by $(E_1^n, E_2^n, \cdots, E_M^n)$ and $(F_1^n, F_2^n, \cdots, F_M^n)$, respectively. For any $k \in [(M-N_2)]$, $l \in [e+1]$, $t \in [n]$, the random variables $\lambda_{kl}(t)$ and $\nu_{kl}(t)$ are derived as,

$$\lambda_{kl}(t) = \begin{cases} L_{k1}^h(t)\left((\bar{\mathbf{E}}_a(t))_{\frac{e}{q}}^{1} \bigtriangledown (\bar{\mathbf{E}}_c(t))_{\frac{m+e}{q}}^{1}\right) & l=1 \\ \left(L_{kl}^g(t)((\bar{\mathbf{E}}_a(t))_{\frac{e}{q}}^{1} \bigtriangledown \bar{\mathbf{E}}_c(t))\right)_{\frac{a_{kl}-1}{q}}^{\frac{a_{kl}}{q}} & 1 < l \leq e+1 \end{cases} \tag{298}$$

$$\nu_{kl}(t) = \begin{cases} L_{k1}^h(t)\left((\bar{\mathbf{F}}_a(t))_{\frac{e}{q}}^{1} \bigtriangledown (\bar{\mathbf{F}}_c(t))_{\frac{m+e}{q}}^{1}\right) & l=1 \\ \left(L_{kl}^g(t)((\bar{\mathbf{F}}_a(t))_{\frac{e}{q}}^{1} \bigtriangledown \bar{\mathbf{F}}_c(t))\right)_{\frac{a_{kl}-1}{q}}^{\frac{a_{kl}}{q}} & 1 < l \leq e+1 \end{cases} \tag{299}$$

where for any $k \in [(M-N_2)]$ we assume that a_{kl} are arbitrary distinct decreasing numbers belonging to the set $\{m+1, m+2, \cdots, q\}$, i.e., we assume that for any $2 \leq l < l' \leq e+1$, $a_{kl} > a_{kl'}$ and $\{a_{k2}, a_{k3}, \cdots, a_{k,e+1}\} \in \{m+1, m+2, \cdots, q\}$. Without loss of generality, let us assume that $a_{kl} = m+e+2-l$. The random variables $\mathbf{E}(t)$, $\mathbf{E}_a(t)$, $\mathbf{E}_c(t)$, $\mathbf{F}(t)$, $\mathbf{F}_a(t)$ and $\mathbf{F}_c(t)$ are also defined as,

$$\mathbf{E}(t) = \begin{bmatrix} E_1(t) & E_2(t) & \cdots & E_M(t) \end{bmatrix}^T \tag{300}$$

$$\mathbf{E}_a(t) = [\mathbf{E}(t)]_{0\to M-N_2} \tag{301}$$

$$\mathbf{E}_c(t) = [\mathbf{E}(t)]_{N_1\to M-N_1} \tag{302}$$

$$\mathbf{F}(t) = \begin{bmatrix} F_1(t) & F_2(t) & \cdots & F_M(t) \end{bmatrix}^T \tag{303}$$

$$\mathbf{F}_a(t) = [\mathbf{F}(t)]_{0\to M-N_2} \tag{304}$$

$$\mathbf{F}_c(t) = [\mathbf{F}(t)]_{N_1\to M-N_1} \tag{305}$$

Note that for any $m \in [M]$ and $t \in [n]$, $\bar{E}_m(t), \bar{F}_m(t) \in \{0, 1, \cdots, \bar{P}\}$, see (111). In the next step we bound $\mathbb{P}(\lambda^{[n]} \in \mathcal{S}_{\nu^{[n]}})$ from above. We wish to bound the probability that the images of these two

codewords align, or in other words $\bar{\mathbf{U}}(\lambda^n, W, \mathcal{G}) = \bar{\mathbf{U}}(\nu^n, W, \mathcal{G})$. Thus, for any $k \in [K]$ and $t \in [n]$ we have,

$$L_k^g(t)\left(\bar{\mathbf{E}}_a(t) \bigtriangledown (\bar{\mathbf{E}}_c(t))^{\frac{1}{m}}_{\frac{m}{q}}\right) = L_k^g(t)\left(\bar{\mathbf{F}}_a(t) \bigtriangledown (\bar{\mathbf{F}}_c(t))^{\frac{1}{m}}_{\frac{m}{q}}\right) \tag{306}$$

From (306), we have,

$$\begin{aligned} & \sum_{j=1}^{M-N_2} \lfloor g_{kj}(t)E_j(t)\rfloor + \sum_{j=N_1+1}^{M} \lfloor g_{kj}(t)(E_j(t))^{\frac{1}{m}}_{\frac{m}{q}}\rfloor \\ = & \sum_{j=1}^{M-N_2} \lfloor g_{kj}(t)F_j(t)\rfloor + \sum_{j=N_1+1}^{M} \lfloor g_{kj}(t)(F_j(t))^{\frac{1}{m}}_{\frac{m}{q}}\rfloor \qquad (307) \\ \Rightarrow & \left|\sum_{j=1}^{M-N_2} g_{kj}(t)(F_j(t)-E_j(t)) + \sum_{j=N_1+1}^{M} g_{kj}(t)((F_j(t))^{\frac{1}{m}}_{\frac{m}{q}} - (E_j(t))^{\frac{1}{m}}_{\frac{m}{q}})\right| \leq M \qquad (308) \end{aligned}$$

where (308) follows from (307) as for any real number x, $|x - \lfloor x \rfloor| < 1$. For any $j \in s_M \stackrel{\text{def}}{=} \{1, 2, \cdots, (M-N_2), N_1+1, \cdots, M\}$, define $A_j(t)$ as,

$$A_j(t) = \begin{cases} F_j(t) - E_j(t) & 1 \leq j \leq (M-N_2) \\ (F_j(t))^{\frac{1}{m}}_{\frac{m}{q}} - (E_j(t))^{\frac{1}{m}}_{\frac{m}{q}} & N_1 < j \leq M \end{cases} \tag{309}$$

For any $k \in [K], t \in [n], l \in \{N\}$ and any fixed values of $g_{k1}(t), \cdots, g_{k(l-1)}(t), g_{k(l+1)}(t), \cdots, g_{kM}(t)$ the random variable $g_{kl}(t)A_l(t)$ must take values within an interval of length no more than $2M$. Therefore, for any $k \in [K], t \in [n], l \in s_M$ if $A_l(t) \neq 0$, then $g_{kl}(t)$ must take values in an interval of length no more than $\frac{2M}{|A_l(t)|}$, the probability of which is no more than $\frac{2Mf_{\max}}{|A_l(t)|}$. The probability of alignment is bounded by

$$\begin{aligned} \mathbb{P}(\lambda^{[n]} \in \mathcal{S}_{\nu^{[n]}}) &\leq \left\{\prod_{t=1, \max_{j\in s_M}|A_j(t)|\neq 0}^{n} \frac{2Mf_{\max}}{\max_{j\in s_M}|A_j(t)|}\right\}^K \qquad (310) \\ &= \prod_{k=1}^{K} \prod_{t=1, \max_{j\in s_M}|A_j(t)|\neq 0}^{n} \frac{2Nf_{\max}}{\max_{j\in s_M}|A_j(t)|} \qquad (311) \end{aligned}$$

E.4 Bounding the Average Size of Aligned Image Sets

Let us assume $n = 1$, $K = 1$ as the generalization to arbitrary n and K follows similar to proof of Theorem 4 in [5]. Without loss of generality let us drop the time index (t). Thus, our goal is to prove that,

$$\begin{aligned} & \sum_{|\lambda_{11}-\nu_{11}|\in\mathcal{Z}_{11},\cdots,|\lambda_{1,e+1}-\nu_{1,e+1}|\in\mathcal{Z}_{1,e+1}, \max_{j\in s_M}|A_j|\neq 0} \frac{2Mf_{\max}}{\max_{j\in s_M}|A_j|} \\ & \leq c_1 + c_2 \log \bar{P} \end{aligned} \tag{312}$$

where A_j is defined from $A_j(t)$ in (309) by dropping the time index (t). Note that, (249) is concluded from (296) and (312) as $\log\log\bar{P} = o\,(\log\bar{P})$. $\mathcal{Z}_{1l}$ is defined as the set $\{0\}\cup[M+1+\lfloor M\Delta\bar{P}^{1}_{\frac{e}{q}}\rfloor]$ for $l=1$ and the set $\{0\}\cup[\bar{P}^{\frac{1}{q}}]$ for $l\in[e+1], l\neq 1$. We define $\breve{\Delta}_l$ as,

$$\breve{\Delta}_j \triangleq \begin{cases} (\bar{E}_j)^{1}_{\frac{e}{q}} - (\bar{F}_j)^{1}_{\frac{e}{q}} & 1\leq j\leq M-N_2 \\ (\bar{E}_j)^{1}_{\frac{m+e}{q}} - (\bar{F}_j)^{1}_{\frac{m+e}{q}} & N_1 < j\leq M \end{cases} \tag{313}$$

where (313) is derived from (298) and (299) dropping the time index (t). Consider the following two cases of $\max_{j\in s_M}|\breve{\Delta}_j|\geq 2$ and $\max_{j\in s_M}\breve{|}\Delta_j|\leq 1$.

1. $\max_{j\in s_M}|\breve{\Delta}_j|\geq 2$
   Any number $X\in\mathcal{X}_\delta$ can be written as $(X)^{\delta}_{a}\bar{P}^{a}+X_a$ for any non-negative number a less than δ. Thus, when $1\leq j\leq(M-N_2)$, the term $|E_j-F_j|$ is bounded from below as,

$$\begin{aligned} |A_j| &= |E_j-F_j| \\ &= |\breve{\Delta}_j\bar{P}^{\frac{e}{q}} + (E_j)_{\frac{e}{q}} - (F_j)_{\frac{e}{q}}| && (314)\\ &\geq |\breve{\Delta}_j\bar{P}^{\frac{e}{q}}| - |(E_j)_{\frac{e}{q}} - (F_j)_{\frac{e}{q}}| && (315)\\ &\geq |\breve{\Delta}_j\bar{P}^{\frac{e}{q}}| - \bar{P}^{\frac{e}{q}} && (316)\\ &= (|\breve{\Delta}_j|-1)\bar{P}^{\frac{e}{q}} && (317)\end{aligned}$$

   and if $N_1<j\leq M$, the term $|(E_j)^{1}_{\frac{m}{q}} - (F_j)^{1}_{\frac{m}{q}}|$ is bounded from below as,

$$\begin{aligned} |A_j| &= |(E_j)^{1}_{\frac{m}{q}} - (F_j)^{1}_{\frac{m}{q}}| \\ &= |\breve{\Delta}_j\bar{P}^{\frac{e}{q}} + (E_j)_{\frac{e}{q}} - (F_j)_{\frac{e}{q}}| && (318)\\ &\geq |\breve{\Delta}_j\bar{P}^{\frac{e}{q}}| - |(E_j)_{\frac{e}{q}} - (F_j)_{\frac{e}{q}}| && (319)\\ &\geq (|\breve{\Delta}_j|-1)\bar{P}^{\frac{e}{q}} && (320)\end{aligned}$$

   Moreover, from (298), (299), and (313), the term $|\lambda_{11}-\nu_{11}|$ is bounded from above as follows,

$$|\lambda_{11}-\nu_{11}| < M + M\Delta\max_{j\in s_M}|\breve{\Delta}_j| \tag{321}$$

   The left side of (312) is bounded as,

$$\begin{aligned} &\sum_{|\lambda_{11}-\nu_{11}|\in\mathcal{Z}_{11},\cdots,|\lambda_{1,e+1}-\nu_{1,e+1}|\in\mathcal{Z}_{1,e+1},\max_{j\in s_M}|A_j|\neq 0}\frac{2Mf_{\max}}{\max_{j\in s_M}|A_j|} \\ \leq &\sum_{|\lambda_{11}-\nu_{11}|\in\mathcal{Z}_{11},\cdots,|\lambda_{1,e+1}-\nu_{1,e+1}|\in\mathcal{Z}_{1,e+1}}\frac{2Mf_{\max}}{\max_{j\in s_M}(|\breve{\Delta}_j|-1)\bar{P}^{\frac{e}{q}}} \\ \leq &\sum_{|\lambda_{11}-\nu_{11}|\in\{0\}\cup[M+M\Delta],|\lambda_{12}-\nu_{12}|\in\mathcal{Z}_{12},\cdots,|\lambda_{1,e+1}-\nu_{1,e+1}|\in\mathcal{Z}_{1,e+1}}\frac{2Mf_{\max}}{\bar{P}^{\frac{e}{q}}} \\ &+\sum_{|\lambda_{11}-\nu_{11}|\in\mathcal{Z}^{\star}_{11},|\lambda_{12}-\nu_{12}|\in\mathcal{Z}_{12},\cdots,|\lambda_{1,e+1}-\nu_{1,e+1}|\in\mathcal{Z}_{1,e+1}}\frac{2M^2\Delta f_{\max}}{(|\lambda_{11}-\nu_{11}|-M-M\Delta)\bar{P}^{\frac{e}{q}}} \end{aligned} \tag{322}$$

$$\tag{323}$$

$$\begin{aligned}
&\leq \frac{2Mf_{\max}(1+M+M\Delta)\prod_{j=2}^{e+1}|\mathcal{Z}_{1j}|}{\bar{P}^{\frac{e}{q}}} \\
&\quad + \frac{2M^2\Delta f_{\max}\prod_{j=2}^{e+1}|\mathcal{Z}_{1j}|}{\bar{P}^{\frac{e}{q}}} \sum_{|\lambda_{11}-\nu_{11}|\in\mathcal{Z}_{11}^{\star}} \frac{1}{(|\lambda_{11}-\nu_{11}|-M-M\Delta)} && (324) \\
&\leq 2^{e+1}M^2 f_{\max}(1+\Delta)(2+\log{(2+M+M\Delta\bar{P}^{1}_{\frac{e}{q}})}) && (325)
\end{aligned}$$

where $\mathcal{Z}_{11}^{\star}$ is defined as the set of $\bar{\mathcal{Z}}_{11}\cap[M+M\Delta]^c$. (322) follows from (317) and (320). Note that $\max_{j\in s_M}|A_j|$ is positive number as $\max_{j\in s_M}|\breve{\Delta}_j|\geq 2$. (323) is obtained from (321) and (324) is true as for any summation $\sum_{a\in S_a,b\in S_b} f(a,b)$ and the real-valued function $f(a,b)$ we have,

$$\sum_{a\in S_a,b\in S_b} f(a,b) \leq |S_a|\sum_{b\in S_b}\max_{a\in S_a} f(a,b) \leq |S_a||S_b|\max_{a\in S_a,b\in S_b} f(a,b) \tag{326}$$

Finally, (325) is concluded as the partial sum of harmonic series can be bounded above by logarithmic function i.e., $\sum_{i=1}^{n}\frac{1}{n}\leq 1+\ln n$.

2. $\max_{j\in s_M}|\breve{\Delta}_j|\leq 1$

   First of all note that from (321), the term $|\lambda_{11}-\nu_{11}|$ only gets values from the set $\{0\}\cup[M+M\Delta]$. Let us define $\hat{\Delta}_{lj}$ as,

   $$\hat{\Delta}_{lj} \triangleq \begin{cases} (\bar{E}_j)^{1}_{\frac{e-l+1}{q}} - (\bar{F}_j)^{1}_{\frac{e-l+1}{q}} & 1\leq j\leq M-N_2 \\ (\bar{E}_j)^{1}_{\frac{e+m-l+1}{q}} - (\bar{F}_j)^{1}_{\frac{e+m-l+1}{q}} & N_1<j\leq M \end{cases} \tag{327}$$

   and define the number $l^{\star}$ as the smallest integer from the set $\{2,3,\cdots,e+1\}$ where

   $$\max_{j\in\{N_1+1,\cdots,M\}}|\hat{\Delta}_{l^{\star}j}|\geq 2 \tag{328}$$

   Consider the following two cases.

   (a) $l^{\star}$ doesn't exist.

   If there doesn't exist any $l\in\{2,\cdots,e+1\}$ and $j\in\{\max(N_1,M-N_2)+1,\cdots,M\}$ satisfying the condition (328), i.e., $\forall l,j,l\in\{2,\cdots,e+1\},j\in\{\max(N_1,M-N_2)+1,\cdots,M\},\hat{\Delta}_{lj}\in\{-1,0,1\}$, then we have,

   $$\begin{aligned}
   &|L^{g}_{kl}(t)((\bar{\mathbf{E}}_a(t))^{1}_{\frac{e}{q}}\bigtriangledown\bar{\mathbf{E}}_c(t)) - L^{g}_{kl}(t)((\bar{\mathbf{F}}_a(t))^{1}_{\frac{e}{q}}\bigtriangledown\bar{\mathbf{F}}_c(t))| \\
   &\leq 2M\Delta\bar{P}^{\frac{e+m-l+1}{q}} && (329) \\
   &\leq 2M\Delta\bar{P}^{\frac{a_{1l}-1}{q}} && (330)
   \end{aligned}$$

   Thus, each of the variables $|\lambda_{1l}-\nu_{1l}|$ is bounded by $2M\Delta+1$ for any $l\in\{2,\cdots,e+1\}$. Therefore, (312) is true as the summation in (312) is the summation of positive numbers less than $2Mf_{\max}$ over at most $(2M\Delta+1)^{e+1}$ numbers, see (326).

(b) $2 \leq l^\star \leq e+1$.
Similar to (317) and (320), when $1 \leq j \leq (M-N_2)$, the term $|E_j - F_j|$ is bounded from below as,

$$\begin{aligned} |A_j| &= |E_j - F_j| \\ &= |\hat{\Delta}_{l^\star j}\bar{P}^{\frac{e-l^\star+1}{q}} + (E_j)_{\frac{e-l^\star+1}{q}} - (F_j)_{\frac{e-l^\star+1}{q}}| \quad (331) \\ &\geq (|\hat{\Delta}_{l^\star j}| - 1)\bar{P}^{\frac{e-l^\star+1}{q}} \quad (332) \end{aligned}$$

and if $N_1 < j \leq M$, the term $|(E_j)^1_{\frac{m}{q}} - (F_j)^1_{\frac{m}{q}}|$ is bounded from below as,

$$\begin{aligned} |A_j| &= |(E_j)^1_{\frac{m}{q}} - (F_j)^1_{\frac{m}{q}}| \\ &= |\hat{\Delta}_{l^\star j}\bar{P}^{\frac{e-l^\star+1}{q}} + (E_j)_{\frac{e-l^\star+1}{q}} - (F_j)_{\frac{e-l^\star+1}{q}}| \quad (333) \\ &\geq (|\hat{\Delta}_{l^\star j}| - 1)\bar{P}^{\frac{e-l^\star+1}{q}} \quad (334) \end{aligned}$$

The left side of (312) is bounded as,

$$\begin{aligned} & \sum_{|\lambda_{11}-\nu_{11}|\in\{0\}\cup[M+M\Delta],|\lambda_{12}-\nu_{12}|\in\mathcal{Z}_{12},\cdots,|\lambda_{1,e+1}-\nu_{1,e+1}|\in\mathcal{Z}_{1,e+1},\max_{j\in s_M}|A_j|\neq 0} \\ & \frac{2Mf_{\max}}{\max_{j\in s_M}|A_j|} \\ \leq\ & (M+1+M\Delta)(2M\Delta+1)^{l^\star-2}\frac{(1+\bar{P}^{\frac{1}{q}})^{e-l^\star+1}}{\bar{P}^{\frac{e-l^\star+1}{q}}} \\ & \times \sum_{|\lambda_{1l^\star}-\nu_{1l^\star}|\in\mathcal{Z}_{1l^\star}} \frac{2Mf_{\max}}{\max_{j\in s_M}(|\hat{\Delta}_{l^\star j}|-1)} \quad (335) \\ \leq\ & (M+1+M\Delta)(2M\Delta+1)^{l^\star-2}\frac{(1+\bar{P}^{\frac{1}{q}})^{e-l^\star+1}}{\bar{P}^{\frac{e-l^\star+1}{q}}} \\ & \times \left(2M^2(1+\Delta)f_{\max} + \sum_{|\lambda_{1l^\star}-\nu_{1l^\star}|\in\bar{\mathcal{Z}}_{1l^\star}} \frac{2M^2\Delta f_{\max}}{|\lambda_{1l^\star}-\nu_{1l^\star}|-M-M\Delta}\right) \quad (336) \\ \leq\ & 2^{e+1}(M+1+M\Delta)(2M\Delta+1)^{l^\star-2}M^2 f_{\max}(1+\Delta)(2+\log\bar{P}^{\frac{1}{q}}) \quad (337) \\ \leq\ & c_1 + c_2\log\bar{P} \quad (338) \end{aligned}$$

for some constants c_1 and c_2 not depending on $\bar{P}$. Note that, $\bar{\mathcal{Z}}_{1l^\star}$ is defined as the set $\mathcal{Z}_{1l^\star} \cap [M+M\Delta]^c$ and (336) follows similar to (323).

F Proof of Lemma 6

As the functions $g_{rsl}(\beta_1,\beta_2)$ are bounded continuous functions on s_β for any $r \in [Q_1], s \in [Q_2], l \in \{a,b,c,d\}$ for any positive real number ϵ there exists some δ that if $|\beta_1 - \beta_1'| < \delta$,$|\beta_2 - \beta_2'| < \delta$ then

$$|g_{rsl}(\beta_1,\beta_2) - g_{rsl}(\beta_1',\beta_2')| \leq \epsilon, \forall r \in [Q_1], s \in [Q_2], l \in \{a,b,c,d\} \quad (339)$$

Therefore, from Lemma 3 we have,

$$\begin{aligned} & | \quad H(L^g_{rs}((\mathbf{X}^n_a)^{g_{rsa}(\beta_1,\beta_2)} \bigtriangledown (\mathbf{X}^n_b)^{g_{rsb}(\beta_1,\beta_2)} \bigtriangledown (\mathbf{X}^n_c)^{g_{rsc}(\beta_1,\beta_2)}), \forall s \in [Q_2]) \\ & \quad -H(L^g_{rs}((\mathbf{X}^n_a)^{g_{rsa}(\beta'_1,\beta'_2)} \bigtriangledown (\mathbf{X}^n_b)^{g_{rsb}(\beta'_1,\beta'_2)} \bigtriangledown (\mathbf{X}^n_c)^{g_{rsc}(\beta'_1,\beta'_2)}), \forall s \in [Q_2])| \\ \leq & \quad nMs\epsilon \log \bar{P} + n\ o\ (\log \bar{P}) \end{aligned} \tag{340}$$

Now as the sum and multiplications of bounded continuous functions are continuous, $S(\beta_1, \beta_2)$ is continuous function on s_β.

References

[1] V. Cadambe and S. Jafar, "Interference Alignment and the Degrees of Freedom of the K user Interference Channel," *IEEE Transactions on Information Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[2] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[3] A. Lapidoth, S. Shamai, and M. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," in *Proceedings of 43rd Annual Allerton Conference on Communications, Control and Computing*, Sep. 28-30, 2005.

[4] A. G. Davoodi and S. A. Jafar, "Aligned image sets under channel uncertainty: Settling conjectures on the collapse of degrees of freedom under finite precision CSIT," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5603–5618, 2016.

[5] ——, "Sum-set inequalities from aligned image sets: Instruments for robust GDoF bounds," *arXiv preprint arXiv:1703.01168*, 2017.

[6] C. Hao, B. Rassouli, and B. Clerckx, "Achievable DoF regions of MIMO networks with imperfect CSIT," *IEEE Transactions on Information Theory*, vol. 63, no. 10, pp. 6587–6606, 2017.

[7] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Transactions on Information Theory*, vol. 52, pp. 3936–3964, Sep. 2006.

[8] S. Jafar and A. Goldsmith, "Isotropic fading vector broadcast channels: the scalar upperbound and loss in degrees of freedom," *IEEE Transactions on Information Theory*, vol. 51, no. 3, pp. 848–857, March 2005.

[9] C. Huang, S. A. Jafar, S. Shamai, and S. Vishwanath, "On Degrees of Freedom Region of MIMO Networks without Channel State Information at Transmitters," *IEEE Transactions on Information Theory*, no. 2, pp. 849–857, Feb. 2012.

[10] C. S. Vaze and M. K. Varanasi, "The degree-of-freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5354–5374, 2012.

[11] A. G. Davoodi and S. A. Jafar, "Transmitter Cooperation under Finite Precision CSIT:A GDoF Perspective," *IEEE Transactions on Information Theory*, 2016.

[12] A. G. Davoodi, and S. A. Jafar, "GDoF of the MISO BC: Bridging the gap between finite precision and perfect CSIT," *IEEE Transactions on Information Theory*, pp. 1297–1301, 2016.

[13] W. Yao and M. K. Varanasi, "Degrees of freedom of the two-user MIMO broadcast channel with private and common messages under hybrid CSIT models," *IEEE Transactions on Information Theory*, May 2017.

[14] ——, "k-user symmetric $M \times N$ MIMO interference channel under finite precision CSIT: A GDoF perspective," *IEEE Transactions on Information Theory*, July 2018.