Title
Degrees of Freedom Region of the (M, N-1, N-2) MIMO Broadcast Channel with Partial CSIT: An Application of Sum-set Inequalities

Permalink
https://escholarship.org/uc/item/94n8s3sd

Journal
2019 IEEE INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY (ISIT), 2019-July

ISSN
2157-8095

Authors
Davoodi, Arash Gholami
Jafar, Syed A
IEEE

Publication Date
2019

DOI
10.1109/ISIT.2019.8849793

License
https://creativecommons.org/licenses/by/4.0/ 4.0

Peer reviewed
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 \(\beta_1, \beta_2\), 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 in [6] for certain parameter regimes (mainly $N_1 = N_2 = 1$), 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 bounds, as well as AIS arguments. However, the general DoF region characterization remains open. 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 $x \geq 0$, 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 $X_{r,s}$ if there is no cause of ambiguity.

For any vector $V = [v_1 \cdots v_k]^T$ and non-negative integer numbers $m$ and $n$ less than $k$, let us define the notation $V_{m \to n}$ as follows,

$$ V_{m \to n} \doteq \begin{cases} \begin{bmatrix} v_{m+1} & \cdots & v_{m+n} \end{bmatrix}^T, & m + n \leq k \\ \begin{bmatrix} v_{m+1} & \cdots & v_k \\ v_1 & \cdots & v_{m+n-k} \end{bmatrix}^T, & k < m + n \end{cases} $$ (1)
For any two vectors $V = [v_1 \cdots v_k]^T$ and $W = [w_1 \cdots w_k]^T$ define their concatenation as

$$V \triangledown W \triangleq [v_1 \cdots v_k \ v_1 w_1 \cdots w_k]^T$$

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

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

The following definitions, inherited from [3], are replicated here for the sake of completeness.

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

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

where $\tilde{P}^\lambda_i$ is a compact notation for $\lfloor \sqrt{P^\lambda_i} \rfloor$. 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 $X$, $\lambda_1$ and $\lambda_2$, define $(X)_{\lambda_1}$ and $(X)^{\lambda_2}_{\lambda_1}$ as,

$$\begin{align*}
(X)_{\lambda_1} & \triangleq X - \tilde{P}^{\lambda_1} \begin{bmatrix} X \\ \tilde{P}^{\lambda_1} \end{bmatrix} \\
(X)^{\lambda_2}_{\lambda_1} & \triangleq \begin{bmatrix} X - \tilde{P}^{\lambda_2} \\ \tilde{P}^{\lambda_2} \end{bmatrix} \begin{bmatrix} X \\ \tilde{P}^{\lambda_2} \end{bmatrix}^{-1}
\end{align*}$$

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

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

$$\begin{align*}
(X)^{\lambda} & \triangleq (X)^{1-\lambda}_{1-\lambda}.
\end{align*}$$

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

$$\begin{align*}
(V)_{\lambda_1} & \triangleq [(v_1)_{\lambda_1} \ (v_2)_{\lambda_1} \ \cdots \ (v_k)_{\lambda_1}]^T \\
(V)^{\lambda_2}_{\lambda_1} & \triangleq [(v_1)^{\lambda_2}_{\lambda_1} \ (v_2)^{\lambda_2}_{\lambda_1} \ \cdots \ (v_k)^{\lambda_2}_{\lambda_1}]^T
\end{align*}$$
Figure 1: Conceptual depiction of an arbitrary variable \(X \in \mathcal{X}_{\lambda_1 + \lambda_2 + \lambda_3}\), and its power-level partitions \((X)_{\lambda_1}\), \((X)_{\lambda_1 + \lambda_2}\) and \((X)_{\lambda_1 + \lambda_2 + \lambda_3}\).

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

1. The magnitudes of all the random variables in \(G\) are bounded away from infinity, i.e., there exists a constant \(\Delta < \infty\) such that for all \(g \in G\) we have \(|g| \leq \Delta\).

2. There exists a finite positive constant \(f_{\max}\), such that for all finite cardinality disjoint subsets \(G_1, G_2\) of \(G\), the joint probability density function of all random variables in \(G_1\), conditioned on all random variables in \(G_2\), exists and is bounded above by \(f_{\max} |G_1|\).

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

**Definition 5 (Arbitrary Channel Set \(H\))** Let \(H\) be a set of real-valued constants with magnitudes bounded away from infinity, i.e., for all \(h \in H\) 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 \(L^g_j(x_i, 1 \leq i \leq k), L^h_j(x_i, 1 \leq i \leq k), L^{g\vec{\gamma}\vec{\delta}}_j(x_i, 1 \leq i \leq k)\) and \(L^{h\vec{\gamma}\vec{\delta}}_j(x_i, 1 \leq i \leq k)\) to represent,

\[
L^g_j(x_1, x_2, \cdots, x_k) = \sum_{1 \leq i \leq k} [g_j, x_i] 
\]

\[
L^h_j(x_1, x_2, \cdots, x_k) = \sum_{1 \leq i \leq k} [h_j, x_i] 
\]

\[
L^{g\vec{\gamma}\vec{\delta}}_j(x_1, x_2, \cdots, x_k) = \sum_{1 \leq i \leq k} [g_j(x_i), \gamma_i]_{\vec{\delta}_i} 
\]

\[
L^{h\vec{\gamma}\vec{\delta}}_j(x_1, x_2, \cdots, x_k) = \sum_{1 \leq i \leq k} [h_j(x_i), \gamma_i]_{\vec{\delta}_i} 
\]
for distinct random variables $g_{ij} \in \mathcal{G}$, arbitrary real valued constants $h_{ji} \in \mathcal{H}$, and arbitrary real valued constants $0 \leq \delta_i \leq \gamma_i \leq \eta_i$. For the vector $V = [v_1 \ v_2 \ \cdots \ v_k]^T$ we similarly define the notations $L_j^g(V)$ and $L_j^h(V)$ to represent,

$$L_j^g(V) = \sum_{1 \leq i \leq k} |g_{ij}v_j|$$

$$L_j^h(V) = \sum_{1 \leq i \leq k} |h_{ij}v_j|$$

(13)

(14)

Noting that these functions are \textit{approximately} (because of the \lceil \cdot \rceil operations) linear, for simplicity we refer to them as linear combinations. In particular, we will refer to $L^g$ functions as \textit{random} linear combinations and to $L^h$ functions as \textit{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.

\textbf{Definition 7} For the linear combinations

$$A = L_i^g(x_i, 1 \leq i \leq k) \quad \text{and} \quad B = L_i^h(x_i, 1 \leq i \leq k)$$

we define $\mathcal{T}(A)$, $\mathcal{T}(B)$, $\mathcal{T}((A)_\lambda^\mu)$ and $\mathcal{T}((B)_\mu)$ as,

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

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

(15)

(16)

(17)

Note that the terminology from Definition [9] is invoked in Definition [7]. Figure 2 provides a visual illustration of $L_i^g$ and $\mathcal{T}(A)$. From the definition of $\mathcal{T}(A)$ and $\mathcal{T}(B)$ in (16), it follows that,

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

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

(18)

(19)

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

$$\mathcal{T}((B)_\mu) \in \{ a : a \in \mathbb{Z}, |a| \leq \bar{\Delta} \mathcal{T}((B)_\mu) \}$$

(20)

(21)

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

\section{3 Sum-set Inequalities: Previous Results in [5]}

Let us recall Theorem 4 of [5].

\footnote{Consider the terms $A = L_i^g(x_i, 1 \leq i \leq k)$ and $(B)_\mu = (L_i^g(x_i, 1 \leq i \leq k))_\mu^\lambda$ and let us bound them as follows.

$$|A| = \sum_{1 \leq i \leq k} |g_{i}(x_i)| \leq k\Delta \max_{1 \leq i \leq k} \gamma_i \bar{\Delta} \leq k\Delta \bar{\Delta} \mathcal{T}(A) \leq k\Delta \bar{\Delta} \mathcal{T}(B) \leq k\Delta \bar{\Delta} \mathcal{T}((B)_\mu)$$

(22)

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

(23)}
Then for any acceptable random variable $W$, $\forall x_1 \in \mathcal{X}_{\eta_1}$ 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 $(\gamma_i, \delta_i)$ are only used to further trim the size of $x_i$, yielding $(x_i)_{\delta_i}^\gamma$ as the trimmed versions. These trimmed variables are then combined with arbitrary coefficients to produce $A = L^\gamma \delta$. Finally, note that $T(A)$ represents the size (power level) of the largest trimmed variable involved in $L^\gamma \delta$.

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

$$Z_k(t) = L_1^\gamma(t)(X_1(t), X_2(t), \ldots, X_N(t))$$

$$Z_k(t) = L_1^{\gamma_1, \delta_1}(t)\left(\sum_{i=1}^{k} \lambda_{kr}, i \in I_k, j \in [N]\right)$$

$$Z_k(t) = L_1^{\gamma_2, \delta_2}(t)\left(\sum_{i=1}^{k} \lambda_{kr}, i \in I_2, j \in [N]\right)$$

$$\vdots$$

$$Z_k(t) = L_1^{\gamma_k, \delta_k}(t)\left(\sum_{i=1}^{k} \lambda_{kr}, i \in I_k, j \in [N]\right)$$

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 $s \in \{1, 2, \ldots, l_k - 1\}$,

$$T(Z_{k,s+1}) + T(Z_{k,s+2}) + \ldots + T(Z_{k,l_k}) \leq \lambda_{k,1} + \lambda_{k,2} + \ldots + \lambda_{k,m(k,s-1)}$$

then for any acceptable random variable $W$, $\forall x_1 \in \mathcal{X}_{\eta_1}$ 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 $(\gamma_i, \delta_i)$ are only used to further trim the size of $x_i$, yielding $(x_i)_{\delta_i}^\gamma$ as the trimmed versions. These trimmed variables are then combined with arbitrary coefficients to produce $A = L^\gamma \delta$. Finally, note that $T(A)$ represents the size (power level) of the largest trimmed variable involved in $L^\gamma \delta$.

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

$$Z_k(t) = L_1^\gamma(t)(X_1(t), X_2(t), \ldots, X_N(t))$$

$$Z_k(t) = L_1^{\gamma_1, \delta_1}(t)\left(\sum_{i=1}^{k} \lambda_{kr}, i \in I_k, j \in [N]\right)$$

$$Z_k(t) = L_1^{\gamma_2, \delta_2}(t)\left(\sum_{i=1}^{k} \lambda_{kr}, i \in I_2, j \in [N]\right)$$

$$\vdots$$

$$Z_k(t) = L_1^{\gamma_k, \delta_k}(t)\left(\sum_{i=1}^{k} \lambda_{kr}, i \in I_k, j \in [N]\right)$$

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 $s \in \{1, 2, \ldots, l_k - 1\}$,

$$T(Z_{k,s+1}) + T(Z_{k,s+2}) + \ldots + T(Z_{k,l_k}) \leq \lambda_{k,1} + \lambda_{k,2} + \ldots + \lambda_{k,m(k,s-1)}$$

then for any acceptable random variable $W$,

$$H(Z_1^{[\eta_1]}, \ldots, Z_K^{[\eta_K]} | W, \mathcal{G}) \geq H(Z_1^{[\eta_1]}, \ldots, Z_K^{[\eta_K]} | W) + Kn o(\log P)$$

Let $\mathcal{G}(Z) \subset \mathcal{G}$ denote the set of all bounded density channel coefficients that appear in $Z_1^{[\eta_1]}, \ldots, Z_K^{[\eta_K]}$, and let $W$ be a random variable such that conditioned on any $\mathcal{G} \subset (\mathcal{G}/\mathcal{G}(Z)) \cup \{W\}$, 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, \(X_1(t), X_2(t), X_3(t) \in X_{\text{max}} = \{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4\}, \ (M = 4)\), 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

$$N_1 \leq N_2 \leq M \leq N_1 + N_2,$$

(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 inner bound 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 $X$ is partitioned as follows.

$$X_a(t) = [X(t)]_{0\rightarrow M-N_2}$$

(31)

$$X_b(t) = [X(t)]_{(M-N_2)\rightarrow (N_1+N_2-M)}$$

(32)

$$X_c(t) = [X(t)]_{N_1\rightarrow (M-N_1)}$$

(33)

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

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].
The dimensions of these symbols are listed as follows.

\[
\begin{align*}
Y_1(t) &= \sqrt{P} G_{1ab}(t) \left[ X_a(t) \bigtriangledown X_b(t) \right] + \sqrt{P^{1-\beta_1}} G_{1c}(t) X_c(t) + \Gamma_1(t) \\
Y_2(t) &= \sqrt{P} G_{2bc}(t) \left[ X_b(t) \bigtriangledown X_c(t) \right] + \sqrt{P^{1-\beta_2}} G_{2a}(t) X_a(t) + \Gamma_2(t)
\end{align*}
\]

(34) (35)

The dimensions of these symbols are as follows.

\[
\begin{align*}
Y_1(t), \Gamma_1(t) : N_1 \times 1, & \quad Y_2(t), \Gamma_2(t) : N_2 \times 1, \\
G_{1ab}(t) : N_1 \times N_1, & \quad G_{1c}(t) : N_1 \times (M - N_1), \\
G_{2bc}(t) : N_2 \times N_2, & \quad G_{2a}(t) : N_2 \times (M - N_2).
\end{align*}
\]

(36) (37) (38)

Here, over channel use \( t \in \mathbb{N} \), the vector of symbols seen at Receiver \( i, i \in \{1, 2\} \), is \( Y_i(t) \), and the vector of symbols sent from the transmitter is \( X(t) = X_a(t) \bigtriangledown X_b(t) \bigtriangledown X_c(t) \). Channel matrices \( G_{1ab}(t), G_{2bc}(t) \) correspond to directions along which no zero-forcing is possible, while \( G_{1c}, G_{2a} \) 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 \( \beta_1, \beta_2 \) 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. \( \Gamma_i(t) \) are the zero-mean unit variance additive white Gaussian noise terms seen at outputs \( Y_i(t) \), independent of all inputs and channel realizations. The input vector \( X(t) \) 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 \( \epsilon \).

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 $C(P)$ are standard. The DoF region is defined as

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

(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 $N_2 \leq M$:

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

$$
d_1 \leq N_1,
$$

(40)

$$
d_2 \leq N_2,
$$

(41)

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

(42)

$$
d_1 + d_2 \leq N_2 + (M - N_2)\beta_2,
$$

(43)

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

(44)

where $\beta_o$ is defined as,

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

(45)

2. For $N_2 > M$:

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

$$
d_1 \leq N_1,
$$

(46)

$$
d_1 + d_2 \leq M
$$

(47)

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

(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) 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 [3].

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 3 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 3.3.1 and Section 3.3.2 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 $(\beta_1, \beta_2) = (1/2, 2/3)$

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,

$$D_1 = \left\{(d_1, d_2) \in \mathbb{R}^2_+ : d_1 \leq 2, \quad d_2 \leq 3, \quad \frac{d_1}{2} + \frac{d_2}{3} \leq \frac{3}{2}, \quad d_1 + d_2 \leq 3 + \frac{7}{9}\right\}$$

and $\beta_0 = \frac{7}{18}$ 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 $\mathbf{X}(t) = \mathbf{X}_a(t) \uplus \mathbf{X}_c(t)$, and outputs $\mathbf{Y}_1(t), \mathbf{Y}_2(t)$.

$$\mathbf{Y}_1(t) = [\mathbf{Y}_{11}(t) \quad \mathbf{Y}_{12}(t)]^T$$

---

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.
\[ \bar{Y}_{1r}(t) = \mathcal{L}^{g}_{r(t)} \left( \bar{X}_a(t) \triangledown (\bar{X}_c(t))^{1/2} \right), \forall r \in [2] \]  
(51)

\[ \bar{Y}_{2}(t) = [\bar{Y}_{21}(t) \quad \bar{Y}_{22}(t) \quad \bar{Y}_{23}(t)]^T \]  
(52)

\[ \bar{Y}_{2r}(t) = \mathcal{L}^{g}_{2r(t)} \left( (\bar{X}_a(t))^{1/3} \triangledown \bar{X}_c(t) \right), \forall r \in [3] \]  
(53)

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

\[ \bar{X}_a(t) = [\bar{X}_1(t) \quad \bar{X}_2(t)]^T \]  
(54)

\[ \bar{X}_c(t) = [\bar{X}_3(t) \quad \bar{X}_4(t) \quad \bar{X}_5(t)]^T \]  
(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 \((\beta_1, \beta_2) = (\frac{1}{2}, \frac{2}{3})\),

\[ 2H(\bar{Y}_{2}^{[n]} | W_1, \mathcal{G}) \leq 3H(\bar{Y}_{1}^{[n]} | W_1, \mathcal{G}) - H((\bar{Y}_{1}^{[n]})^{1/3} | W_1, \mathcal{G}) + 3n \log \bar{P} + n o (\log \bar{P}) \]  
(56)

See Figure 5 for an accompanying illustration for Lemma 1. The proof of Lemma 1 is presented in Appendix A.
6.3 Proof of the Bound $d_1 + d_2 \leq 3 + \frac{7}{5}$

1. Starting from Fano’s Inequality for the first receiver and suppressing $n o(\log(P))$ terms that are inconsequential for DoF, we have,

\[
3nR_1 \leq 3I(\tilde{Y}_1^n; W_1 | \mathcal{G}) \\
= 2H(\tilde{Y}_1^n | \mathcal{G}) - 3H(\tilde{Y}_1^n | W_1, \mathcal{G}) + H(\tilde{Y}_1^n | \mathcal{G}) \tag{57}
\]

\[
= 2H(\tilde{Y}_1^n | \mathcal{G}) - 3H(\tilde{Y}_1^n | W_1, \mathcal{G}) + H((\tilde{Y}_1^n)^{1/3} | \mathcal{G}) \tag{58}
\]

\[
= 2H(\tilde{Y}_1^n | \mathcal{G}) - 3H(\tilde{Y}_1^n | W_1, \mathcal{G}) + H((\tilde{Y}_1^n)^{1/3} | \mathcal{G}) + H((\tilde{Y}_1^n)^{2/3} | (\tilde{Y}_1^n)^{1/3}, \mathcal{G}) \tag{59}
\]

\[
\leq \left( 4 + \frac{4}{3} \right) n \log \tilde{P} - 3H(\tilde{Y}_1^n | W_1, \mathcal{G}) + H((\tilde{Y}_1^n)^{1/3} | \mathcal{G}) \tag{60}
\]

\[
\leq \left( 7 + \frac{4}{3} \right) n \log \tilde{P} - 2H(\tilde{Y}_2^n | W_1, \mathcal{G}) + H((\tilde{Y}_1^n)^{1/3} | \mathcal{G}) - H((\tilde{Y}_1^n)^{1/3} | W_1, \mathcal{G}) \tag{61}
\]

\[
\leq \left( 7 + \frac{4}{3} \right) n \log \tilde{P} - 2H(\tilde{Y}_2^n | W_1, \mathcal{G}) + H((\tilde{Y}_1^n)^{1/3} | W_2, \mathcal{G}) \tag{62}
\]

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., $2H(\tilde{Y}_1^n | \mathcal{G}) \leq 4n \log \tilde{P}$ and $H((\tilde{Y}_1^n)^{2/3} | (\tilde{Y}_1^n)^{1/3}, \mathcal{G}) \leq \frac{4}{3} n \log \tilde{P}$. \[(61)\] is obtained from Lemma \[(1)\]. \[(62)\] follows from the property that for independent random variables $B$ and $C$, $I(A; B) \leq I(A; B | C)$. As a result, we have $I((\tilde{Y}_1^n)^{1/3}; W_1 | \mathcal{G}) \leq I((\tilde{Y}_1^n)^{1/3}; W_1 | W_2, \mathcal{G}) \leq H((\tilde{Y}_1^n)^{1/3} | W_2, \mathcal{G})$.

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

\[
3nR_2 \leq I(\tilde{Y}_2^n; W_2 | \mathcal{G}) + 2I(\tilde{Y}_2^n; W_2 | W_1, \mathcal{G}) \tag{63}
\]

\[
\leq H(\tilde{Y}_2^n | \mathcal{G}) - H(\tilde{Y}_2^n | W_2, \mathcal{G}) + 2H(\tilde{Y}_2^n | W_1, \mathcal{G}) \leq 3n \log \tilde{P} - H(\tilde{Y}_2^n | W_2, \mathcal{G}) + 2H(\tilde{Y}_2^n | W_1, \mathcal{G}). \tag{64}
\]

\[(64)\] 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., $H(\tilde{Y}_2^n | \mathcal{G}) \leq 3n \log \tilde{P}$.

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

\[
3nR_1 + 3nR_2 \leq \left( 10 + \frac{4}{3} \right) n \log \tilde{P} - H(\tilde{Y}_2^n | W_2, \mathcal{G}) + H((\tilde{Y}_1^n)^{1/3} | W_2, \mathcal{G}) \tag{65}
\]

\[
\leq \left( 10 + \frac{4}{3} \right) n \log \tilde{P} \tag{66}
\]

where \[(66)\] is obtained by a direct application of the sumset inequalities of \[(5)\], as explained below. From \[(66)\], the sum DoF bound $d_1 + d_2 \leq 3 + \frac{7}{5}$ follows immediately.
4. Finally, we explain how (66) is implied by the sumset inequalities of [5]. Specifically, we need to prove that
\[
H((\hat{Y}_1^{|n|})^{1/3} | W_2, G) \leq H(\hat{Y}_2^{|n|} | W_2, G).
\]
In order to apply the result of Theorem 4 of [5] (i.e., Theorem 4 of [5]) to our setting, let us set \( M = 1, K = 2, \lambda_{1,1} = \lambda_{2,1} = 1, t_1 = t_2 = 1, \) and \( I_{11} = I_{21} = \{1\} \). Then, the inequality (29) reduces to,
\[
H(Z_1^{|n|}, Z_2^{|n|} | W, G) \geq H(Z_{11}^{|n|}, Z_{21}^{|n|} | W) + n \sigma(\log P).
\]
Let us specialize \( W = W_2 \) and define \( Z_1(t), Z_2(t), Z_{11}(t), Z_{21}(t), t \in [n], \) as,
\[
\bar{Y}_{21}(t) = L_{21}(t)(\langle X_1(t) \rangle^{1/3}, \langle X_2(t) \rangle^{1/3}, \bar{X}_3(t), \bar{X}_4(t), \bar{X}_5(t)) = Z_1(t)
\]
\[
\bar{Y}_{22}(t) = L_{22}(t)(\langle X_1(t) \rangle^{1/3}, \langle X_2(t) \rangle^{1/3}, \bar{X}_3(t), \bar{X}_4(t), \bar{X}_5(t)) = Z_2(t)
\]
\[
(\bar{Y}_{11}(t))^{1/3} = L_{21}(t)(\langle \hat{X}_1(t) \rangle^{1/3}, \langle \hat{X}_2(t) \rangle^{1/3}) = Z_{11}(t)
\]
\[
(\bar{Y}_{12}(t))^{1/3} = L_{22}(t)(\langle \hat{X}_1(t) \rangle^{1/3}, \langle \hat{X}_2(t) \rangle^{1/3}) = Z_{21}(t)
\]
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 4. In other words, Theorem 4 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 ‘\( \triangleq \)’ 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 ‘\( \triangleq \)’ in (70) and (71) is justified.

\[
(\bar{Y}_{11}(t))^{1/3} = \begin{bmatrix} \bar{Y}_{11}(t) - P \left[ \frac{\bar{Y}_{11}(t)}{P} \right] \\ \frac{\bar{Y}_{11}(t)}{P^{2/3}} \end{bmatrix} \begin{bmatrix} P^{2/3} \\ - P \left[ \frac{\bar{Y}_{11}(t)}{P} \right] \end{bmatrix} + \delta_a(t)
\]
\[
= \frac{\bar{Y}_{11}(t)}{P^{2/3}} - \delta_b(t) + \delta_a(t)
\]
\[
= \left[ \frac{g_1(t)X_1(t)}{P^{2/3}} \right] + \left[ \frac{g_2(t)X_2(t)}{P^{2/3}} \right] + \sum_{k=3}^5 \left[ \frac{g_k(t)X_k(t)}{P^{2/3}} \right] - \delta_b(t) + \delta_a(t)
\]
\[
= \left[ \frac{g_1(t)X_1(t)}{P^{2/3}} \right] + \left[ \frac{g_2(t)X_2(t)}{P^{2/3}} \right] + \delta_c(t) - \delta_b(t) + \delta_a(t)
\]
\[
= Z_{11}(t) + \delta_{11}(t)
\]
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 \( [A] + [B] - [A + B] \in \{-1, 0, 1\} \) for any real numbers \( A \) and \( B \). Next, consider \( \delta_b \), whose entropy is bounded as follows.

\[
H(\delta_b) \leq H\left( \left[ \frac{\bar{Y}_{11}(t)}{P} \right] \right) \leq O(1)
\]
\[ \text{(78)} \] is true as \(|\bar{Y}_{11}(t)| \leq 5\Delta \bar{P}\). Similarly, \(H(\delta_c) \leq O(1)\).
Thus, from \(\text{(181)}\) we have,
\[ H(Z_{11}^{[n]}, Z_{21}^{[n]} | W_2, G) \leq H(Z_{11}^{[n]}, Z_{22}^{[n]} | W_2, G) + n \log(\bar{P}) \tag{79} \]
\[ \Rightarrow H((\bar{Y}_1^{[n]})^{1/3} | W_2, G) \leq H(\bar{Y}_{21}^{[n]}, \bar{Y}_{22}^{[n]} | W_2, G) + n \log(\bar{P}) \tag{80} \]
\[ \leq H(\bar{Y}_2^{[n]} | W_2, G) + n \log(\bar{P}) \tag{81} \]

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

For the two-user \((M, N_1, N_2) = (4, 1, 3)\) with \((\beta_1, \beta_2) = \left(\frac{1}{4}, \frac{1}{2}\right)\) levels of partial CSIT, from Theorem 2 the DoF region is computed as,
\[ D_1 = \left\{ (d_1, d_2) \in \mathbb{R}^2^+: \ d_1 \leq 1, \ d_2 \leq 3, \ d_1 + \frac{d_2}{3} \leq \frac{5}{4}, \ d_1 + d_2 \leq 3 + \frac{7}{16} \right\} \tag{82} \]
and \(\beta_o = \frac{1}{16}\) from \(\text{(45)}\). The challenge is to prove the bound \(\text{(44)}\), i.e., \(d_1 + d_2 \leq 3 + \frac{7}{16}\).

7.1 Deterministic Model

Similar to Section 6.1, the deterministic transformation produces a BC with input \(\bar{X}(t) = \bar{X}_a(t) \triangledown \bar{X}_c(t)\), and outputs \(\bar{Y}_1(t), \bar{Y}_2(t)\).
\[ \bar{Y}_1(t) = L^g_{1(t)} \left( \bar{X}_a(t) \triangledown (\bar{X}_c(t))^{3/4} \right) \tag{83} \]
\[ \bar{Y}_2(t) = [\bar{Y}_{21}(t) \quad \bar{Y}_{22}(t)]^T \tag{84} \]
\[ \bar{Y}_{2r}(t) = L^g_{2r(t)} \left( (\bar{X}_a(t))^{1/2} \triangledown \bar{X}_c(t) \right), \forall r \in [2] \tag{85} \]
where \(\bar{X}_a(t)\) and \(\bar{X}_c(t)\) are defined as,
\[ \bar{X}_a(t) = \bar{X}_1(t) \tag{86} \]
\[ \bar{X}_c(t) = [\bar{X}_2(t) \quad \bar{X}_3(t)]^T \tag{87} \]
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 \((\beta_1, \beta_2) = \left(\frac{1}{4}, \frac{1}{2}\right)\),
\[ H(\bar{Y}_2^{[n]} | W_1, G) \leq 4H(\bar{Y}_1^{[n]} | W_1, G) - 3H((\bar{Y}_1^{[n]})^{1/2} | W_1, G) + \frac{3}{4} n \log \bar{P} + n \log(\bar{P}) \tag{88} \]

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

\[
4nR_1 \leq H(Y_1[n] | G) - H(Y_1[n] | W_1, G) + 3H(Y_1[n] | G) - 3H(Y_1[n] | W_1, G) \tag{89}
\]

\[
= H(Y_1[n]) - H(Y_1[n] | W_1, G)
+ 3H((Y_1[n])_{1/2}, (Y_1[n])_{1/2} | G) - 3H((Y_1[n])_{1/2}, (Y_1[n])_{1/2} | W_1, G) \tag{90}
\]

\[
= H(Y_1[n]) - H(Y_1[n] | W_1, G)
+ 3H((Y_1[n])_{1/2} | G) - 3H((Y_1[n])_{1/2} | W_1, G)
+ 3H((Y_1[n])_{1/2} | (Y_1[n])_{1/2}, G) - 3H((Y_1[n])_{1/2} | (Y_1[n])_{1/2}, W_1, G) \tag{91}
\]

\[
\leq \frac{5}{2} n \log \bar{P} - H(Y_1[n]) - H(Y_1[n] | W_1, G)
+ 3H((Y_1[n])_{1/2} | G) - 3H((Y_1[n])_{1/2} | W_1, G) \tag{92}
\]

\[
\leq \frac{13}{4} n \log \bar{P} - H(Y_1[n], Y_2[n], Y_3[n], Y_4[n] | W_1, G)
+ 3H((Y_1[n])_{1/2} | G) - 3H((Y_1[n])_{1/2} | W_1, G)
+ n o(\log \bar{P}) \tag{93}
\]

\[
\leq \frac{13}{4} n \log \bar{P} - H(Y_1[n], Y_2[n], Y_3[n], Y_4[n] | W_1, G)
+ 3H((Y_1[n])_{1/2} | W_2, G) + n o(\log \bar{P}) \tag{94}
\]

where (89) 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., \(3H(\bar{Y}_1)_{1/2} | (Y_1)_{1/2}, G) \leq \frac{3}{4} n \log \bar{P}, H(\bar{Y}_1 | G) \leq n \log \bar{P}\). (93) is obtained as from Lemma 2 and the chain rule we have

\[
H(\bar{Y}_1, \bar{Y}_2, \bar{Y}_3 | W_1, G) \\
\leq 4H(\bar{Y}_1 | W_1, G) - 3H((\bar{Y}_1)_{1/2} | W_1, G) + \frac{3}{4} n \log \bar{P} + n o(\log \bar{P}) \\
= H(\bar{Y}_1 | W_1, G) + 3H((\bar{Y}_1)_{1/2} | (\bar{Y}_1)_{1/2}, W_1, G) + \frac{3}{4} n \log \bar{P} + n o(\log \bar{P}) \quad (95)
\]

(94) is true because \(I(A; B) \leq I(A; B | C)\) when \(B\) is independent of \(C\). As a result, we have \(I((\bar{Y}_1)_{1/2}; W_1 | G) \leq H((\bar{Y}_1)_{1/2} | W_2, G)\).

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

\[
nR_2 \leq I(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23}; W_2 | G) \\
= H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | G) - H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | W_2, G) \quad (96)
\]

\[
nR_2 \leq I(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23}; W_2 | W_1, G) \\
= H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | W_1, G) \quad (97)
\]

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

\[
4nR_2 \leq 3H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | G) - 3H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | W_2, G) + H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | W_1, G) \\
\leq 9n \log \bar{P} - 3H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | W_2, G) + H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | W_1, G) \quad (98)
\]

(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., \(H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | G) \leq 2n \log \bar{P}\).

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

\[
4nR_1 + 4nR_2 \leq \frac{49}{4} n \log \bar{P} - 3H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | W_2, G) + 3H((\bar{Y}_1)_{1/2} | W_2, G) + n o(\log \bar{P}) \quad (99)
\]

\[
\leq \frac{49}{4} n \log \bar{P} + n o(\log \bar{P}) \quad (100)
\]

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 \((Y_1(t))^{1/2}\) is a bounded density linear combination of random variables \((\bar{X}_1(t))^{1/2}, (\bar{X}_2(t))^{1/2}, (\bar{X}_3(t))^{1/2}, (\bar{X}_4(t))^{1/2}\) 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

\[
H((\bar{Y}_1)_{1/2} | W_2, G) \\
\leq H(\bar{Y}_{21} | W_2, G) + n o(\log \bar{P}) \quad (101)
\]

\[
\leq H(\bar{Y}_{21}, \bar{Y}_{22}, \bar{Y}_{23} | W_2, G) + n o(\log \bar{P}) \quad (102)
\]
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 $\bar{Y}_1(t)$ and $\bar{Y}_2(t)$, which are defined as follows.

\[
\bar{Y}_1(t) = [\bar{Y}_{11}(t) \ \bar{Y}_{12}(t) \ \cdots \ \bar{Y}_{1N_1}(t)]^T
\]

(103)

\[
\bar{Y}_{1r}(t) = L_{1r}^0(t) (\bar{X}_a(t) \bigtriangledown \bar{X}_b(t) \bigtriangledown (\bar{X}_c(t))^\dagger), \forall r \in [N_1]
\]

(104)

\[
\bar{Y}_2(t) = [\bar{Y}_{21}(t) \ \bar{Y}_{22}(t) \ \cdots \ \bar{Y}_{2N_2}(t)]^T
\]

(105)

\[
\bar{Y}_{2r}(t) = L_{2r}^0(t) \left( (\bar{X}_a(t))^n \bigtriangledown \bar{X}_b(t) \bigtriangledown \bar{X}_c(t) \right), \forall r \in [N_2]
\]

(106)

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

\[
\bar{X}(t) = [\bar{X}_1(t), \bar{X}_2(t), \cdots, \bar{X}_M(t)]
\]

(107)

\[
\bar{X}_a(t) = [\bar{X}(t)]_{0 \rightarrow M - N_2}
\]

(108)

\[
\bar{X}_b(t) = [\bar{X}(t)]_{M - N_2 \rightarrow N_1 + N_2 - M}
\]

(109)

\[
\bar{X}_c(t) = [\bar{X}(t)]_{N_1 \rightarrow M - N_1}
\]

(110)

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

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

(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 \leq N_2 [14]] Define the two random variables $\bar{U}_1$ and $\bar{U}_2$ as,

\[
\bar{U}_1 = \left( U_{11}^{[n]}, U_{12}^{[n]}, \cdots, U_{1N_1}^{[n]} \right)
\]

(112)

\[
\bar{U}_2 = \left( U_{21}^{[n]}, U_{22}^{[n]}, \cdots, U_{2N_2}^{[n]} \right)
\]

(113)

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

\[
U_{1j}(t) = L_{1j}^0(t) \left( (\bar{V}_1(t))^\eta_{-\lambda_1} \bigtriangledown (\bar{V}_2(t))^\eta_{-\lambda_2} \bigtriangledown \cdots \bigtriangledown (\bar{V}_l(t))^\eta_{-\lambda_l} \right), \forall j \in [N_1]
\]

(114)

\[
U_{2j}(t) = L_{2j}^0(t) \left( (\bar{V}_1(t))^\eta_{-\lambda_1} \bigtriangledown (\bar{V}_2(t))^\eta_{-\lambda_2} \bigtriangledown \cdots \bigtriangledown (\bar{V}_l(t))^\eta_{-\lambda_l} \right), \forall j \in [N_2]
\]

(115)

where $\bar{V}_i(t) = [\bar{V}_{i1}(t) \ \cdots \ \bar{V}_{iM}(t)]^T$, $\bar{V}_{im}(t) \in \mathcal{X}_\eta$ are all independent of $\mathcal{G}$, and $0 \leq \lambda_{1i}, \lambda_{2i} \leq \eta$ for all $i \in \{1, \cdots, \lambda\}$. 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,

\[
H(\tilde{U}_1 \mid W, \mathcal{G}) - H(\tilde{U}_2 \mid W, \mathcal{G}) \\
\leq n \left( (N_1 - \sum_i M_i)(\lambda_{1,s+1} - \lambda_{2,s+1}) + \sum_i M_i(\lambda_{1i} - \lambda_{2i})^+ \right) \log \bar{P} + n \circ (\log \bar{P}) \tag{116}
\]

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 \( \bar{Y}_1 \) into \( \bar{Y}_{1a}, \bar{Y}_{1\bar{a}} \)

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

\[
\bar{Y}_{1\bar{a}}(t) = [\bar{Y}_{11\bar{a}}(t) \ \bar{Y}_{12\bar{a}}(t) \ \cdots \ \bar{Y}_{1N_1+N_2-M\bar{a}}(t)]^T \tag{117}
\]

\[
\bar{Y}_{1a}(t) = [\bar{Y}_{11a}(t) \ \bar{Y}_{12a}(t) \ \cdots \ \bar{Y}_{1M-N_2\bar{a}}(t)]^T \tag{119}
\]

\[
\bar{Y}_{1\bar{a}}(t) = L_{1\bar{a}}^a(t) (\bar{X}_a(t) \cup (\bar{X}_c(t))_{\beta_1}^1), \forall r \in [N_1 + N_2 - M] \tag{118}
\]

\[
\bar{Y}_{1a}(t) = L_{1a}^a(t) (\bar{X}_a(t) \cup \bar{X}_b(t) \cup (\bar{X}_c(t))_{\beta_1}), \forall r \in [M - N_2] \tag{120}
\]

#### 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,

\[
d_1 + d_2 \leq N_2 + (M - N_2)\beta_o \tag{121}
\]

where \( \beta_o \) is equal to,

\[
\beta_o = \frac{N_1 - N_2 + (N_2 - N_1)\beta_2 + (M - N_1)\beta_1}{M - N_1} \tag{122}
\]

Corresponding to Lemma\(^{[\text{I}]}\) 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,

\[
\hat{N}_1 H(\hat{Y}_2^n \mid W_1, \mathcal{G}) \leq (\hat{N}_1 + \hat{N}_2) H(\hat{Y}_2^n \mid W_1, \mathcal{G}) - \hat{N}_2 H(\hat{Y}_{1\bar{a}}^n, (\bar{Y}_{1a}^n)^{1-\beta_2} \mid W_1, \mathcal{G}) + n\hat{N}_0 \log \bar{P} + n \circ (\log \bar{P}) \tag{123}
\]

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

\[
\hat{N}_0 = \hat{N}_1(\hat{N}_2 + \hat{N}_1)\beta_1 \tag{124}
\]

\[
\hat{N}_1 = M - N_2 \tag{125}
\]

\[
\hat{N}_2 = N_2 - N_1 \tag{126}
\]

---

\(^{4}\)Let \( \mathcal{G}(U) \subset \mathcal{G} \) denote the set of all bounded density channel coefficients that appear in \( \tilde{U}_1, \tilde{U}_2 \). \( W \) 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\(^{[\text{I}]}\).
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 \(n \log(P)\) terms that are inconsequential for DoF, we have,

\[
n(\hat{N}_1 + \hat{N}_2)R_1
\]

\[
\leq (\hat{N}_1 + \hat{N}_2)I(W_1; \tilde{Y}_1^{[n]} | \mathcal{G})
\]  

(127)

\[
= \hat{N}_1H(\tilde{Y}_1^{[n]} | \mathcal{G}) - (\hat{N}_1 + \hat{N}_2)H(\tilde{Y}_1^{[n]} | W_1, \mathcal{G}) + \hat{N}_2H(\tilde{Y}_1^{[n]} | \mathcal{G})
\]  

(128)

\[
= \hat{N}_1H(\tilde{Y}_1^{[n]} | \mathcal{G}) - (\hat{N}_1 + \hat{N}_2)H(\tilde{Y}_1^{[n]} | W_1, \mathcal{G}) + \hat{N}_2H((\tilde{Y}_1^{[n]}_{\alpha_2}, (\tilde{Y}_1^{[n]}_{\alpha_2})^{1-\beta_2}, \tilde{Y}_1^{[n]}_{\alpha_2} | \mathcal{G})
\]  

(129)

\[
\leq n\hat{N}_1N_1 \log(P) - (\hat{N}_1 + \hat{N}_2)H(\tilde{Y}_1^{[n]} | W_1, \mathcal{G}) + \hat{N}_2H((\tilde{Y}_1^{[n]}_{\alpha_2}, (\tilde{Y}_1^{[n]}_{\alpha_2})^{1-\beta_2} | \mathcal{G})
\]  

(130)

\[
\leq n(\hat{N}_1N_1 + \hat{N}_1\hat{N}_2\beta_2) \log(P) - (\hat{N}_1 + \hat{N}_2)H(\tilde{Y}_1^{[n]} | W_1, \mathcal{G}) + \hat{N}_2H((\tilde{Y}_1^{[n]}_{\alpha_2}, (\tilde{Y}_1^{[n]}_{\alpha_2})^{1-\beta_2} | \mathcal{G})
\]  

(131)

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 \(\tilde{Y}_1^{[n]}\) cannot be more than \(nN_1 \log(P) + n \log(P)\), 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.,

\[
\hat{N}_2H((\tilde{Y}_1^{[n]}_{\alpha_2}, (\tilde{Y}_1^{[n]}_{\alpha_2})^{1-\beta_2}, \mathcal{G}) \leq n\hat{N}_2\hat{N}_1\beta_2 \log(P) + n \log(P))\].

Applying Lemma 4 to (131) produces (132). Finally, (134) is true because \(I(A; B) \leq I(A; B | C) \leq H(B | C)\) whenever \(A\) is independent of \(C\).

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

\[
n(\hat{N}_1 + \hat{N}_2)R_2
\]

\[
\leq \hat{N}_1I(\tilde{Y}_2^{[n]}; W_2 | W_1, \mathcal{G}) + \hat{N}_2I(\tilde{Y}_2^{[n]}; W_2 | \mathcal{G})
\]  

(135)

\[
\leq \hat{N}_1I(\tilde{Y}_2^{[n]}; W_1, \mathcal{G}) + \hat{N}_2I(\tilde{Y}_2^{[n]}; W_2, \mathcal{G})
\]  

(136)

\[
\leq \hat{N}_1H(\tilde{Y}_2^{[n]}; W_1, \mathcal{G}) + n\hat{N}_2\hat{N}_1\beta_2 \log(P) - \hat{N}_2H(\tilde{Y}_2^{[n]}; W_2, \mathcal{G})
\]  

(137)

(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.,

\[
H(\tilde{Y}_2^{[n]} | \mathcal{G}) \leq n\hat{N}_2 \log(P) + n \log(P).
\]

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

\[
n(\hat{N}_1 + \hat{N}_2)R_1 + n(\hat{N}_1 + \hat{N}_2)R_2
\]

\[
\leq n(\hat{N}_0 + \hat{N}_1N_1 + \hat{N}_1\hat{N}_2\beta_2 + \hat{N}_2N_2) \log(P) - \hat{N}_2H(\tilde{Y}_2^{[n]}; W_2, \mathcal{G})
\]

\[
+ \hat{N}_2H(\tilde{Y}_1^{[n]}_{\alpha_2}, (\tilde{Y}_1^{[n]}_{\alpha_2})^{1-\beta_2} | W_2, \mathcal{G})
\]  

(137)
In this section we prove the bound (44) for general (44).

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,

\[
d_1 + d_2 \leq N_2 + (M - N_2)\beta_o
\]

where \( \beta_o \) is equal to,

\[
\beta_o = \frac{\beta_1\beta_2(M - N_2)}{(N_2 - N_1)(1 - \beta_1) + (M - N_2)\beta_2}
\]

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,

\[
\begin{align*}
\tilde{N}_1 H(\tilde{Y}_1^n | W_1, G) \\
\leq (\tilde{N}_1 + \tilde{N}_2)H(Y_1^n | W_1, G) - \tilde{N}_2 H(\tilde{Y}_{1a}, (\tilde{Y}_{1b})^{1-\beta_2} | W_1, G) + n\tilde{N}_0 \log \tilde{P} + n o \log \tilde{P}
\end{align*}
\]

where the numbers \( \tilde{N}_0, \tilde{N}_1 \) and \( \tilde{N}_2 \) are defined as,

\[
\begin{align*}
\tilde{N}_0 &= \tilde{N}_1(M - N_1)\beta_1 \\
\tilde{N}_1 &= (M - N_2)\beta_2 \\
\tilde{N}_2 &= (N_2 - N_1)(1 - \beta_1)
\end{align*}
\]

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,

\[
\begin{align*}
n(\tilde{N}_1 + \tilde{N}_2)R_1 \\
\leq n(\tilde{N}_0 + \tilde{N}_1 N_1 + \tilde{N}_1 \tilde{N}_2) \log \tilde{P} - \tilde{N}_1 H(\tilde{Y}_1^n | W_1, G) + \tilde{N}_2 H(\tilde{Y}_{1a}, (\tilde{Y}_{1b})^{1-\beta_2} | W_2, G) + n o \log \tilde{P}
\end{align*}
\]

where the difference between (134) and (146) is due to the difference in the definitions of \((\tilde{N}_1, \tilde{N}_2)\) versus \((\tilde{N}_1, \tilde{N}_2)\).
2. Similarly, for the second receiver we have,
\[
\begin{align*}
n(\tilde{N}_1 + \tilde{N}_2)R_2 &\leq \tilde{N}_2H(\tilde{Y}_2^n | \mathcal{G}) - \tilde{N}_2H(\tilde{Y}_2^n | W_2, \mathcal{G}) + \tilde{N}_1H(\tilde{Y}_2^n | W_1, \mathcal{G}) \\
&\leq n\tilde{N}_2N_2 \log \tilde{P} - \tilde{N}_2H(\tilde{Y}_2^n | W_2, \mathcal{G}) + \tilde{N}_1H(\tilde{Y}_2^n | W_1, \mathcal{G})
\end{align*}
\] (147)

3. Summing the inequalities (146) and (147) results in,
\[
\begin{align*}
n(\tilde{N}_1 + \tilde{N}_2)R_1 + n(\tilde{N}_1 + \tilde{N}_2)R_2 &\leq n(\tilde{N}_0 + \tilde{N}_1N_1 + \tilde{N}_1\tilde{N}_2 + \tilde{N}_2N_2) \log \tilde{P} - \tilde{N}_2H(\tilde{Y}_2^n | W_2, \mathcal{G}) \\
&+ \tilde{N}_2H(\tilde{Y}_1^{[a]}), (\tilde{Y}_1^{[b]})^{1-\beta_2} | W_2, \mathcal{G}) + n o (\log \tilde{P}) \\
&\leq n(\tilde{N}_0 + \tilde{N}_1N_1 + \tilde{N}_1\tilde{N}_2 + \tilde{N}_2N_2) \log \tilde{P} + n o (\log \tilde{P})
\end{align*}
\] (148)

(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 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 $o(\log(P))$ that are inconsequential for DoF, we proceed as follows.

\[2H(\tilde{Y}_2^n | W_1, G) + H((\tilde{Y}_1^n)^{1/3} | W_1, G)\]
\[= 2H((\tilde{Y}_2^n)^{1/2}, (\tilde{Y}_2^n)^{1/2} | W_1, G) + H((\tilde{Y}_1^n)^{1/3} | W_1, G)\]  
(150)
\[\leq 2H((\tilde{Y}_2^n)^{1/2} | W_1, G) + H((\tilde{Y}_1^n)^{1/3} | W_1, G) + 3n \log \tilde{P}\]  
(151)
\[\leq 2H((\tilde{Y}_3^n)^{1/2}, (\tilde{X}_4^n)^{1/2}, (\tilde{X}_4^n)^{1/2} | W_1, G) + H((\tilde{Y}_1^n)^{1/3} | W_1, G) + 3n \log \tilde{P}\]  
(152)
\[\leq 2H((\tilde{Y}_3^n)^{1/3}, (\tilde{X}_4^n)^{1/2}, (\tilde{X}_3^n)^{1/2} | W_1, G) + H((\tilde{Y}_1^n)^{1/3} | W_1, G) + 3n \log \tilde{P}\]  
(153)
\[= 2H((\tilde{X}_3^n)^{1/2}, (\tilde{X}_4^n)^{1/2}, (\tilde{X}_3^n)^{1/2} | (\tilde{Y}_1^n)^{1/3}, W_1, G) + 3H((\tilde{Y}_1^n)^{1/3} | W_1, G) + 3n \log \tilde{P}\]  
(154)
\[\leq H((\tilde{X}_3^n)^{1/2}, (\tilde{X}_4^n)^{1/2} | (\tilde{Y}_1^n)^{1/3}, W_1, G) + H((\tilde{X}_3^n)^{1/2}, (\tilde{X}_5^n)^{1/2} | (\tilde{Y}_1^n)^{1/3}, W_1, G)\]
\[+ H((\tilde{X}_4^n)^{1/2}, (\tilde{X}_5^n)^{1/2} | (\tilde{Y}_1^n)^{1/3}, W_1, G) + 3H((\tilde{Y}_1^n)^{1/3} | W_1, G) + 3n \log \tilde{P}\]  
(155)
\[= H((\tilde{X}_3^n)^{1/2}, (\tilde{X}_4^n)^{1/2}, (\tilde{Y}_1^n)^{1/3} | W_1, G) + H((\tilde{X}_3^n)^{1/2}, (\tilde{X}_5^n)^{1/2} | (\tilde{Y}_1^n)^{1/3}, W_1, G)\]
\[+ H((\tilde{X}_4^n)^{1/2}, (\tilde{X}_5^n)^{1/2}, (\tilde{Y}_1^n)^{1/3} | W_1, G) + 3n \log \tilde{P}\]  
(156)
\[\leq 3H(\tilde{Y}_1^n | W_1, G) + 3n \log \tilde{P}\]  
(157)

(150) follows from Definition 2 and (151) is true because $Y_2$ has 3 antennas and the entropy of the bottom half of the signal power levels on each of them can altogether contribute at most $\frac{3}{2}n \log \tilde{P}$. Next, (152) is true because $(Y_2)^{1/2}$ is a function of $(\tilde{X}_3^n)^{1/2}, (\tilde{X}_4^n)^{1/2}, (\tilde{X}_5^n)^{1/2}$, 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 $C$,

\[2H(A, B, C) \leq H(A, B) + H(A, C) + H(B, C)\]  
(158)

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

\[H((\tilde{X}_3^n)^{1/2}, (\tilde{X}_4^n)^{1/2}, (\tilde{Y}_1^n)^{1/3} | W_1, G) \leq H(\tilde{Y}_1^n | W_1, G) + n \log \tilde{P}\]  
(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,

\[Z_{11}(t) = (\tilde{Y}_{11}(t))^{1/3} \leq L_{11}(t)((\tilde{X}_1(t))^{1/3}, (\tilde{X}_2(t))^{1/3})\]  
(160)
\[Z_{12}(t) = L_{12}(t)((\tilde{X}_1(t))^{1/2}, (\tilde{X}_2(t))^{1/2}, (\tilde{X}_3(t))^{1/2}, (\tilde{X}_4(t))^{1/2}, (\tilde{X}_5(t))^{1/2}) = (\tilde{X}_3(t))^{1/2}\]  
(161)
\[Z_{21}(t) = (\tilde{Y}_{21}(t))^{1/3} \leq L_{21}(t)((\tilde{X}_1(t))^{1/3}, (\tilde{X}_2(t))^{1/3})\]  
(162)
\[Z_{22}(t) = L_{22}(t)((\tilde{X}_1(t))^{1/2}, (\tilde{X}_2(t))^{1/2}, (\tilde{X}_3(t))^{1/2}, (\tilde{X}_4(t))^{1/2}, (\tilde{X}_5(t))^{1/2}) = (\tilde{X}_5(t))^{1/2}\]  
(163)
\[Z_1(t) = \tilde{Y}_{11}(t) = L_{11}(t)(\tilde{X}_1(t), \tilde{X}_2(t), (\tilde{X}_3(t))^{1/2}, (\tilde{X}_4(t))^{1/2}, (\tilde{X}_5(t))^{1/2})\]  
(164)
\[Z_2(t) = \tilde{Y}_{12}(t) = L_{12}(t)(\tilde{X}_1(t), \tilde{X}_2(t), (\tilde{X}_3(t))^{1/2}, (\tilde{X}_4(t))^{1/2}, (\tilde{X}_5(t))^{1/2})\]  
(165)
Note that for any $i, j$ the linear combination $L_{ij}^h(t)$ is arbitrary linear combination of random variables satisfying the condition in Definition 6. Thus, some of the coefficients in $L_{ij}^h(t)$ can be chosen to be zero, e.g., we choose $L_{12}^h(t)((\bar{X}_1(t))_{1/2}, (\bar{X}_2(t))_{1/2}, (\bar{X}_3(t))_{1/2}, (\bar{X}_4(t))_{1/2}, (\bar{X}_5(t))_{1/2})$ to be $(\bar{X}_3(t))_{1/2}$. Thus, from Theorem 1 we have,

$$H(Z_1^n, Z_2^n, Z_3^n, Z_4^n | W_1, \mathcal{G}) \leq H(Z_1^n, Z_2^n | W_1, \mathcal{G}) + o(\log \bar{P}) \quad (166)$$

$$\Rightarrow H((\bar{X}_3^n))_{1/2}, (\bar{X}_5^n))_{1/2}, (\bar{Y}_1^n))_{1/2} | W_1, \mathcal{G}) \leq H(\bar{Y}_1^n | W_1, \mathcal{G}) + o(\log \bar{P}) \quad (167)$$

See Figure 8 and 9

![Figure 8](image)

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

### B Proof of Lemma 2

Note that $2H((\bar{Y}_2^n)_{1/4}, (\bar{Y}_2^n)_{1/4} | W_1, \mathcal{G}) \leq \log \bar{P}$. Therefore, it is sufficient to prove,

$$2H((\bar{Y}_2^n)_{3/4}, (\bar{Y}_2^n)_{3/4} | W_1, \mathcal{G}) + 3H((\bar{Y}_1^n)_{1/2} | W_1, \mathcal{G}) \leq 5H(\bar{Y}_1^n | W_1, \mathcal{G}) + o(\log \bar{P}) \quad (168)$$

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

$$C_i(t) = \begin{cases} 
(\bar{X}_5(t))_{3/4}, i = 1 \\
(\bar{X}_5(t))_{1/2}, i = 2 \\
L_3^{g}((\bar{X}_1(t))_{1/4}, (\bar{X}_2(t))_{1/4}, (\bar{X}_3(t))_{1/4}, (\bar{X}_4(t))_{1/4}, (\bar{X}_5(t))_{1/4}), i = 3 \\
(\bar{X}_4(t))_{3/4}, i = 4 \\
(\bar{X}_4(t))_{1/2}, i = 5 \\
L_6^{g}((\bar{X}_1(t))_{1/4}, (\bar{X}_2(t))_{1/4}, (\bar{X}_3(t))_{1/4}, (\bar{X}_4(t))_{1/4}, (\bar{X}_5(t))_{1/4}), i = 6 \\
C_1(t), i = 7 
\end{cases} \quad (169)$$

Starting from the left side of (168), we have
\[
2H((\tilde{Y}_{21}^{[n]})^{3/4}, (\tilde{Y}_{22}^{[n]})^{3/4} \mid W_1, \mathcal{G}) + 3H((\tilde{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G})
= 2H(C_1^{[n]}, C_2^{[n]}, C_3^{[n]}, C_4^{[n]}, C_5^{[n]}, C_6^{[n]} \mid W_1, \mathcal{G}) + 3H((\tilde{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G})
\leq 2H((\tilde{Y}_1^{[n]})^{1/2}, C_2^{[n]}, C_3^{[n]}, C_4^{[n]}, C_5^{[n]}, C_6^{[n]} \mid W_1, \mathcal{G}) + 3H((\tilde{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G}) + n\, o \,(\log \bar{P}) \quad (\text{170})
= 2H(C_2^{[n]}, C_3^{[n]}, C_4^{[n]}, C_5^{[n]}, C_6^{[n]} \mid (\tilde{Y}_1^{[n]})^{1/2}, W_1, \mathcal{G}) + 5H((\tilde{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G}) + n\, o \,(\log \bar{P}) \quad (\text{172})
\leq \sum_{i=2}^{6} H(C_i^{[n]}, C_{i+1}^{[n]} \mid (\tilde{Y}_1^{[n]})^{1/2}, W_1, \mathcal{G}) + 5H((\tilde{Y}_1^{[n]})^{1/2} \mid W_1, \mathcal{G}) + n\, o \,(\log \bar{P}) \quad (\text{173})
= \sum_{i=2}^{6} H((\tilde{Y}_1^{[n]})^{1/2}, C_i^{[n]}, C_{i+1}^{[n]} \mid W_1, \mathcal{G}) + n\, o \,(\log \bar{P}) \quad (\text{174})
\leq 5H(\tilde{Y}_1^{[n]} \mid W_1, \mathcal{G}) + n\, o \,(\log \bar{P}) \quad (\text{175})
\]

(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,

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

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,

\[
H(Z_1^{[n]} \mid W, \mathcal{G}) \geq H(Z_{11}^{[n]} \mid W) + n\, o(\log \bar{P}). \quad (\text{177})
\]
Figure 10: The random variables \((\bar{X}_1(t))^{1/4}, (\bar{X}_2(t))^{1/4}, (\bar{X}_3(t))^{1/4}, (\bar{X}_4(t))^{3/4}, (\bar{X}_5(t))^{3/4}\) and their partitions \(C_i(t)\) are specified.

where

\[
Z_1(t) = (\bar{Y}_1(t))^{1/2} \quad (178)
\]

\[
Z_{11}(t) = C_1(t) \quad (179)
\]

\[
W = C_2^{[n]}, C_3^{[n]}, C_4^{[n]}, C_5^{[n]}, C_6^{[n]}, W_1 \quad (180)
\]

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

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

where

\[
Z_1(t) = \bar{Y}_1(t) \quad (182)
\]

\[
Z_{11}(t) = (\bar{Y}_1(t))^{1/2} \quad (183)
\]

\[
W = C_2^{[n]}, C_3^{[n]}, C_4^{[n]}, C_5^{[n]}, C_6^{[n]}, W_1 \quad (184)
\]

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

\[
\mathcal{T}(Z_{1,2}) + \mathcal{T}(Z_{1,3}) \leq \lambda_{1,1} + \lambda_{1,2} \quad (187)
\]

\[
\mathcal{T}(Z_{1,3}) \leq \lambda_{1,1} \quad (188)
\]
Note that $T(C_i(t)) = \frac{1}{i}$ from Figure 10. For instance, illustration of

$$H((\tilde{Y}_{1}^{[n]} | W_1, \mathcal{G}) \leq H(\tilde{Y}_{1}^{n} | W_1, \mathcal{G})$$

(189)

is shown in Figure 11

C  Proof of Lemma 4

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

$$\tilde{Y}_{1c}(t) = \begin{bmatrix} \bar{y}_{11c}(t) & \bar{y}_{12c}(t) & \cdots & \bar{y}_{1,N_1+N_2-M,c}(t) \end{bmatrix}^T \quad (190)$$

$$\bar{y}_{1rc}(t) = L_{3r}(t) \left( (\bar{X}_a(t))^{1-\beta_2} \nabla \bar{X}_b(t) \nabla (\bar{X}_c(t))^{1-\beta_1} \right), \forall r \in [N_1 + N_2 - M]$$

(191)

$$\tilde{Y}_{1d}(t) = \begin{bmatrix} \bar{y}_{11d}(t) & \bar{y}_{12d}(t) & \cdots & \bar{y}_{1,M-N_2,d}(t) \end{bmatrix}^T \quad (192)$$

$$\bar{y}_{1rd}(t) = L_{3r}(t) \left( (\bar{X}_a(t) \nabla (\bar{X}_c(t)))^{1-\beta_1} \right), \forall r \in [M - N_2]$$

(193)

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

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

$$\bar{Y}_{2a}(t) = \begin{bmatrix} \bar{y}_{21a}(t) & \bar{y}_{22a}(t) & \cdots & \bar{y}_{2,N_1+N_2-M,a}(t) \end{bmatrix}^T \quad (194)$$

$$\bar{y}_{2ra}(t) = L_{4r}(t) \left( (\bar{X}_a(t))^{1-\beta_2} \nabla \bar{X}_b(t) \right), \forall r \in [N_1 + N_2 - M]$$

(195)

$$\bar{Y}_{2b}(t) = \begin{bmatrix} \bar{y}_{21b}(t) & \bar{y}_{22b}(t) & \cdots & \bar{y}_{2,M-N_2,b}(t) \end{bmatrix}^T \quad (196)$$

$$\bar{y}_{2rb}(t) = L_{4r}(t) \left( (\bar{X}_a(t))^{1-\beta_2} \nabla \bar{X}_c(t) \right), \forall r \in [M - N_1]$$

(197)

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

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

$$\hat{N}_1 H(\bar{Y}_{2a}^{[n]}), \bar{Y}_{2b}^{[n]} | W_1, \mathcal{G}) = n \hat{N}_0 \log \hat{P}$$

(198)

$$\leq \hat{N}_1 H((\bar{Y}_{2b}^{[n]}))_{\beta_1} | W_1, \mathcal{G})$$

(199)

$$\leq \hat{N}_1 H(\bar{Y}_{2a}^{[n]}, (\bar{Y}_{2b}^{[n]}))_{\beta_1} | W_1, \mathcal{G})$$

(200)

where (200) is true from the chain rule.

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

\[ H((\bar{Y}_{1d}^n)_{\beta_2} \mid \bar{Y}_{1c}^n, (\bar{Y}_{1d}^n)^{1-\beta_2}, W_1, \mathcal{G}) = H(\bar{Y}_{1c}^n \mid W_1, \mathcal{G}) - H(\bar{Y}_{1c}^n, (\bar{Y}_{1d}^n)^{1-\beta_2} \mid W_1, \mathcal{G}) \]

(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,

\[
\begin{align*}
\hat{N}_1 H(\tilde{Y}_{2a}^c, (\tilde{Y}_{2b}^c)^{1-\beta_1} \mid W_1, \mathcal{G}) + \hat{N}_2 H(\tilde{Y}_{1c}^n, (\tilde{Y}_{1d}^n)^{1-\beta_2} \mid W_1, \mathcal{G}) \\
\leq (\hat{N}_1 + \hat{N}_2) H(\tilde{Y}_1^c, (\tilde{Y}_1^n)^{1-\beta_2} \mid W_1, \mathcal{G}) + n o \left( \log \hat{P} \right)
\end{align*}
\]  

(202)

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

\[
\tilde{C}_i(t) = \begin{cases} 
\tilde{X}_{i-\hat{N}_1-\hat{N}_2-M}(t) & 0 < i \leq \hat{N}_1 + \hat{N}_2 \\
\tilde{C}_{i-\hat{N}_1-\hat{N}_2}(t) & i > \hat{N}_1 + \hat{N}_2
\end{cases}
\]  

(203)

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

\[
\begin{align*}
\hat{N}_1 H(\tilde{Y}_{2a}^c, (\tilde{Y}_{2b}^c)^{1-\beta_1} \mid W_1, \mathcal{G}) + \hat{N}_2 H(\tilde{Y}_{1c}^n, (\tilde{Y}_{1d}^n)^{1-\beta_2} \mid W_1, \mathcal{G}) \\
\leq \hat{N}_1 H(\tilde{Y}_{1c}^n, (\tilde{Y}_{1d}^n)^{1-\beta_2} \mid W_1, \mathcal{G}) + \hat{N}_2 H(\tilde{Y}_{1c}^n, (\tilde{Y}_{1d}^n)^{1-\beta_2} \mid W_1, \mathcal{G}) \\
+ n o \left( \log \hat{P} \right)
\end{align*}
\]  

(204)

(205)

(206)

(207)

(208)

(209)

(210)

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

\[
\hat{U}_1 = (\tilde{Y}_{2a}^c) 
\]  

(211)

\[
\hat{U}_2 = (\tilde{Y}_{1c}^n) 
\]  

(212)

\[
W = (\tilde{Y}_{2b}^c)^{1-\beta_1}, (\tilde{Y}_{1d}^n)^{1-\beta_2}
\]  

(213)

From (116), (204) is concluded as all the \((\lambda_{1i} - \lambda_{2j})^+\) 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 $C_i^{[n]}$ in (230) and (208) follows from sub-modularity properties of the entropy function, i.e., for any $m$ random variables $\{X_1, X_2, \ldots, X_m\}$ where we define $X_{k+m}$ as $X_k$ for positive numbers $k$ we have,

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

if $n \leq m$. (209) is true from the chain rule. (210) is concluded from Theorem 1. From (193), define the set $\mathcal{S}$ continuous functions on $\mathcal{S}$

**Proof of Lemma 5**

Define the set $s_\beta$ 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,

$$S(\beta_1, \beta_2) \overset{\text{def}}{=} \sum_{r=1}^{Q_1} f_r(\beta_1, \beta_2)$$

$$= H(L_{r\delta} g_{rs}(X_a^r \delta a(\beta_1, \beta_2)) \setminus (X_b^r \delta a \beta_1, \beta_2) \setminus (X_c^r \delta a \beta_1, \beta_2), \forall s \in [Q_2])$$

(219)

is a continuous function on the set $s_\beta$ under the conditions specified in the following lemma.

**Lemma 6** $S(\beta_1, \beta_2)$ is a continuous function on the set $s_\beta$ if $f_r(\beta_1, \beta_2)$, $g_{rs}(\beta_1, \beta_2)$ are bounded continuous functions on $s_\beta$ 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 $(\beta_1, \beta_2) = \left(\frac{m}{q}, \frac{e}{q}\right)$ where $m + e \leq q$. From the definition of $Y_{1d}(t)$, $Y_{1d}(t)$, $Y_{2a}(t)$, $Y_{2c}(t)$ and $Y_{2d}(t)$ in (191)-(197) we have,

$$\tilde{N}_1H(Y_2^{[n]} \delta a, Y_2^{[n]} \delta c, Y_2^{[n]} \delta d \mid W_1, G) - n\tilde{N}_0 log \tilde{P}$$

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

If $f(x_1, x_2, \ldots, x_n) \leq A, \forall (x_1, x_2, \ldots, x_n) \in \mathbb{Q}^n$

then $f(x_1, x_2, \ldots, x_n) \leq A, \forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$  \quad (220)

where (220) follows from continuity of function $f(x_1, x_2, \ldots, x_n)$ on $\mathbb{R}^n$. Therefore, proving that a function $f(x_1, x_2, \ldots, x_n)$ on $\mathbb{R}^n$ is bounded by some number $A$ is equivalent to proving that the function is bounded by $A$ for $(x_1, x_2, \ldots, x_n) \in \mathbb{Q}^n$. Thus, proving (142) for the set $\{(\beta_1, \beta_2) \in \mathbb{R}^{2+}; \beta_1 + \beta_2 \leq 1\}$ is equivalent to proving
\[ \hat{N}_1 H (Y_{2a}^{[n]}, Y_{2c}^{[n]}, Y_{2d}^{[n]} | W_1, \mathcal{G}) - n \hat{N}_1 (M - N_1) \beta_1 \log \hat{P} \]  
\[ \leq \hat{N}_1 H (Y_{2a}^{[n]}, Y_{2c}^{[n]}, Y_{2d}^{[n]} | W_1, \mathcal{G}) - \hat{N}_1 H ((\bar{Y}_{2c})_{\beta_1}, (\bar{Y}_{2d})_{\beta_1} | W_1, \mathcal{G}) \]  
\[ \leq \hat{N}_1 H (Y_{2a}^{[n]}, (Y_{2c}^{[n]})_{\beta_1}, (Y_{2d}^{[n]})_{\beta_1} | W_1, \mathcal{G}) \]  
(222)  
(223)  
(224)

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

\[ \bar{Y}_{2c}(t) = \lfloor \bar{Y}_{2b}(t) \rfloor_{0 \to (M - N_2)} \]  
\[ \bar{Y}_{2d}(t) = \lfloor \bar{Y}_{2b}(t) \rfloor_{(M - N_2) \to N_2 - N_1} \]  
(225)  
(226)

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

\[ \hat{N}_1 H (Y_{2a}^{[n]}, (Y_{2c}^{[n]})_{\frac{1}{q}}, (Y_{2d}^{[n]})_{\frac{1}{q}} | W_1, \mathcal{G}) + \hat{N}_2 H (Y_{1c}^{[n]}, (Y_{1d}^{n})_{\frac{1}{q}} | W_1, \mathcal{G}) \]  
\[ \leq (\hat{N}_1 + \hat{N}_2) H (Y_{1}^{[n]} | W_1, \mathcal{G}) + n \circ (\log \hat{P}) \]  
(227)

where the numbers \( \hat{N}_1 \) and \( \hat{N}_2 \) can be rewritten as,

\[ \hat{N}_1 = (M - N_2) \frac{e}{q} \]  
\[ \hat{N}_2 = (N_2 - N_1) \frac{q - m}{q} \]  
(228)  
(229)

Before preceding to proof of (227), let us define the random variables \( C_i^{[n]} \) as the distinct \( \frac{1}{q} \) power levels of \( (Y_{2c}^{[n]})_{\frac{1}{q}} \) and \( (Y_{2d}^{[n]})_{\frac{1}{q}} \), i.e.,

\[ C_i(t) = \begin{cases} 
(Y_{2c}^{[n]})_{\frac{1}{q}}^{i - \frac{1}{q}}(t) & 1 \leq i \leq q \hat{N}_1 \\
(Y_{2d}^{[n]})_{\frac{1}{q}}^{i - \frac{1}{q}}(t) & q \hat{N}_1 < i < q(\hat{N}_1 + \hat{N}_2) \\
\hat{C}_{i-q(\hat{N}_1+\hat{N}_2)}(t) & q(\hat{N}_1 + \hat{N}_2) < i
\end{cases} \]  
(230)

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

\[ \hat{N}_1 H (Y_{2a}^{[n]}, (Y_{2c}^{[n]})_{\frac{1}{q}}, (Y_{2d}^{[n]})_{\frac{1}{q}} | W_1, \mathcal{G}) \]  
\[ \leq \hat{N}_1 H (Y_{2a}^{[n]}, (Y_{2c}^{[n]})_{\frac{1}{q}}, (Y_{2d}^{[n]})_{\frac{1}{q}} | W_1, \mathcal{G}) + \hat{N}_2 H (Y_{1c}^{[n]}, (Y_{1d}^{n})_{\frac{1}{q}} | W_1, \mathcal{G}) \]  
\[ = \hat{N}_1 H (Y_{2a}^{[n]}, (Y_{2c}^{[n]})_{\frac{1}{q}}, (Y_{2d}^{[n]})_{\frac{1}{q}} | W_1, \mathcal{G}) \]  
(231)

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

\[ \hat{N}_1 H (Y_{2c}^{[n]} | W_1, \mathcal{G}) \]  
\[ \leq \hat{N}_1 H (Y_{2c}^{[n]} | W_1, \mathcal{G}) + \hat{N}_2 H ((Y_{1b})_{\beta_2}^{[n]} | Y_{1a}^{[n]}, (Y_{1b}^{[n]})^{1-\beta_2} | W_1, \mathcal{G}) + n \hat{N}_0 \log \hat{P} + n \circ (\log \hat{P}) \]  
(221)
\[ \leq \tilde{N}_1 H(\tilde{Y}[n]_{1c}, (\tilde{Y}_{2c})^{m+c}_q, (\tilde{Y}_{1d})^{1}_q, (\tilde{Y}_{2d})^{m+c}_q | W_1, \mathcal{G}) + \tilde{N}_2 H(\tilde{Y}[n]_{1c}, (\tilde{Y}_{1d})^{1}_q | W_1, \mathcal{G}) \]
\[ + n o (\log \tilde{P}) \]  

(232)

\[= \tilde{N}_1 H((\tilde{Y}[n]_{2c})^{m+c}_q, (\tilde{Y}_{2d})^{1}_q | \tilde{Y}[n]_{1c}, (\tilde{Y}_{1d})^{1}_q, W_1, \mathcal{G}) + (\tilde{N}_1 + \tilde{N}_2) H(\tilde{Y}[n]_{1c}, (\tilde{Y}_{1d})^{1}_q | W_1, \mathcal{G}) \]
\[+ n o (\log \tilde{P}) \]  

(233)

\[= (M - N_2) \frac{e}{q} H(C[n]_1, C[n]_2, \ldots, C[n]_{(M-N_2)+e} \mid (\tilde{Y}[n]_{1c}, (\tilde{Y}_{1d})^{1}_q, W_1, \mathcal{G}) \]
\[+ ((M - N_2) \frac{e}{q} + (N_2 - N_1) \frac{q - m}{q}) H(\tilde{Y}[n]_{1c}, (\tilde{Y}_{1d})^{1}_q | W_1, \mathcal{G}) + n o (\log \tilde{P}) \]  

(234)

\[\leq \frac{1}{q} \sum_{j=1}^{(M - N_2)e + (N_2 - N_1)(q - m)} H(C[n]_{i_1}, C[n]_{i_2}, \ldots, C[n]_{(M - N_2)e} \mid (\tilde{Y}[n]_{1c}, (\tilde{Y}_{1d})^{1}_q, W_1, \mathcal{G}) \]
\[+ ((M - N_2) \frac{e}{q} + (N_2 - N_1) \frac{q - m}{q}) H(\tilde{Y}[n]_{1c}, (\tilde{Y}_{1d})^{1}_q | W_1, \mathcal{G}) + n o (\log \tilde{P}) \]  

(235)

\[= \frac{1}{q} \sum_{j=1}^{(M - N_2)e + (N_2 - N_1)(q - m)} H(\bar{C}[n]_{i_j}, \bar{C}[n]_{i_{j+1}}, \ldots, \bar{C}[n]_{i_j + (M - N_2)e}, \tilde{Y}[n]_{1c}, (\tilde{Y}_{1d})^{1}_q | W_1, \mathcal{G}) \]
\[+ n o (\log \tilde{P}) \]  

(236)

\[\leq \frac{(M - N_2)e + (N_2 - N_1)(q - m)}{q} H(\bar{Y}[n]_{1c} \mid W_1, \mathcal{G}) + n o (\log \tilde{P}) \]  

(237)

\[= (\tilde{N}_1 + \tilde{N}_2) H(\bar{Y}[n]_{1c} \mid W_1, \mathcal{G}) + n o (\log \tilde{P}) \]  

(238)

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

\[ \bar{U}_1 = \left( \begin{array}{c}
\bar{Y}[n]_{2a}
\bar{Y}[n]_{2c}
\end{array} \right) \]  

(239)

\[ \bar{U}_2 = \left( \begin{array}{c}
\bar{Y}[n]_{1c}
\bar{Y}[n]_{1d}
\end{array} \right) \]  

(240)

\[ W = \left( \begin{array}{c}
(\bar{Y}[n]_{2c})^{m+c}_q
(\bar{Y}_{2d})^{m+c}_q
\end{array} \right) \]  

(241)

From (116), (232) is concluded as all the \((\lambda_1 - \lambda_2)^+\) are zero in the right side of (116). (233) is true from the chain rule, (234) is obtained from the definition of \( C[n]_i \) 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 4 of Theorem 4 in [5], i.e.,

\[ H(\bar{C}[n]_{i_1}, \bar{C}[n]_{i_{j+1}}, \ldots, \bar{C}[n]_{i_j + (M - N_2)e}, \bar{Y}[n]_{1c} \mid W_1, \mathcal{G}) \leq H(\bar{Y}[n]_{1d} \mid \bar{Y}[n]_{1c}, W_1, \mathcal{G}) + n o (\log \tilde{P}) \]  

(242)

From (193), define the random variables \( Z_r(t) \), and \( Z_{r1}(t) \) for all \( r \in [(M - N_2)] \) and \( t \in [n] \) as

\[ Z_r(t) = \bar{y}_{1rd}(t) \]  

(245)

\[ Z_{r1}(t) = (\bar{y}_{1rd}(t))^{1}_q \]  

(246)
\[ Z_{r,m+1}(t) = \tilde{C}_{i_j+(r-1)e+m-1}(t), \forall m \in [e] \]  
\[ \lambda_{ri} \]  
where \( \lambda_{ri} \) is derived for any \( i \in [e+1] \) as  
\[ \lambda_{ri} = \begin{cases} \frac{1}{q} & 0 < i \leq e \\ \frac{q}{e} & i = e + 1 \end{cases} \]  

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)),

\[ H(Z_1^{[n]}, \cdots, Z_{(M-N_2)+(e+1)}^{[n]} \mid W) - H(Z_1^{[n]}, \cdots, Z_{(M-N_2)}^{[n]} \mid W, \mathcal{G}) \leq +n \, o(\log \bar{P}) \]  

**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

\[ Z = L^0_1(X_1, X_2) \]  
\[ Z' = L^0_2(X_1, X_2) \]  
\[ Z_a = (Z)^{1.8}_{0.8} \]  
\[ Z_b = (Z')^{1.8}_{0.8} \]  

First of all, note that we can assume that \( Z_a = (Z)^{1.8}_{0.8} = L^0_1((X_1)^{1.8}_{0.8}, (X_2)^{1.8}_{0.8}) \) which is concluded similar to (70). Let us prove that,

\[ H(Z_a, Z_b) - H(Z \mid \mathcal{G}) \leq o(\log \bar{P}) \]  

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

\[ H((Z)^{1.8}_{0.8}, (Z')^{1.8}_{0.8}) - H(Z \mid \mathcal{G}) \leq o(\log \bar{P}) \]  

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

\[ H(U') - H(U \mid \mathcal{G}) \leq o \, (\log \bar{P}) \]  

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

\[ H(U') - H(U \mid \mathcal{G}) \leq \log \{ \mathbb{E}_\mathcal{G} | \mathcal{S}_\nu(\mathcal{G}) | \} \]  

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

\[ \mathbb{E}_\mathcal{G} \{| \mathcal{S}_\nu(\mathcal{G}) | \} = \sum \lambda \mathbb{P}(\lambda) \]  

where \( \mathbb{P}(\lambda) \) is defined as the probability that \( \lambda \) and \( \nu \) correspond to the same \( U \).
E.1.2 Bounding the Average Size of Aligned Image Sets

Consider the following two cases of max

Now, let us bound \( \mathbb{P}(\lambda \in S_{\nu}) \) 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

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 \( \frac{4f_{\text{max}}}{|E_2 - F_2|} \) if \( E_2 \neq F_2 \). The probability of alignment is bounded by \( \frac{4f_{\text{max}}}{\max(|E_1 - F_1|, |E_2 - F_2|)} \) 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

for some positive constant \( c \) not depending on \( P \) which results in \( H(U') - H(U | \mathcal{G}) \leq \log(c \log P) \). \( Z_i \) is defined as the support of the random variable \( |\lambda_i - \nu_i| \).

Define \( \Delta_j \) as follows

Consider the following two cases of max

Any number \( X \in X_{\delta} \) can be written as \( (X)_{\alpha}^{\delta} + X_{\alpha} \) for any non-negative number \( \alpha \) less than \( \delta \). Thus, we have

Therefore, \( \mathbb{E}_{\mathcal{G}} |S_{\nu}(\mathcal{G})| \) is bounded as follows.

\footnote{From Definition we know that \( Z_1 = \{a : a \in Z, |a| \leq 2\Delta P^{0.8}\}, Z_2 = \{a : a \in Z, |a| \leq 2\Delta P^{0.8}\}.}
Let us first compute the term \( \left| \sum_{\lambda_1 - \nu_1 \in \mathbb{Z}, |\lambda_2 - \nu_2| \in \mathbb{Z}_2} f_{\max} \right| \) is true as the partial sum of harmonic series can be bounded above by logarithmic function. Therefore, we bound the expected value of \( E_b \) as follows

\[
\sum_{\lambda_1 - \nu_1 \in \mathbb{Z}, |\lambda_2 - \nu_2| \in \mathbb{Z}_2} f_{\max} \max(|E_1 - F_1|, |E_2 - F_2|)
\]

where (274) is true as,

\[
|\lambda_1 - \nu_1| = |L_2^g((E_1)_{0.8}^1, (E_2)_{0.8}^1) - L_1^g((F_1)_{0.8}^1, (F_2)_{0.8}^1)| 
\]

\[
\leq 4 + 2\Delta \max(|(E_1)_{0.8}^1 - (F_1)_{0.8}^1, (E_2)_{0.8}^1 - (F_2)_{0.8}^1|) \]

(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,

\[
\sum_{a \in S_a, b \in S_b} f(a, b) \leq |S_a| \sum_{b \in S_b} \max_{a \in S_a} f(a, b) \leq |S_a||S_b| \max_{a \in S_a, b \in S_b} f(a, b) \]

(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 \( E_b \) as follows

\[
\sum_{\lambda_1 - \nu_1 |z_1, |\lambda_2 - \nu_2| \in \mathbb{Z}_2} f_{\max} \max(|E_1 - F_1|, |E_2 - F_2|)
\]

\[
\leq \sum_{|\lambda_2 - \nu_2| \in \mathbb{Z}_2} 8f_{\max}(2 + \Delta) \max(|E_1 - F_1|, |E_2 - F_2|)
\]

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

\[
|\lambda_2 - \nu_2| = |(L_2^g(E_1, E_2))_{0.8} - (L_2^g(F_1, F_2))_{0.8}|
\]

\[
= |L_2^g(E_1, E_2) - \overline{P}^{0.8}[\frac{L_2^g(E_1, E_2)}{P^{0.8}f_1} - L_2^g(F_1, F_2) + \overline{P}^{0.8}\frac{L_2^g(F_1, F_2)}{P^{0.8}}]|
\]

\[
= I_1|L_2^g(E_1, E_2) - L_2^g(F_1, F_2)| + I_2\overline{P}^{0.8}\left(\frac{L_2^g(E_1, E_2)}{P^{0.8}} - \frac{L_2^g(F_1, F_2)}{P^{0.8}}\right)
\]

\[
= I_1|L_2^g(E_1, E_2) - L_2^g(F_1, F_2)| + I_2\overline{P}^{0.8}I_3
\]
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

\[
\sum_{|\lambda_2 - \nu_2| \in \mathbb{Z}_2} 8f_{\text{max}}(2 + \Delta) \leq \sum_{|\lambda_2 - \nu_2| \in \mathbb{Z}_2} 8f_{\text{max}}(2 + \Delta) \leq 16 \Delta f_{\text{max}}(2 + \Delta) 2(4\Delta + 13) \log(2\Delta P^{0.8}) \leq O(\log \bar{P})
\]

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

\[
|L^2_2(E_1, E_2) - L^2_2(F_1, F_2)| \leq 4 + 2\Delta \max(|E_1 - F_1|, |E_2 - F_2|)
\]

(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

\[
\sum_{i=1}^{a} \max(1, |i - b(n)n - c(n)|, 1) \leq a(2M + 5) \log((a + M)n + M)
\]

(290) is true as we count each number at most \( a(2M+5) \) times, e.g., consider the term \( \frac{1}{i} \). \( i \) 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 \( \bar{U}' \) and \( \bar{U} \) as \((Z_1^{[n]}, \cdots, Z_2^{[n]}{M-N_2,(e+1)}) \) and \((Z_1^{[n]}, \cdots, Z_{M-N_2}^{[n]}{M-N_2)} \), respectively. Let us prove,

\[
H(\bar{U}' | W, \mathcal{G}) - H(\bar{U} | W, \mathcal{G}) \leq n o (\log \bar{P})
\]

(291)

We are only interested in the difference of entropies of \( \bar{U}' \) and \( \bar{U} \) conditioned on \( W \) and \( \mathcal{G} \). Similar to the AIS approach in [5], we first claim that from the functional dependence, \( \bar{U} \) can be made a function of \( \bar{U}' \), \( W, \mathcal{G} \). Consider some instance of \( \bar{U}' \), 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 \( \bar{U}' \) which result in the same \( \bar{U} \) as \( \nu^{[n]} \). Since uniform distribution maximizes the entropy,

\[
D_{\Delta} \triangleq H(\bar{U}' | W, \mathcal{G}) - H(\bar{U} | W, \mathcal{G})
\]

\[
\leq H(\bar{U}' | \bar{U}, W, \mathcal{G})
\]

\[
\leq E_{\mathcal{G}} \{\log |\mathcal{S}_{\nu^{[n]}}(W, \mathcal{G})|\}
\]

\[
= E_W \{E_{\mathcal{G}} \{\log |\mathcal{S}_{\nu^{[n]}}(W, \mathcal{G})| | W\}\}
\]

\[
\leq \max_{w \in \mathcal{W}} E_{\mathcal{G}} \{\log |\mathcal{S}_{\nu^{[n]}}(W, \mathcal{G})| | W = w\}
\]

(295)
where $\mathcal{W}$ is defined as the support of the random variable $W$. We are only interested in the difference of entropies of $\bar{U}'$ and $\bar{U}$ conditioned on $W$ and $\mathcal{G}$, i.e., $H(\bar{U}' | W, \mathcal{G}) - H(\bar{U} | W, \mathcal{G})$. Similar to the AIS approach in [5], we start with functional dependence. From the functional dependence argument, without loss of generality $\bar{U}$ can be made a function of $\bar{U}', W, \mathcal{G}$. So, from (296), $E_\mathcal{G}\{|S_{\nu^n}(W, \mathcal{G})| \mid W = w\}$ 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 $\lambda^n$, or in the other words,

$$E_\mathcal{G}\{|S_{\nu^n}(W, \mathcal{G})| \mid W = w\} = \sum_{\lambda^n} \mathbb{P}(\lambda^n)$$

(297)

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

### E.3 Bounding the Probability of Image Alignment

Given $\mathcal{G}$ and $W = w$, consider two distinct instances of $\bar{U}'$ denoted as $\lambda^{[n]} = (\lambda_1^{[n]}, \ldots, \lambda_{(M-N_2)}^{[n]}, \lambda_{e+1}^{[n]})$ and $\nu^{[n]} = (\nu_1^{[n]}, \ldots, \nu_{(M-N_2)}^{[n]}, \nu_{e+1}^{[n]})$ produced by corresponding realizations of codewords $(X_1^n, X_2^n, \ldots, X_M^n)$ denoted by $(E_1^n, E_2^n, \ldots, E_M^n)$ and $(F_1^n, F_2^n, \ldots, F_M^n)$, 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,

$$\lambda_{kl}(t) = \begin{cases} L_{kl}^h(t) \left( (E_{a}(t))_{\frac{1}{q}} \triangledown (E_{c}(t))_{\frac{1}{q+m+c}} \right) & l = 1 \\ L_{kl}^h(t)(\bar{E}_{a}(t))_{\frac{1}{q}} \triangledown \bar{E}_{c}(t)) & 1 < l \leq e + 1 \end{cases}$$

(298)

$$\nu_{kl}(t) = \begin{cases} L_{kl}^h(t) \left( (F_{a}(t))_{\frac{1}{q}} \triangledown (F_{c}(t))_{\frac{1}{q+m+c}} \right) & l = 1 \\ L_{kl}^h(t)(\bar{F}_{a}(t))_{\frac{1}{q}} \triangledown \bar{F}_{c}(t)) & 1 < l \leq e + 1 \end{cases}$$

(299)

where for any $k \in [(M-N_2)]$ we assume that $a_{kl}$ are arbitrary distinct decreasing numbers belonging to the set $\{m + 1, m + 2, \ldots, q\}$, i.e., we assume that for any $2 \leq l < l' \leq e + 1$, $a_{kl} > a_{k'l'}$ and $\{a_{k2}, a_{k3}, \ldots, a_{e,k+1}\} \subseteq \{m + 1, m + 2, \ldots, q\}$. 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,

$$\mathbf{E}(t) = [E_1(t) \ E_2(t) \ \cdots \ E_M(t)]^T$$

(300)

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

(301)

$$\mathbf{E}_c(t) = [\mathbf{E}(t)]_{M-N_2 \rightarrow M-N_1}$$

(302)

$$\mathbf{F}(t) = [F_1(t) \ F_2(t) \ \cdots \ F_M(t)]^T$$

(303)

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

(304)

$$\mathbf{F}_c(t) = [\mathbf{F}(t)]_{M-N_2 \rightarrow M-N_1}$$

(305)

Note that for any $m \in [M]$ and $t \in [n]$, $\bar{E}_m(t), \bar{F}_m(t) \in \{0, 1, \ldots, \bar{P}\}$, see (111). In the next step we bound $\mathbb{P}(\lambda^{[n]} \in S_{\nu^{[n]}})$ from above. We wish to bound the probability that the images of these two
codewords align, or in other words \( \tilde{U}(\lambda^n, W, \mathcal{G}) = \tilde{U}(\nu^n, W, \mathcal{G}) \). Thus, for any \( k \in [K] \) and \( t \in [n] \) we have,

\[
L_k^q(t) \left( E_a(t) \nabla (E_c(t)) \right) = L_k^q(t) \left( F_a(t) \nabla (F_c(t)) \right)
\]

From (306), we have,

\[
\sum_{j=1}^{M-N_2} |g_{kj}(t)E_j(t)| + \sum_{j=N_1+1}^{M} |g_{kj}(t)(E_j(t))^{1/m}_q| = \sum_{j=1}^{M-N_2} |g_{kj}(t)F_j(t)| + \sum_{j=N_1+1}^{M} |g_{kj}(t)(F_j(t))^{1/m}_q| \]

\[
\Rightarrow |\sum_{j=1}^{M-N_2} g_{kj}(t)(F_j(t) - E_j(t)) + \sum_{j=N_1+1}^{M} g_{kj}(t)((F_j(t))^{1/m}_q - (E_j(t))^{1/m}_q)| \leq M
\]

where (308) follows from (307) as for any real number \( x \), \( |x - [x]| < 1 \). For any \( j \in s_M = [1, 2, \ldots, (M - N_2), N_1 + 1, \ldots, M] \), define \( A_j(t) \) as,

\[
A_j(t) = \begin{cases} 
F_j(t) - E_j(t) & 1 \leq j \leq (M - N_2) \\
(F_j(t))^{1/m}_q - (E_j(t))^{1/m}_q & N_1 < j \leq M 
\end{cases}
\]

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

\[
P(\lambda^n \in S_{\nu^n}) \leq \prod_{t=1, \max_{j \in s_M} |A_j(t)| \neq 0} \sum_{k=1}^{K} \prod_{t=1, \max_{j \in s_M} |A_j(t)| \neq 0}^n \frac{2M \max_{j \in s_M} |A_j(t)|}{2M \max_{j \in s_M} |A_j(t)|}
\]

**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,

\[
\sum_{|\lambda_{11} - \nu_{11}| \in \mathbb{Z}_{11}, \ldots, |\lambda_{1,\epsilon+1} - \nu_{1,\epsilon+1}| \in \mathbb{Z}_{1,\epsilon+1}, \max_{j \in s_M} |A_j(t)| \neq 0} \frac{2M \max_{j \in s_M} |A_j(t)|}{2M \max_{j \in s_M} |A_j(t)|} \leq c_1 + c_2 \log P
\]
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 \( \log \log \hat{P} = o(\log \hat{P}) \). \( Z^*_l \) is defined as the set \( \{0\} \cup [M + 1 + [M\Delta \hat{P}^2_{\hat{\eta}}]] \) for \( l = 1 \) and the set \( \{0\} \cup [\hat{P}^2_{\hat{\eta}}] \) for \( l \in [e + 1], l \neq 1 \). We define \( \Delta_l \) as,

\[
\Delta_l = \begin{cases} 
(\hat{E}_j)_{\frac{1}{q}} - (\hat{F}_j)_{\frac{1}{q}} & 1 \leq j \leq M - N_2 \\
(\hat{E}_j)_{\frac{1}{m+\epsilon}} - (\hat{F}_j)_{\frac{1}{m+\epsilon}} & N_1 < j \leq M 
\end{cases} 
\tag{313}
\]

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

1. \( \max_{j \in s_M} |\Delta_j| \geq 2 \)

Any number \( X \in X_\delta \) can be written as \( (X)^{\delta} P^\alpha + X_\alpha \) for any non-negative number \( \alpha \) less than \( \delta \). Thus, when \( 1 \leq j \leq (M - N_2) \), the term \( |E_j - F_j| \) is bounded from below as,

\[
|A_j| = |E_j - F_j| \\
= |\Delta_j P^\xi_{\hat{\eta}} + (E_j)_{\frac{1}{q}} - (F_j)_{\frac{1}{q}}| \\
\geq |\Delta_j P^\xi_{\hat{\eta}}| - |(E_j)_{\frac{1}{q}} - (F_j)_{\frac{1}{q}}| \\
\geq (|\Delta_j| - 1)P^\xi_{\hat{\eta}} 
\tag{314}
\]

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

\[
|A_j| = |(E_j)_{\frac{1}{m}} - (F_j)_{\frac{1}{m}}| \\
= |\Delta_j P^\xi_{\hat{\eta}} + (E_j)_{\frac{1}{q}} - (F_j)_{\frac{1}{q}}| \\
\geq |\Delta_j P^\xi_{\hat{\eta}}| - |(E_j)_{\frac{1}{q}} - (F_j)_{\frac{1}{q}}| \\
\geq (|\Delta_j| - 1)P^\xi_{\hat{\eta}} 
\tag{315}
\]

Moreover, from (298), (299), and (313), the term \( |\lambda_{i1} - \nu_{i1}| \) is bounded from above as follows,

\[
|\lambda_{i1} - \nu_{i1}| < M + M\Delta \max_{j \in s_M} |\Delta_j| 
\tag{316}
\]

The left side of (312) is bounded as,

\[
\sum_{|\lambda_{i1} - \nu_{i1}| \in Z_{i1}, \ldots, |\lambda_{i1,e+1} - \nu_{i1,e+1}| \in Z_{i1,e+1}, \max_{j \in s_M} |A_j| \neq 0} \frac{2Mf_{max}}{\max_{j \in s_M} |A_j|} \\
\leq \sum_{|\lambda_{i1} - \nu_{i1}| \in Z_{i1}, \ldots, |\lambda_{i1,e+1} - \nu_{i1,e+1}| \in Z_{i1,e+1}} \frac{2Mf_{max}}{\max_{j \in s_M} (|\Delta_j| - 1)P^\xi_{\hat{\eta}}} \\
\leq \sum_{|\lambda_{i1} - \nu_{i1}| \in \{0\} \cup [M + M\Delta], |\lambda_{i1} - \nu_{i1}| \in Z_{i1}, \ldots, |\lambda_{i1,e+1} - \nu_{i1,e+1}| \in Z_{i1,e+1}} \frac{2Mf_{max}}{P^\xi_{\hat{\eta}}} \\
+ \sum_{|\lambda_{i1} - \nu_{i1}| \in Z_{i1}, |\lambda_{12} - \nu_{12}| \in Z_{12}, \ldots, |\lambda_{i1,e+1} - \nu_{i1,e+1}| \in Z_{i1,e+1}} \frac{2M^2f_{max}}{(|\lambda_{11} - \nu_{11}| - M - M\Delta)P^\xi_{\hat{\eta}}}
\tag{322}
\]
2. max

Finally, (325) is concluded as the partial sum of harmonic series can be bounded above by logarithmic function i.e.,

where \( Z_{11}^* \) is defined as the set of \( \tilde{Z}_{11} \cap [M + M\Delta]^c \). (322) follows from (317) and (320). Note that \( \max_{j \in s_M} |A_j| \) is positive number as \( \max_{j \in s_M} |\Delta_j| \geq 2 \). (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,

\[
\sum_{a \in S_a, b \in S_b} f(a, b) \leq |S_a| \sum_{a \in S_a} \max_{b \in S_b} f(a, b) \leq |S_a||S_b| \max_{a \in S_a, b \in S_b} f(a, b)
\]

(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. \( \max_{j \in s_M} |\Delta_j| \leq 1 \)

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 \( \hat{\Delta}_{ij} \) as,

\[
\hat{\Delta}_{ij} \triangleq \begin{cases} 
(\tilde{E}_j)^{1 \left\lfloor \frac{j}{q} \right\rfloor} - (\tilde{F}_j)^{1 \left\lfloor \frac{j}{q} \right\rfloor} & 1 \leq j \leq M - N_2 \\
(\tilde{E}_j)^{1 \left\lfloor \frac{j}{q} \right\rfloor} - (\tilde{F}_j)^{1 \left\lfloor \frac{j}{q} \right\rfloor} & N_1 < j \leq M
\end{cases}
\]

(327)

and define the number \( l^* \) as the smallest integer from the set \( \{2, 3, \ldots, e + 1\} \) where

\[
\max_{j \in \{N_1+1, \ldots, M\}} |\hat{\Delta}_{l,j}| \geq 2
\]

(328)

Consider the following two cases.

(a) \( l^* \) doesn’t exist.

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

\[
|L_{kl}(t)(\overline{E}_a(t))^{\frac{1}{q}} \nabla \overline{E}_c(t)) - L_{kl}(t)(\overline{F}_a(t))^{\frac{1}{q}} \nabla \overline{F}_c(t))| \leq 2M\Delta P^{\frac{e + m - l + 1}{q}}
\]

(329)

\[
\leq 2M\Delta P^{\frac{e + m - l - 1}{q}}
\]

(330)

Thus, each of the variables \( |\lambda_{l} - \nu_{l}| \) is bounded by \( 2M\Delta + 1 \) for any \( l \in \{2, \ldots, e + 1\} \). 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) \( 2 \leq l^* \leq e + 1 \).

Similar to (317) and (320), when \( 1 \leq j \leq (M - N_2) \), the term \(|E_j - F_j|\) is bounded from below as,
\[
|A_j| = |E_j - F_j| \\
= |\Delta_{t^*} \hat{P} \frac{e^{-l^*}}{q} + (E_j)_{\frac{e^{-l^*}}{q}} - (F_j)_{\frac{e^{-l^*}}{q}}| \\
\geq (|\Delta_{t^*} - 1|) \hat{P} \frac{e^{-l^*}}{q} \tag{331}
\]
and if \( N_1 < j \leq M \), the term \(|(E_j)_{\frac{1}{2}} - (F_j)_{\frac{1}{2}}|\) is bounded from below as,
\[
|A_j| = |(E_j)_{\frac{1}{2}} - (F_j)_{\frac{1}{2}}| \\
= |\Delta_{t^*} \hat{P} \frac{e^{-l^*}}{q} + (E_j)_{\frac{e^{-l^*}}{q}} - (F_j)_{\frac{e^{-l^*}}{q}}| \\
\geq (|\Delta_{t^*} - 1|) \hat{P} \frac{e^{-l^*}}{q} \tag{333}
\]

The left side of (312) is bounded as,
\[
\sum_{|\lambda_{11} - \nu_{11}| \in \{0\} \cup [M + M\Delta]} \sum_{|\lambda_{12} - \nu_{12}| \in \mathbb{Z}_{12} \cdots, |\lambda_{1, e + 1} - \nu_{1, e + 1}| \in \mathbb{Z}_{1, e + 1}, \max_{j \in s_M} |A_j| \neq 0} \frac{2Mf_{\max}}{\max_{j \in s_M} |A_j|} \\
\leq (M + 1 + M\Delta)(2M\Delta + 1)^{t^* - 2} \frac{(1 + \hat{P})_{\frac{1}{2}} e^{-l^* + 1}}{\hat{P} \frac{e^{-l^*}}{q}} \tag{335}
\]
\[
\leq (M + 1 + M\Delta)(2M\Delta + 1)^{t^* - 2} \frac{(1 + \hat{P})_{\frac{1}{2}} e^{-l^* + 1}}{\hat{P} \frac{e^{-l^*}}{q}} \times (2M^2(1 + \Delta)f_{\max} + \sum_{|\lambda_{11} - \nu_{11}| \in \mathbb{Z}_{11}} \frac{2M^2\Delta f_{\max}}{\max_{j \in s_M} |A_j|}) \tag{336}
\]
\[
\leq 2e^{e+1}(M + 1 + M\Delta)(2M\Delta + 1)^{t^* - 2}M^2f_{\max}(1 + \Delta)(2 + \log \hat{P}_{\frac{1}{2}}) \tag{337}
\]
\[
\leq c_1 + c_2 \log \hat{P} \tag{338}
\]

for some constants \( c_1 \) and \( c_2 \) not depending on \( \hat{P} \). Note that, \( \hat{Z}_{11^*} \) is defined as the set \( \hat{Z}_{11^*} \cap [M + M\Delta]^c \) and (336) follows similar to (323).

**F Proof of Lemma [6]**

As the functions \( g_{rs}(\beta_1, \beta_2) \) are bounded continuous functions on \( s_\beta \) for any \( r \in [Q_1], s \in [Q_2], l \in \{a, b, c, d\} \) for any positive real number \( \epsilon \) there exists some \( \delta \) that if \( |\beta_1 - \beta_1'| < \delta, |\beta_2 - \beta_2'| < \delta \) then
\[
|g_{rs}(\beta_1, \beta_2) - g_{rs}(\beta_1', \beta_2')| \leq \epsilon, \forall r \in [Q_1], s \in [Q_2], l \in \{a, b, c, d\} \tag{339}
\]
Therefore, from Lemma 3 we have,
\[
\begin{align*}
|H(L_g((X^n_a)_{g_{rs}}(\beta_1,\beta_2) \triangledown (X^n_b)_{g_{rsb}}(\beta_1,\beta_2) \triangledown (X^n_c)_{g_{rsc}}(\beta_1,\beta_2)), \forall s \in [Q_2]) & - H(L_g((X^n_a)_{g_{rs}}(\beta'_1,\beta'_2) \triangledown (X^n_b)_{g_{rsb}}(\beta'_1,\beta'_2) \triangledown (X^n_c)_{g_{rsc}}(\beta'_1,\beta'_2)), \forall s \in [Q_2])| \\
\leq nM \epsilon \log \bar{P} + n o (\log \bar{P})
\end{align*}
\]
(340)

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

References


