Interference Alignment and a Noisy Interference Regime for Many-to-One Interference Channels

Viveck R. Cadambe, Syed A. Jafar

Center for Pervasive Communications and Computing
Electrical Engineering and Computer Science
University of California Irvine,
Irvine, California, 92697, USA
Email: vcadambe@uci.edu, syed@uci.edu

Abstract

We study the capacity of discrete memoryless many-to-one interference channels, i.e., $K$ user interference channels where only one receiver faces interference. For a class of many-to-one interference channels, we identify a noisy interference regime, i.e., a regime where random coding and treating interference as noise achieves sum-capacity. Specializing our results to the Gaussian MIMO many-to-one interference channel, which is a special case of the class of channels considered, we obtain new capacity results. Firstly, while previous results characterized noisy interference regimes for many-to-one interference channels with inputs having average power constraints, we show that this remains valid for a more general class of inputs. This more general class of inputs includes the practical scenario of the inputs being restricted to fixed finite-size constellations such as PSK or QAM. Secondly, we extend noisy interference results previously studied in interference channels with single antenna nodes at all transmitters, to MIMO and parallel many-to-one interference channels. Finally, while previous results considered the Gaussian interference channel with full channel state information (CSI) at all nodes, we provide a noisy interference regime for fading Gaussian many-to-one interference channels without CSI at the transmitters.

While the many-to-one interference channel requires interference alignment, which in turn requires structured codes in general, we argue that in the noisy interference regime, interference is implicitly aligned by random coding irrespective of the input distribution. As a byproduct of our study, we identify a second class of many-to-one interference channels (albeit deterministic) where random coding is optimal (though interference is not treated as noise). We attribute the optimality of random coding in this second class of channels to the resolvability of the multiple interferers at the receiver which precludes the possibility of interference alignment and hence obviates the need of structured codes.

I. Introduction

The idea of interference alignment has been recently discovered to play a significant role in the characterization of capacity of wireless interference networks [1]–[4]. Interference alignment is the idea that signals are designed so that they overlap at receivers where they cause interference while remaining distinguishable at receivers where they are desired. In network communication scenarios where receivers face interference from multiple sources, alignment compacts the space occupied by the multiple interfering signals and results in increased rates for the messages desired at the receiver. Therefore, optimal code design for interference channels typically involves a conflict between the need for interference management via alignment at undesired receivers, and the need to maximize rates at the desired receiver. This conflict is clearly reflected in the contrast between interference channels, and channels that do not require alignment, viz. point-to-point, multiple access (MAC) and broadcast (BC) channels. For instance, it is well known that the identically and independently distributed (i.i.d.) circularly symmetric Gaussian distributions on the inputs, which maximizes the differential entropy of the received signals, achieve capacity in the point to point, MAC and BC channels. However, in contrast, in interference channels, asymmetric complex signaling [5], and structured (lattice) codes [4], [6]–[9] have been shown to be useful, especially because they align interference. In fact, the lack of a complete understanding of the limits of interference alignment is among the primary hurdles in capacity characterizations of wireless networks. In this paper, we will provide a finer understanding of interference alignment, and characterize the sum-capacity of a class of discrete memoryless (many-to-one) interference channels.

The results demonstrating the need for explicit interference alignment via lattice-coding/asymmetric complex signaling contrast with the “noisy interference” results for the Gaussian $K$ user interference channel found recently,
Noisy interference condition:

\[ H_{12}X_2 + H_{13}X_3 + Z_1 \]

is a degraded version of \((Y_2, Y_3)\).

Fig. 1. The 3-user Gaussian Many-to-one Interference Channel and the Noisy Interference Condition.

presented in references [10]–[12]. These references showed that for the \(K\) user interference channel, if the channel gains satisfy certain conditions, then, using random codebooks with circularly symmetric Gaussian distributions for all messages and treating interference as noise at all receivers is sum-capacity optimal. Put differently, these results indicate that in certain scenarios, explicit alignment in the form of structured (lattice) coding or asymmetric signaling is not necessary, and random coding is optimal, even though there is potential for alignment with receivers facing multiple interferers. One of the main goals of this work is a better understanding of why random (Gaussian) coding is optimal in the noisy interference regime, in spite of the opportunity for alignment. We study this question in the setting of the discrete memoryless many-to-one interference channel - the interference channel where only one receiver faces interference - which is the simplest setting where a receiver faces multiple interferers. The main result of this work is the characterization of a noisy interference regime for a class of discrete memoryless many-to-one interference channels. The noisy interference condition obtained here can be loosely described as follows: In the many-to-one interference channel, if the effective interference (with noise) seen by the only receiver facing interference is a stochastically degraded version of the set of received signals at all other receivers, then, random coding at all the transmitters and interference being treated as noise at the receiver facing interference achieves sum-capacity.

The above result, which will be expressed rigorously later (Section III), holds for a broad class of discrete memoryless many-to-one interference channels including the Gaussian many-to-one interference channel. From the perspective of Gaussian interference channels, we make two observations. Firstly, our main result captures the noisy interference regime for the single-antenna Gaussian many-to-one interference channel found previously in references [10]–[12]. Secondly, while results of references [10]–[12] are mainly restricted to the single-antenna Gaussian interference channels with classical assumptions on the model, such as full channel state information (CSI) at all nodes and average power constraints on the inputs, our result described above holds for a broader class of discrete memoryless many-to-one interference channels, and is therefore more robust to the system model. Specializing our main result to the Gaussian setting enables us to extend the noisy interference results to scenarios of practical importance not captured by such classical assumptions and thus not previously considered. Before summarizing such extensions, we first describe how our main result summarized (in italics) above captures the noisy interference regime for many-to-one interference channels previously discovered in [10]–[12].

Consider a \(K\) user Gaussian many-to-one interference channel (Fig. 1), whose inputs and outputs can be expressed as

\[
Y_\tau = H_{ii}(\tau)X_i(\tau) + Z(\tau), \quad i = 2, 3, \ldots, K
\]

\[
Y_1(\tau) = \sum_{i=1}^{K} H_{1i}(\tau)X_i(\tau) + Z_1(\tau), \quad i \in K_1
\]

where, corresponding to the \(\tau\)th symbol, \(X_i(\tau)\) is the complex scalar input at Transmitter \(i\), \(Y_\tau(\tau)\) is the complex
scalar output at Receiver $j$ and $Z_j(\tau)$ represents the zero-mean unit-variance circularly symmetric additive white Gaussian noise (AWGN) variable at Receiver $j$. $H_{ji}$ is a complex scalar representing the channel gain between Transmitter $i$ and Receiver $j$. As is standard in the interference channel, Transmitter $i$ has a message to Receiver $j$, which is independent of the message at, and unknown to other transmitters (and unknown to all receivers prior to communication). Then, in this channel, with an average power constraint on the inputs, it is shown in [10]–[12], that if

$$\sum_{j=2}^{K} \frac{|H_{ji}|^2}{|H_{jj}|^2} \leq 1,$$  

(1)

then, circularly symmetric Gaussian inputs and treating interference as noise at all receivers is sum-capacity optimal. We note that the condition in (1) is a special case of the conditions stated in our main result above (in italics), i.e., when (1) holds, the effective interference at Receiver 1, $V = \sum_{j=2}^{K} H_{1j} X_j + Z_1$, is a stochastically degraded version of the set of signals received at all other receivers, i.e., $(Y_2, Y_3, \ldots, Y_K)$. This is because $V$ is a degraded version of $\sum_{i=2}^{K} \frac{H_{ij}}{H_{jj}} Y_j$, which is, obviously, a degraded version of $(Y_2, Y_3, \ldots, Y_K)$. Thus, we have shown that, the noisy interference regime of [10]–[12] is included in the noisy interference regime found in our main result. In Appendix B we show that, the two regimes - the regime described in (1) and the regime described by our main result - are in fact equivalent. While our results are equivalent to the results of previous works in the context of the classical single-antenna Gaussian many-to-one interference channels, as mentioned earlier, our result holds for a more general class of Gaussian many-to-one interference channels. Specifically, our results extend the noisy interference regime to many-to-one interference channels beyond the classical assumptions. We summarize such extensions below.

- Previous works [10]–[12] consider Gaussian interference channels where the input alphabet is continuous and there is an average power constraint on the input codewords. For these channels, the references show that for certain values of channel gains, random Gaussian codebooks and treating interference as noise is optimal when there is an average power constraint on the input codewords. In practice, however, input signals are typically restricted to fixed finite-size constellations such as PSK, QAM etc. It is not clear whether the noisy interference results results carry forward to the more practical setting of the inputs being constrained to fixed constellations. In fact, there remained open the question of whether there even exists a non-trivial set of channel gains where random coding and treating interference as noise achieves sum-capacity in this setting. In this work, we settle this open question by showing that the noisy interference regime remains valid in the Gaussian many-to-one interference channel even if the inputs are restricted to fixed constellations. In other words, if the channel gains satisfy the conditions of (1), then, random codebooks generated i.i.d with the appropriate distribution at the inputs, and treating interference as noise at the receiver facing interference achieves sum-capacity - even if the inputs are restricted to fixed constellations. Therefore, in the noisy interference regime, the capacity characterization problem is essentially reduced to the problem of determining the optimal single-letter distribution on the inputs.

- The results of [10]–[12] are for interference channels with a single antenna at each node - the question of the existence and characterization of noisy interference regimes for MIMO interference channels remains open. In this work, we (partially) address this open question by characterizing a noisy interference regime for the MIMO Gaussian many-to-one interference channel. Note that extensions of [10]–[12] have been proposed to two user MIMO interference channels [13], [14]. Our result differs from the result of [13], [14] in that, we present a noisy interference regime for the $K$-user interference channel, albeit not fully connected (since we only consider the many-to-one interference channel). It must be noted that the noisy interference regime for the MIMO setting also remains valid for input signals being restricted to finite constellations.

- Previous noisy interference results are presented for the case where the channel is constant (i.e., not fading), and when transmitters and receivers have channel state information (CSI). In this paper, we obtain a noisy interference regime for the fading Gaussian many-to-one interference channel where transmitters do not have CSI, and only the receivers have CSI.

1We have so far only discussed the optimality of random coding and treating interference as noise in our main result. References [10]–[12] also show the optimality using the circularly symmetric Gaussian distribution in the noisy interference regime; this optimality will be shown for Gaussian channels in our characterization as well in a formal description of our result in Section III.
Previous results [15], [16] have shown that parallel (i.e. multi-carrier) Gaussian interference channels (including many-to-one interference channels), unlike point-to-point, multiple access and broadcast channels, are in general inseparable, i.e., joint coding over the multiple carriers is required to achieve sum-capacity in parallel interference channels. While parallel interference channels are in general inseparable, under certain special conditions, they are separable, i.e., separate (independent) coding over the various carriers (and in fact, treating interference as noise) achieves sum-capacity. Such conditions have been identified for parallel single-antenna Z interference channels in [17] and for the (fully-connected) 2-user Gaussian interference channels in [18]. In this paper, we extend the results of [17] to Gaussian MIMO many-to-one interference channels. In particular, we show that, under the special case that the many-to-one interference channels formed over each of the carriers forming the parallel channel satisfies our noisy interference conditions, the channel is separable from a sum-capacity perspective. For example, in the single-antenna Gaussian many-to-one interference channel, if the channel gains on each of the carriers satisfy the condition of (1), then separate random coding and treating interference as noise achieves sum-capacity. Therefore, in this case, with an average power constraint on the input, the sum-capacity of the parallel many-to-one interference channel is the sum of the capacities of the various individual carriers under an optimal power allocation - much like the point-to-point, MAC and BC channels. Further, this separability result is not limited to the average power constraint on the inputs, and holds even for inputs of fixed finite constellations. It must be noted that our main result automatically implies that random coding and treating interference as noise over such a channel (where each carrier satisfies our noisy interference criterion) achieves sum-capacity, because the required degradedness condition holds for the parallel channel. But our main result described above does not, however, imply their separability - the property that the optimal distribution used in random coding has the input over each carrier independent of the input of the other carriers. The separability is an additional result shown in Section III-B.

Why is explicit interference alignment not required in noisy interference regimes?

An important insight to emerge from this work is that in the noisy interference regime, interference is aligned implicitly via random codes. The idea of interference alignment with random codes can be understood in the following setting. Consider a receiver receiving multiple signals coded from a codebook generated in the classical random coding fashion. If the cardinalities of the codebooks corresponding to these signals lie in the achievable random coding rate region (with the corresponding input distributions) of the multiple access channel formed at the receiver, then the receiver can resolve these multiple signals with high probability. In other words, the signals are not aligned. On the other hand, if the cardinalities of the codebooks lie outside this achieved rate region of the multiple access channel formed at the receiver, then the signals align. In fact, in this scenario, the signals cannot be resolved uniquely at the receiver, because the signals align. While alignment is not a desirable phenomenon if the receiver intends to resolve the signals as is the case in the multiple access channel, it is beneficial if the signals are interfering at the receiver as is the case in interference channels. In the noisy interference regime for the many-to-one interference channel, we show that because of a degraded nature of the channel, interference can be aligned with random codes for any distribution on the inputs. In particular, in the Gaussian channel, interference is aligned, even with random coding and with the circularly symmetric Gaussian distribution; thus, alignment is implicit in this case.

It must be noted that the optimality of random coding in the noisy interference regime is desirable from two perspectives. First, the generality of random coding argument enables us to present results for a fairly broad class of channels which may or may not be linear (though we later specialize our results to the linear Gaussian setting). Secondly, the optimality of random codes enables a single-letter characterization for the capacity, unlike in channels which need structured codes, where single-letter characterizations may not even exist [19].

The idea of implicit interference alignment in the noisy interference regime is examined more closely in Section IV by specializing the noisy interference regime to a deterministic many-to-one interference channel. In this setting, we observe that if a random code transmits at sufficiently high rates, then the interference becomes noisy and the extent of alignment via random codes is optimal. The simpler setting of the deterministic channel, apart from enabling a better understanding of the idea of alignment via random coding, allows two other interesting insights into interference alignment. Firstly, we find that random coding achieves capacity in a scenario where the multiple interferers are resolvable at the receiver facing interference (The idea of resolvable interference has been earlier used to determine the capacity of a class of symmetric deterministic interference channels in [20]). The resolvability
The rate of the code is the tuple $(R_1, R_2, \ldots, R_K)$, where $R_i = \log |W_i|$, with $|W_i|$ denoting the cardinality of the message set $W_i$. A rate-tuple $\tilde{R}$ is said to be achievable if there exists a sequence of codes, all of rate $\tilde{R}$, of interference precludes the possibility of interference alignment which enables a characterization of its capacity region. Secondly, a combination of insights from the noisy interference regime and the resolvable interference regime enables us to provide, in Section III-C a (partial) answer to the question: How many bits of additional rate can interference alignment provide on the many-to-one interference channel?

We now proceed to the next section where we formally define the discrete memoryless many-to-one interference channel - the basic setting of all the results of this paper.

II. SYSTEM MODEL: A CLASS OF DISCRETE MEMORYLESS MANY-TO-ONE INTERFERENCE CHANNELS

The $K$ user discrete memoryless many-to-one interference channel (Figure 2) is defined by a set of $K$ inputs $X_i \in \mathcal{X}_i$, and a set of $K$ outputs $Y_i \in \mathcal{Y}_i$ for $i = 1, 2, \ldots, K$. In the class of many-to-one interference channels considered, the outputs $Y_i, i = 2, 3, \ldots, K$ are generated using the distributions $p_{Y_i|X_i}$. The output $Y_1$ is generated as

$$Y_1 = f_1(X_1, V),$$

where $V \in \mathcal{V}$ is generated using $p_{V|X_2, X_3, \ldots, X_K}$. We assume that $V$ is invertible from $(Y_1, X_1)$, i.e., there exists a function $f_1^{-1}$ such that

$$V = f_1^{-1}(Y_1, X_1).$$

There are $K$ independent messages, with message $W_i \in W_i$ generated at source $i \in \{1, 2, \ldots, K\}$, with each message being uniformly distributed over the corresponding message set. A code of length $T$ symbols consists of encoding functions (or equivalently, codebooks) $\phi_i : \mathcal{W}_i \rightarrow \mathcal{X}_i^T$ and decoding functions $\psi_i : \mathcal{Y}_i^T \rightarrow \mathcal{W}_i$ for all $i \in \{1, 2, \ldots, K\}$. It is assumed that all the codebooks, i.e., all the mappings $\phi_i, i = 1, 2, \ldots, K$, are known to all the decoders. We restrict our study to channels and constraints on codewords which ensure that one of the following two sets of quantities exist

- $H(Y_i^T)$ and $H(Y_i^T|X_i^T)$ exist for $i = 1, 2, \ldots, K$. Note that using $i = 1$, and (2), this automatically implies that $H(V^T)$ exists. Also note that the this condition captures all channels where the alphabets $\mathcal{X}_i, \mathcal{Y}_i$ are finite.
- $h(Y_i^T)$ and $h(Y_i^T|X_i^T)$ exist for $i = 1, 2, \ldots, K$. Note that using $i = 1$, and (2), this automatically implies that $h(V^T)$ exists.

The average probability of error of the code $P_e^{(T)}$ is defined to be the probability that the set of decoded messages is not identical to the set of encoded messages, i.e.,

$$P_e^{(T)} \triangleq \Pr \left( \{ \exists i \in \{1, 2, \ldots, K\} | \psi_i(Y_i^T) \neq W_i \} \right).$$

The rate of the code is the tuple $\tilde{R} = (R_1, R_2, \ldots, R_K)$, where $R_i = \log |W_i|$, with $|W_i|$ denoting the cardinality of the message set $W_i$. A rate-tuple $\tilde{R}$ is said to be achievable if there exists a sequence of codes, all of rate $\tilde{R}$.
such that average probability of error vanishes asymptotically, as the sequence index increases. Let \( C \) be the closure of the set of all achievable rate tuples. The sum-capacity \( C_\Sigma \) of the interference channel is defined as

\[
C_\Sigma \triangleq \max_{(R_1, R_2, \ldots, R_K) \in C} \sum_{i=1}^{K} R_i.
\]

**Notation:** We use the notation \( A^T \) to denote \( (A(1), A(2), \ldots, A(T)) \in A^T \) for any random variable \( A \). The calligraphic notation is used to indicate sets. The notation \( N(\mu, \Lambda) \) is used to indicate a circularly symmetric complex Gaussian random vector with mean \( \mu \) and covariance matrix \( \Lambda \). \( I_N \) is used to denote the \( N \times N \) identity matrix.

The following quantities are also used in the paper.

\( K_1 = \{2, 3, \ldots, K\} \)

\( K = K_1 \cup \{1\} \)

\( Y_A = \{Y_i, i \in A\}, A \subseteq K. \)

Before we proceed, the following points must be noted.

- The channel studied here is a natural adaptation, to the many-to-one interference channel setting, of the 2 user interference channel studied in [21].
- In the special case where all the alphabets \( X_i, Y_i, V \) are finite for all \( i \), and all the distribution functions are deterministic (i.e. \( H(Y_i|X_i) = H(V|X_2, X_3, \ldots, X_K) = 0 \) for all \( i = 2, \ldots, K \) irrespective of the input distribution), the channel is an adaptation, to the many-to-one interference channel setting, of the class 2 user deterministic interference channels studied by El Gamal and Costa in [22].
- As we describe next, the MIMO Gaussian many-to-one interference channel is a special case of the class of channels described above.

**The MIMO Gaussian Many-to-One Interference Channel**

Consider the MIMO Gaussian interference channel with \( M_i \) antennas at Transmitter \( i \) and \( N_i \) antennas at Receiver \( i \) so that \( X_i \subseteq \mathbb{C}^{M_i} \), \( Y_i = \mathbb{C}^{N_i} \)

\[
Y_i(\tau) = H_{ii}X_i(\tau) + Z_i(\tau), i \in K_1
\]

\[
Y_1(\tau) = \sum_{i \in K} H_{i1}X_i(\tau) + Z_1(\tau), i \in K_1
\]

where, corresponding to the \( \tau \)th symbol, \( X_i \) is the \( M_i \times 1 \) vector representing the input at Transmitter \( i \), \( Y_i, Z_i \) are the \( N_i \times 1 \) vectors representing the output and the additive white Gaussian noise vectors at Receiver \( i \). We assume that all the noise vectors are circularly symmetric with zero mean and a covariance matrix of identity. This channel can be reduced to the channel defined previously (Figure 2) if we set

\[
V = \sum_{i=2}^{K} H_{i1}X_i + Z_1,
\]

\[
f(X_1, V) = H_{11}X_1 + V,
\]

where \( V \) is a \( N_1 \times 1 \) vector. Note that since the constraints on the inputs \( X_i \) are fairly general, we capture most scenarios of interest such as inputs from a finite fixed-size constellation (Example: PSK, QAM) and inputs with an average and power constraints on the codeword. Further, for the special case where \( X_i = \mathbb{C}^{M_i} \) and an average power constraint is imposed on the input codewords, we will give an explicit expression of the capacity of the channel in the noisy interference regime (to be defined later) in terms of the powers \( P_i \), where

\[
E \left[ \frac{1}{T} \sum_{\tau=1}^{T} ||X_i(\tau)||^2 \right] \leq P_i,
\]

and \( T \) denotes the length of the codeword.
III. A NOISY INTERFERENCE REGIME FOR MANY-TO-ONE INTERFERENCE CHANNELS

Before we proceed to the main result, we introduce a lemma which is useful in the proofs.

Lemma 1: Consider random sequences $A^T, B^T, C^T$ such that $A \in A$, $B \in B$, $C \in C$, and $B^T, C^T$ are generated as $p(B^T, C^T, A^T) = p(A^T) \prod_{\tau=1}^{T} p(B(\tau)|A(\tau))p(C(\tau)|B(\tau))$, where $p(B(\tau)|A(\tau)) = p_{B|A}$ and $p(C(\tau)|B(\tau)) = p_{C|B}$, for all $\tau = 1, 2, \ldots, T$. Note that the sequence $A^T$ does not have to be generated in an i.i.d fashion. If $h(B^T), h(C^T)$ exist, then

$$H(B^T) - H(C^T) \leq \sum_{\tau=1}^{T} (H(B(\tau)) - H(C(\tau)))$$

If $h(B^T), h(C^T)$ exist, then

$$h(B^T) - h(C^T) \leq \sum_{\tau=1}^{T} (h(B(\tau)) - h(C(\tau)))$$

Further, suppose $A = C^M$, $p_{B|A} \sim N(0, \Lambda_1)$ and $p_{C|B} \sim N(0, \Lambda_2)$, or equivalently, let there be variables $Z_i, i = 1, 2$ generated i.i.d according to $Z_i \sim N(0, \Lambda_i), i \in \{1, 2\}$, and $\forall \tau \in \{1, 2, \ldots, T\}$

$$B(\tau) = A(\tau) + Z_1(\tau)$$
$$C(\tau) = B(\tau) + Z_2(\tau)$$

Also, let consider a covariance matrix constraint on the sequence $A^T$, i.e.,

$$E \left[ \frac{1}{T} \sum_{\tau=1}^{T} A(\tau)A(\tau)\right] \leq \Gamma$$

for some covariance matrix $\Gamma$, then i.i.d circularly symmetric Gaussian distribution on $A^T$ maximizes the quantity $h(B^T) - h(C^T)$.

The proof of the lemma is placed in Appendix III. The reader may note that a special case of the above lemma, where $A = B = C = \mathbb{C}$ and $p_{B|A}, p_{C|B}$ are both Gaussian, is used in showing previous noisy interference results for SISO interference channels [10]–[12]. We now present a noisy interference regime for the discrete memoryless many-to-one interference channel considered in this paper.

Theorem 1: In the many-to-one interference channel defined previously, if $p_{Y_i|X_{k_i}}$ is a degraded form of $p_{Y_1|X_1} = \prod_{i \in K_1} p_{Y_i|X_i}$, then, its sum-capacity can be achieved with random coding and treating interference as noise, and it can be expressed as

$$C_{\Sigma} = \max_{i \in K} I(\hat{X}_i; Y_i),$$

(5)

where the maximization is carried over all probability distributions on the input which factorize as $\prod_{i=1}^{K} p_{X_i}$.

The theorem is proved in Appendix III. We now apply the above result to the Gaussian setting below.

Corollary 1: Consider the Gaussian many-to-one interference channel as defined by (2)–(4). If, for $i \in K_1$, there exist covariance matrices $\Lambda_i$ so that

- $\sum_{i \in K_1} \Lambda_i \prec I_{N_i}$, and
- $U_i \overset{\text{i.d.}}{=} H_{1i}X_i + \tilde{Z}_i$ is a stochastically degraded version of $Y_i$, where $\tilde{Z}_i \sim N(0, \Lambda_i), then, a single-letter input distribution on inputs with Receiver 1 treating all interference as noise is sum-capacity optimal. Further, if $\mathcal{X}_i = \mathbb{C}^{M_i}$ with a power constraint $P_i$ on the input codeword $X_i^T$, then the optimal input distribution on the input is Gaussian, i.e., the capacity of the channel can be expressed as

$$C_{\Sigma} = \max_{i \in K_1, p(\Gamma_i) \leq P_i, i \in K} \sum_{i=1}^{K} \log \left( \frac{\det \left( I_{N_i} + \sum_{k \in K} H_{ik} \Gamma_k H_{ik}^H \right)}{\det \left( I_{N_i} + \sum_{k \in K \setminus \{i\}} H_{ik} \Gamma_k H_{ik}^H \right)} \right)$$
with \( H_{ij} = 0 \) if \( i \notin \{1, j\} \).

The proof is almost identical to Theorem 1 on noting that \( V \) is a degraded version of \( U_{K_1} = (U_2, U_3, \ldots, U_K) \), which is in turn, a degraded version of \( Y_{K_1} \), effectively implying that \( V \) is a degraded version of \( Y_{K_1} \). For completeness, we provide the proof in Appendix [V].

Remark: In this case, if the inputs come from a finite constellation such as PSK/QAM, then the capacity characterization essentially involves determination of the optimal input distribution which maximizes [5]. Unlike the point-to-point channel, it is not clear even for symmetric constellations such as BPSK, whether the optimal distribution is uniform on the inputs.

Remark: The noisy interference condition of Theorem 1 and the corresponding sum-capacity characterization remains unchanged, even if Transmitter \( i \in K_1 \) each had an independent message for Receiver 1, along with the usual message for Receiver \( i, \) so that there are \( 2K - 1 \) messages in the system - i.e., even if each link of the many-to-one channel carried a message. In this channel, with the channel satisfying the conditions of the theorem, all the messages to Receiver 1 from Transmitter \( i \neq 1 \) will be set to null so that the channel operates as an interference channel, for sum-capacity. The proof is almost identical as for the interference channel, with minor adaptations which are demonstrated in [23]; the reference showed that the noisy interference regime derived for the two-user interference channel in [10]–[12] remains unchanged even if each transmitter had a message to all receivers to form the two-user \( X \) channel.

A. Examples

Example 1 - SIMO Gaussian Many-to-one Interference Channel: The above theorem generalizes the noisy interference regime for many-to-one interference channels shown in [10]–[12]. To see this, consider the SIMO many-to-one interference channel where all the inputs are one-dimensional scalars, whereas all the outputs are vectors. In this case, note that the channel from Transmitter \( i \) to Receiver \( j \) can be represented by the vector \( H_{ji} \). Without loss of generality, let us assume that \(||H_{1i}||^2 = 1\). In this case, it can be verified that, if

\[ \sum_{i \in K_1} ||H_{1i}||^2 \leq 1, \]

then treating interference as noise is optimal. This can be seen with the auxiliary variables \( \tilde{Z}_i \sim \mathcal{N}(0, ||H_{1i}||^2) \) and, as mentioned in the above corollary, \( U_i = H_{1i}X_i + \tilde{Z}_i \). For the SISO case, the above condition boils down to the conditions specified in [10]–[12].

Example 2 - MIMO Gaussian Many-to-one Interference Channel: Consider a MIMO interference channel where \( M_i = N_i = M, \forall i \in K_1 \) and \( N_1 = 1 \). Since a MIMO \( M \times M \) channel can be decomposed into \( M \) parallel links using singular-value decomposition, we can assume that the channel matrices \( H_{ii}, i \in K_1 \) are diagonal without loss of generality. Let \( H_{1i}^{(k)} \) denote the \( k \)th diagonal entry of \( H_{ii}, \) \( H_{1i}, i \in K_1 \) is a \( M \times 1 \) vector, whose \( k \)th entry, we denote by \( H_{1i}^{(k)} \). Now, if

\[ \sum_{i \in K_1} \sum_{k=1}^M \frac{|H_{1i}^{(k)}|^2}{|H_{1i}^{(k)}|^2} \leq 1, \]

then, treating interference as noise is optimal. This can be noted by setting \( \tilde{Z}_i \sim \mathcal{N}(0, \sum_{k=1}^M \frac{|H_{1i}^{(k)}|^2}{|H_{1i}^{(k)}|^2}) \) and \( U_i = H_{1i}X_i + \tilde{Z}_i, \) for \( i \in K_1 \).

Example 3 - Fading Gaussian Many-to-one Interference Channel without CSIT: Consider a SISO Gaussian interference channel with Rayleigh fading. In this case, let the received signals maybe expressed similar to (3)-(4), where all the quantities are scalars; the only difference being that, in this case, the channel fade \( H_{ij} \) is time-varying. Specifically, the input-output relations can be expressed as

\[ Y_i(\tau) = H_i(\tau)X_i(\tau) + Z_i(\tau), i \in K_1 \]
\[ Y_1(\tau) = \sum_{i \in K} H_{1i}(\tau)X_i(\tau) + Z_1(\tau), i \in K_1, \]

where, all the receivers have 1 antenna so that the inputs \( X_i \), outputs \( Y_i \) and the channel gains \( H_{ij} \) are scalars for all \( i \in K_1 \) and \( j = 1 \) or \( j = i \). We assume that \( H_{ij}(\tau) \) is drawn i.i.d according to a circularly symmetric Gaussian distribution \( \mathcal{N}(0, \sigma_j^2) \). Note that this is the classical Rayleigh fading model, where the magnitude of the fade \( |H_{ij}| \)
is Rayleigh distributed with parameter \(\sigma_{ij}\). We assume that \(H_{ij}\) is independent of \(H_{ij^\prime}\) if \(i \neq i^\prime\) or \(j \neq j^\prime\). The receivers are aware of the channel state information, or equivalently, the effective output at Receiver \(j\), corresponding to the \(r\)th symbol can be expressed as \((Y_{ij}(\tau), H(\tau))\), where \(H(\tau) = \{H_{ij}(\tau) : i = 1, \ldots, K, j = 1, \ldots, K\}\). We assume that the transmitters do not have CSI, so that the input codewords are independent of the channel gains. Suppose \(\sum_{i=1}^{K} \frac{\sigma_{ij}}{\sigma_{ii}} \leq 1\), then, the noisy interference condition is satisfied. This can be verified on noting that, if \(\sigma_{ij}\) satisfy the specified condition, the effective interference at Receiver 1, \((V, H)\), is a degraded version of \((\sum_{i \in K_1} \frac{\sigma_{ij}}{\sigma_{ii}} Y_i, H)\). Thus, this condition on the variances \(\sigma_{ij}\) provides a condition for noisy interference in the fading Gaussian many-to-one interference channel without CSIT.

**Example 4 - Collision-based Interference Channel Model:** Here, we construct a collision-based model for the many-to-one interference channel. Intuitively, the model can be explained as follows. Transmitter \(i \in K_1\) has two choices - it can transmit a symbol from a finite set \(X^i_1\), or it can choose to remain silent. The signal transmitted by Transmitter \(i \in K_1\) is received perfectly at Receiver \(i\). At Receiver 1, there are two possibilities: the signal transmitted by Transmitter 1 can be received perfectly, or it can be erased (due to a collision). The probability of collision/erasure can be designed based on the set of transmitters which are silent. Formally, the model can be constructed as follows. Consider a \(K\)-user many-to-one interference channel, where \(X^i_1 = \{\phi\} \cup X^i_2\), where \(X^i_2\) is a finite set which does not contain the element \(\phi\). The element \(\phi\) is used to indicate the case where user \(i\) remains silent (Note that this symbol can be used in the code at Transmitter \(i\) to convey information to the Receiver \(i\)). The received signals are defined as \(Y_i = X_i\) for \(i \in K_1\), and \(V = \{0, \varepsilon\}\) with \(Y_1 = f_1(X_1, V) = \begin{cases} X_1, & \text{if } V = 0 \\ \varepsilon, & \text{if } V = \varepsilon \end{cases}\).

Also, \(V\) is drawn based on any probability distribution \(p_{V|X_1}\). In this channel, \(V = \varepsilon\) can be interpreted as an occurrence of a collision at Receiver 1. In this channel, clearly, \(V\) is a degraded version of \(X_1\) and therefore, the search for the sum-capacity of this channel is reduced to the search of the optimal single-letter distribution on all the inputs. Note that this model captures the traditional collision based medium access models - this can be seen by setting \(V = \varepsilon\) deterministically, when any of the users that 'collide' with user 1 are transmitting any symbol other than \(\phi\), and setting \(V = 0\) otherwise. With the optimal input distribution, since treating interference as noise is optimal at Receiver 1, the receiver effectively observes a binary erasure channel, with the erasure probability calculated based on the input distribution and \(p_{V|X_1}\).

### B. Separability of Parallel Noisy Discrete Memoryless Many-to-one Interference Channels

In this section, we consider a parallel extension of the class of discrete memoryless many-to-one interference channels introduced in Section II. Specifically, we show that if the many-to-one interference channel formed over each carrier (parallel component) of the parallel channel satisfies the noisy interference condition of Theorem 1, then separate random coding - sending independent random codewords over each carrier - along with treating interference as noise is sum-capacity optimal. We now proceed to describe the model and our result formally.

The class of \(F\)-carrier parallel \(K\)-user discrete many-to-one interference channels considered can be represented by the set of \(K\) (vector) inputs \(X_i = (X^{(1)}_i, X^{(2)}_i, \ldots, X^{(F)}_i) \in \mathcal{X}^{(1)} \times \mathcal{X}^{(2)} \times \mathcal{X}^{(F)}\) and a set of \(K\) (vector) outputs \(Y_i = (Y^{(1)}_i, Y^{(2)}_i, \ldots, Y^{(F)}_i) \in \mathcal{Y}^{(1)} \times \mathcal{Y}^{(2)} \times \mathcal{Y}^{(F)}\) for \(i = 1, 2, \ldots, K\). The \(k\)th component of output, \(Y^{(k)}_i\) is determined by the \(k\)th component of the inputs \(X^{(k)}_j\) for \(j = 1, 2, \ldots, K\). The \(k\)th component of the outputs \(Y^{(k)}_i\) is defined earlier in Section II. Note that the variables \(\tilde{V} = (V^{(1)}, V^{(2)}, \ldots, V^{(F)})\) and the functions \(\mathbf{f}_i = (f^{(1)}_i, f^{(2)}_i, \ldots, f^{(F)}_i)\) are used in defining the output are Receiver 1, with \(V^{(k)}, f^{(k)}_i\) used in defining the \(k\)th component of the output. There are \(K\) messages in the system \(W_1, W_2, \ldots, W_K\). The definition of a code of length \(T\), the probability of error, the corresponding rate of the code, the capacity region and the sum-capacity are defined similar to the Section II. The only difference is that, here, we restrict our study to constraints on the codewords, which ensure that either

- \(H(Y^{(k)}_i | X^{(k)}_i)\) and \(H(Y^{(k)}_i | X^{(k)}_i)\) exist for \(k = 1, 2, \ldots, F\) and \(i = 1, 2, \ldots, K\), or
- \(h(Y^{(k)}_i | X^{(k)}_i)\) and \(h(Y^{(k)}_i | X^{(k)}_i)\) exist for \(k = 1, 2, \ldots, F\) and \(i = 1, 2, \ldots, K\),

where, as before, \(T\) denotes the length of the codeword. The class of parallel discrete memoryless many-to-one interference channels considered here is, in fact, a special case of the class of channels defined in Section II. Before we proceed to our main result, we provide a parallel extension of Lemma II.
Lemma 2: Consider random sequences $\vec{A}^T, \vec{B}^T, \vec{C}^T$ such that $\vec{A} = (A^{(1)}, A^{(2)}, \ldots, A^{(F)})$ and similarly, $\vec{B}$ and $\vec{C}$ represent $F$-dimensional vectors/tuples. $\vec{B}^T, \vec{C}^T$ are generated as

$$p \left( \vec{B}^T, \vec{C}^T, \vec{A}^T \right) = p \left( \vec{A}^T \right) \prod_{k=1}^F \prod_{\tau=1}^T p \left( B^{(k)}(\tau) | A^{(k)}(\tau) \right) p \left( C^{(k)}(\tau) | B^{(k)}(\tau) \right),$$

where $p \left( B^{(k)}(\tau) | A^{(k)}(\tau) \right) = p^{(k)}_{B|A}$, $p \left( C^{(k)}(\tau) | B^{(k)}(\tau) \right) = p^{(k)}_{C|B}$ for all $\tau = 1, 2, \ldots, T, k = 1, 2, \ldots, F$. Note that the sequence $\vec{A}^T$ does not have to be generated in an i.i.d fashion. The sequences $\vec{B}^T, \vec{C}^T$ can be interpreted to be outputs of a physically degraded parallel ($F$-carrier) discrete memoryless broadcast channel whose (vector) input sequence is the $\vec{A}^T$. Also note that $\vec{A}^T \to \vec{B}^T \to \vec{C}^T$. Now, if $H(B^{(k)T}), H(C^{(k)T})$ exist for all $k = 1, 2, \ldots, F$, then

$$H(\vec{B}^T) - H(\vec{C}^T) \leq \sum_{k=1}^F \sum_{\tau=1}^T \left( H \left( B^{(k)}(\tau) \right) - H \left( C^{(k)}(\tau) \right) \right)$$

If $h(B^{(k)T}), h(C^{(k)T})$ exist for $k = 1, 2, \ldots, F$, then

$$h(\vec{B}^T) - h(\vec{C}^T) \leq \sum_{k=1}^F \sum_{\tau=1}^T \left( h \left( B^{(k)}(\tau) \right) - h \left( C^{(k)}(\tau) \right) \right)$$

Further, suppose that the channels describing $\vec{B}$ and $\vec{C}$ are Gaussian parallel broadcast channels, with an average covariance constraint on the input corresponding to each carrier, i.e., with

$$E \left[ \frac{1}{T} \sum_{\tau=1}^T \| A^{(k)}(\tau) A^{(k)}(\tau)^\dagger \right] = \Gamma^{(k)}$$

for some set of covariance matrices $\Gamma^{(k)}, k = 1, 2, \ldots, F$, then i.i.d circularly symmetric Gaussian distribution on $\vec{A}^T$, with each $A^{(k)}$ independent of $A^{(k')} \text{ for } k \neq k'$ maximizes the quantity $h(\vec{B}^T) - h(\vec{C}^T)$.

The proof of the lemma, which is similar to the proof of Lemma 1 is placed in Appendix B.

Corollary 2: Consider a parallel discrete memoryless many-to-one interference channel, where each of these parallel channels satisfy the noisy interference conditions of Theorem 1, i.e., where $V^{(k)}$ is a degraded version of $X^{(k)} \text{ w.r.t } X^{(k)}$. Then separate random coding over each of the parallel carriers and treating interference as noise achieves sum-capacity. The sum-capacity $C_\Sigma$ can therefore be written as

$$C_\Sigma = \sum_{k=1}^F \sum_{i=1}^K I(X^{(k)}_i; Y^{(k)}_i)$$

The proof of the above corollary omitted here, since it is almost identical to the proof of Theorem 1 with Lemma 2 used in the proof, in place of Lemma 1. Since the class of parallel many-to-one interference channels described above is a special case of the class of many-to-one channels described in Theorem 1 the optimality of random coding and treating interference as noise simply follow from the theorem - the additional insight of the above corollary is that the optimal input distribution involves the principle of separate coding, i.e., in the optimal input distribution, $X^{(k)}_i$ is independent of $X^{(k')}_i \text{ for } k \neq k'$ for all $i \in K$. The above corollary automatically implies that a set of parallel Gaussian many-to-one interference channels, each of which satisfies the noisy interference condition, is separable.

IV. ON INTERFERENCE ALIGNMENT IN NOISY INTERFERENCE REGIMES AND OTHER INSIGHTS FROM DETERMINISTIC MANY-TO-ONE INTERFERENCE CHANNELS

The deterministic many-to-one interference channel is the channel as described earlier, where $\mathcal{X}_i, \mathcal{Y}_i, \mathcal{V}$ are all finite and

$$H(Y_i | X_i) = H(V | X_{K_i}) = 0, \forall i \in K_1, \quad (6)$$

for all possible distributions on the input. Note that in this channel, the outputs can be uniquely determined from the set of inputs of the channel. Also note that this class of channels captures the deterministic framework proposed by [24], and studied in the many-to-one interference channel setting in [4]. We next proceed to understand the idea of interference alignment via random codes in this deterministic framework.
A. Discussion: Why is explicit interference alignment not required in the noisy interference regime?

For the deterministic many-to-one channel defined in (1), we present the noisy interference regime in a slightly different form which leads to interesting interpretation later in this section.

**Corollary 3:** In the many-to-one interference channel, if there exists a function $q$ such that $V = q(Y_{K_1})$, then the sum-capacity is given by

$$C_S = \max_{\Pi \in \mathcal{K}} H(Y_1) + H(Y_{K_1}|V)$$

**Proof:** The proof follows from setting $H(Y_i|X_i) = H(V|X_{K_i}) = 0$ for all $i \in K_i$ in Theorem [1] and noting that

$$\sum_{i \in K} H(Y_i) - H(V) = H(Y_{K_1}) - H(V) = H(Y_{K_1}|V)$$

since $Y_i$ is independent of $Y_j$ for $i \neq j$, and $V$ is a deterministic function of $Y_{K_1}$.

**Remark 1:** The above class of channels can be considered to be weak interference channels, because the condition of the result implies that the effective interference at Receiver 1 must be reconstructible from signals received at all receivers $i \neq 1$. The fact that random coding and treating interference as noise in the weak many-to-one interference channels is optimal can also be verified in the class of deterministic many-to-one interference channels studied by Bresler, Parekh and Tse [4].

For a better understanding of the noisy interference regime, let us take a closer look at the idea of interference alignment over the deterministic 3-user many-to-one channel. Over this channel, consider a random coding scheme of length $T$, such that it generates $2^{TR_2}$ typical sequences of $X_2^T$ and $2^{TR_3}$ typical sequences of $X_3^T$. Since messages in the system are distributed uniformly, this means that $H(X_i^T) = TR_i, i = 2, 3$. Now, if $(R_2, R_3)$ lie in the achievable rate region (with these distributions) of the multiple access channel formed with inputs $X_2, X_3$ and output $V$, then the sequences $X_2^T$ and $X_3^T$ are invertible (i.e., decodable) from $V^T$. This is the case because random coding is optimal in the multiple access channel. In other words, the sequences $X_2^T, X_3^T$ are not aligned; in fact, they are resolvable from $V^N$, and each $V^N$ sequence therefore corresponds to a unique $X_2^T, X_3^T$ sequence pair with high probability. Such codewords would satisfy

$$H(V^T) \approx TR_2 + R_3 = H(X_2^T) + H(X_3^T) = H(X_2^T, X_3^T)$$

The approximation sign is used above rather than equality, since the comparison is in an asymptotic sense. Contrary to the above scenario, if $(R_2, R_3)$ lie outside the rate region achieved with these distributions in the multiple access channel, then even with random coding, the sequences $X_2^T, X_3^T$ align. In particular, in the noisy interference regime, $V = q(Y_2, Y_3)$ and in achievable scheme, $R_i = H(Y_i)$, which imply that

$$H(V^T) < TR_2 + R_3 = H(Y_2^T, Y_3^T)$$

as long as $q$ is a non-invertible function (The case of $q$ being invertible falls in the class of the resolvable interference regime discussed later in this section). Note that the above condition holds in the noisy interference regime, for every possible input distribution. Thus, explicit alignment by an appropriate choice of input distribution or using multi-letter based structured coding is not required, and random codes automatically align interference in the above channel. The noisy nature of the interference can be explained by the insights of reference [25], which noted that on a single-user channel, if a random code of a rate higher than a user’s capacity is used, then the signal loses structure in the sense that the output satisfies an equipartition property independent of the codebook used. This loss of structure can be used to explain the noisy nature of the interference, on noting that for any given input distribution, with random coding, $R_i = H(Y_i) \geq I(V; X_i|X_{K_1-i}) = H(V|X_{K_1-i})$ in the noisy interference regime; in other words, from the perspective of a receiver with output $V$, the rate of transmission of the user is higher than the corresponding user’s mutual information and thus the interfering signal loses any structure imposed by its codebook. The additional insight here is that, if the rate of each incoming signal at a receiver is higher than that user’s mutual information, then, not only does the signal lose its structure, but the multiple signals also align. In fact, we will argue later in this section (Section [LV-A]) that the extent of alignment is also the maximum possible in this case. These insights carry through to the Gaussian case as well, where, if each user $i \in K_1$ transmit at rates corresponding their single-user capacity, then the interference is noisy, and is aligned - even by circularly symmetric Gaussian distributions at all inputs.
B. Resolvable Interference Regime for Deterministic Many-to-one Interference Channels

In this class of many-to-one interference channels, interference cannot be aligned since the multiple interferers at the first receiver are resolvable. Since alignment is not possible, random coding achieves the capacity region. We first show inner and outer bounds for the deterministic many-to-one interference channel, respectively, in Theorems 2 and 3. We then find conditions where these bounds are tight to define the resolvable interference regime in Corollary 4. The bounds and the regime are all defined in terms of auxiliary variables $U_{K_1} = (U_2, U_3, \ldots, U_K)$ such that

- $U_i$ is a deterministic function of $X_i$ and
- $V$ is a deterministic function of $(U_2, U_3, \ldots, U_K)$.

Note that $U_i = X_i$ provides a trivial assignment of auxiliary variables $U_i$. However, the bounds can be optimized over all possible choices of $U_i$ satisfying these properties. We now proceed to describe an outer bound on the capacity region of the many-to-one interference channel.

**Theorem 2:** The capacity region of the deterministic many-to-one interference channel lies in the convex hull of the following region, over all possible product distributions $\prod_{i \in K} p_{X_i}(x_i)$.

$$R_1 \leq H(Y_1|V)$$ (7)

$$R_i \leq H(Y_i), i \in K_1$$ (8)

$$R_1 + \sum_{i \in S} R_i \leq H(Y_1|U_{S^c}) + H(Y_{S}|V, U_{S^c}), \forall S \subseteq K_1,$$ (9)

where $S^c$ represents the complement of $S$ w.r.t $K_1$.

The above outer bound is proved in Appendix VI.

We now describe below, a rate achievable in general, in the deterministic many-to-one interference channel using a random coding scheme which does not align interference. The achievable scheme is similar to the one presented in [26] in the context of the deterministic $Z$ channel.

**Theorem 3:** For the deterministic many-to-one interference channel, the convex hull over all product input distributions $\prod_{i \in K} p_{X_i}(x_i)$, of the following rate region is achievable.

$$R_1 \leq H(Y_1|V)$$ (10)

$$R_i \leq H(Y_i), i \in K_1$$ (11)

$$R_1 + \sum_{i \in S} R_i \leq H(Y_1|U_{S^c}) + \sum_{i \in S} H(Y_i|U_i),$$ (12)

$\forall S \subseteq K_1, S = K_1 - S$.

The proof is placed in Appendix VI. It should be noted that the above achieved rate region is loose, in general, with respect to the bound of Theorem 2. However, if $H(U_{S^c}|V) = 0$ for all $S \subseteq K_1$, then the achieved rate region can be verified to be optimal by comparing (7)-(9) with (10)-(12). We state this formally below.

**Corollary 4:** Consider a many-to-one interference channel where $U_{K_1}$ is invertible from $V$, i.e., $H(U_{K_1}|V) = 0$ for all possible input distributions. Then, the capacity region of the many-to-one interference channel is given by (10)-(12).

**Proof:** Note that it is sufficient to show that the right hand sides of (12) and (9) are equal. We show this below.

$$H(Y_1|U_{S^c}) + \sum_{i \in S} H(Y_i|U_i)$$

$$= H(Y_1|U_{S^c}) + H(Y_S|U_S)$$ (13)

$$= H(Y_1|U_{S^c}) + H(Y_S|U_S, U_{S^c})$$ (14)

$$= H(Y_1|U_{S^c}) + H(Y_S|V, U_S, U_{S^c})$$ (15)

$$= H(Y_1|U_{S^c}) + H(Y_S|V, U_{S^c}).$$ (16)

In (13)-(14), we have used the fact that $(Y_i, U_i)$ is independent of $Y_j, U_j$ for $i \neq j, i, j \in K_1$. We have also used the fact that $U_S$ is invertible from $V$ in the final equation above.

Note that $U_j$ can be interpreted as the effective interference caused by Transmitter $i \neq 1$ at Receiver 1. Also, note that with any achievable coding scheme in this channel, Receiver 1 can decode $X_1^T$, and because of (3),...
invert \( V^T \) in this channel. The condition that \( U_K \) is invertible from \( V \) means that all the interfering signals \( U_1^T \) are resolvable at the first receiver, and hence alignment is precluded irrespective of the coding scheme used. Therefore, not surprisingly, the random coding achievable scheme of Theorem 3 is optimal for this class of channels.

It must be noted that, the achievable schemes of Theorem 3 and Corollary 3 are both different random coding schemes. The schemes differ, in particular, in the decoding procedure at Receiver 1 and hence achieve different rates. In the achievable scheme of Theorem 3 the rate region achieved is with Receiver 1 picking the sequence \( X_1^T \) such that \( Y_1^T, U_2^T, X_1^T \) are jointly typical (See Appendix VII). In contrast, in the decoding scheme for Corollary 3 the sequence \( X_1^T \) is decoded as the one such that \( (Y_1^T, X_1^T) \) are jointly typical, i.e., the interference is treated as noise.

C. Discussion: How many bits of additional rate can interference alignment provide?

The achievable scheme of Theorem 3 does not involve interference alignment and is therefore optimal, only when alignment is precluded on the many-to-one interference channel. The resolvability condition of the channel described in Corollary 4 above is precisely one where alignment is precluded. However, in general, if the resolvability condition is not satisfied, then on comparing (7)-(9) with (10)-(12), we can conclude that an additional rate of \( \Delta_S \) should be achieved by alignment for the users belonging in \( S \) for the outer bound to be tight. It is not clear whether this additional rate can be achieved at all, in general, or whether the outer bound is loose. However, the results on the noisy interference regime imply that if the many-to-one interference channel is weak, then this additional rate can be achieved using interference alignment via random coding, and the outer bound is tight in a sum-capacity sense (Compare expression of Corollary 3 with (9)). In other words, the extent of alignment is optimal in the sense that the additional rate benefit provided by alignment is the maximum possible. If the channel is not weak, then \( \Delta_S \) can be interpreted as a bound on the amount of additional rate that can be obtained via alignment for the users in \( S \subseteq K_1 \).

D. An open question: When does a channel have a single-letter capacity characterization?

A clear open problem motivated by this work is a capacity characterization of deterministic many-to-one interference channels. This question is particularly intriguing because it is not clear whether the channel allows a single-letter capacity characterization. Previous works on approximating the capacity of the channel motivate the need of structured (lattice) coding based achievable schemes [4]. It has been discussed in [19] that for channels where structured codes are necessary, single-letter characterizations may not exist. This is because coding schemes such as linear and lattice codes introduce structure as correlations in multiple uses of the channel. Interestingly, single-letter based lattice coding schemes (i.e., single-dimensional lattices) are shown to suffice for a degrees of freedom characterization of almost all interference channels in [9]; however, it has been argued multi-dimensional lattices are useful for finer characterizations of capacity [27]. Thus, the question of existence of single-letter characterizations of interference channels, and more general wireless networks remains wide open. The issue of existence of single-letter capacity characterizations also appears in several broadcast channel scenarios. The study of degrees of freedom of compound broadcast networks [28], [29] suggests the possibility of alignment and structured coding in the channel, whereas, for certain degraded settings, the broadcast (multicast) channel allows single-letter capacity characterizations (See [30] and references therein). Thus, an important open question in network information theory is a better understanding of structured codes, and its impact on capacity characterizations of discrete memoryless channels.

V. Conclusion

We generalize the noisy interference regimes, previously shown in average-power constrained SISO Gaussian interference channels, to the discrete memoryless many-to-one interference channel. In this noisy interference regime, random coding at all transmitters and treating interference as noise at the receiver which faces interference achieves sum-capacity on the many-to-one interference channel. Our generalization enables extension of the noisy interference regimes to the Gaussian MIMO and parallel many-to-one interference channels and the fading Gaussian many-to-one interference channels without CSIT. Unlike previous results which consider an average power constraint on the inputs, we also show that treating interference as noise is optimal in the Gaussian many-to-one interference channel, even if the inputs are constrained to come from fixed finite constellations such as QAM or PSK. Through the lens of interference alignment, we obtain a better understanding of why random (Gaussian) codewords are sufficient to achieve capacity in the noisy interference regime in the Gaussian interference channels. In particular,
we argue that if users transmit, using random coding, at rates higher than the interfering link’s mutual information, then the interference is noisy and the extent of alignment is maximum. Such alignment hence obviates the need for techniques such as structured (lattice) codes which have been shown to be immensely useful in other regimes. We also show that for deterministic many-to-one interference channels, if the interferers are resolvable at the receiver facing interference, random coding achieves capacity since interference alignment is precluded. While we are able to provide single-letter characterizations for certain classes of channels in this paper, the question of the existence of single letter characterizations for wireless networks, in general, remains open.

APPENDIX I

EQUIVALENCE OF OUR NOISY INTERFERENCE REGIME AND THAT OF [10]–[12] FOR SINGLE-ANTENNA GAUSSIAN MANY-TO-ONE INTERFERENCE CHANNELS

We have already shown in the introduction that the noisy interference regime of [10]–[12] is included in the regime described by our result. We here show that our noisy interference regime is no larger than the regime described in [10]–[12]. In particular, we will show here that, if the conditions of (1) are not satisfied, then, the effective interference at Receiver 1 is not a degraded version of the set of all signals at the other receivers. In particular, we will show that, if

$$\sum_{j=2}^{K} |H_{1j}|^2 > 1,$$

(17)

then, for any possible set of values \( \rho_i = E[Z_1 Z_i], i = 2, 3, \ldots, K \) with, there exists an input distributions on \( X_i, i = 1, 2, \ldots, K \) such that

$$I(X_2, X_3, \ldots, X_K; V|Y_2, Y_3, \ldots, Y_K) > 0,$$

(18)

where \( V = \sum_{i=2}^{K} H_{1i}X_i + Z_1 \). To see this, note that, since \( (Z_2, Z_3, \ldots, Z_K) \) is a set of mutually independent Gaussian random variables, the correlation matrix of \((Z_1, Z_2, \ldots, Z_K)\) can be written as

$$
\begin{bmatrix}
1 & \rho_1 & \rho_2 & \cdots & \rho_K \\
\rho_1 & 1 & 0 & & \\
\rho_2 & 0 & 1 & \ddots & \\
\vdots & \vdots & \ddots & \ddots & 0 \\
\rho_K & 0 & \cdots & 0 & 1
\end{bmatrix}
$$

Note that the above matrix has to be a positive semidefinite matrix, whose determinant is non-negative. This implies that

$$\sum_{i=2}^{K} \rho_i^2 \leq 1$$

(17) and the above equation together imply that there exists \( i_0 \in \{2, 3, \ldots, K\} \) such that \( |H_{1i_0}| > |\rho_{i_0}| \). We are now ready to provide the input distribution for which (13) is satisfied. Consider the input distribution where \( X_i = 0 \) deterministically for all \( i \in \{2, 3, \ldots, K\} - \{i_0\} \), and \( X_{i_0} \) is a circularly symmetric zero-mean Gaussian random variable having some positive (non-zero) variance. Then, we can write

$$I(X_2, X_3, \ldots, X_K; V|Y_2, Y_3, \ldots, Y_K)$$

$$= I(X_{i_0}; H_{1i_0}X_{i_0} + Z_1, Z_2, \ldots, Z_{i_0-1}, H_{i_0i_0}X_{i_0} + Z_{i_0}, Z_{i_0+1}, \ldots, Z_K)$$

$$\geq I(X_{i_0}; H_{1i_0}X_{i_0} + Z_1| H_{i_0i_0}X_{i_0} + Z_{i_0})$$

$$\geq 0$$

(19)

(20)

where (19) follows from the independence of \( X_{i_0} \) and \( Z_i, i = \{2, 3, \ldots, K\} \). In (20), the equality is satisfied (for Gaussian \( X_{i_0} \)) only if \( |\rho_{i_0}| = |H_{1i_0}|/|H_{i_0i_0}| \). However this condition is not satisfied for our choice of \( i_0 \), and we have, for this input distribution, \( I(X_2, X_3, \ldots, X_K; V|Y_2, Y_3, \ldots, Y_K) > 0 \), i.e., the interference at Receiver 1 is not a degraded version of the outputs at the other receivers. This shows the equivalence between our characterization of
the noisy interference regime and the characterization in [10]–[12]. It must be noted that we do not claim if (17) is satisfied, that the interference is not noisy - we only claim that it lies outside our characterization of the noisy interference regime. Whether the noisy interference regime characterized by us can be expanded is an interesting open question.

**Appendix II**

**Proof of Lemma 1**

From the Markov chain property on $A, B, C$, we can write

$$I(A^T; B^T) = I(A^T; B^T, C^T)$$
$$I(A^T; B^T) = I(A^T; C^T) + I(A^T; B^T|C^T)$$

$$\Rightarrow h(B^T) - h(C^T) = h(B^T|C^T) + h(B^T|A^T) - h(C^T|A^T) - h(B^T|C^T, A^T)$$

$$(a) \leq \sum_{\tau=1}^{T} h(B(\tau)|C(\tau)) + h(B(\tau)|A(\tau)) - h(C(\tau)|A(\tau)) - h(B(\tau)|C(\tau), A(\tau))$$

$$= \sum_{\tau=1}^{T} h(B(\tau)|C(\tau)) + h(B(\tau)|A(\tau)) - h(C(\tau), B(\tau)|A(\tau))$$

$$= \sum_{\tau=1}^{T} h(B(\tau)|A(\tau)) + h(B(\tau), C(\tau)) - h(C(\tau)) - h(C(\tau), B(\tau)|A(\tau))$$

$$= \sum_{\tau=1}^{T} h(B(\tau)) - I(B(\tau); A(\tau)) + I(B(\tau), C(\tau); A(\tau)) - h(C(\tau))$$

$$(b) = \sum_{\tau=1}^{T} h(B(\tau)) - h(C(\tau)),$$

where in $(a)$, we have used the chain rule and the fact that conditioning cannot increase differential entropy in the first two terms, and the following facts in the final two terms above

$$p(C^T|A^T) = \prod_{\tau=1}^{T} p(C(\tau)|A(\tau)),$$

$$p(B^T|C^T, A^T) = \frac{p(B^T, C^T|A^T)}{p(C^T|A^T)} = \prod_{\tau=1}^{T} \frac{p(B(\tau), C(\tau)|A(\tau))}{p(C(\tau)|A(\tau))} = \prod_{\tau=1}^{T} p(B(\tau)|C(\tau), A(\tau)).$$

In $(b)$, we have used the Markov chain property on $A, B, C$ which implies that $I(A(\tau); B(\tau), C(\tau)) = I(A(\tau); B(\tau))$. Now, if $p_B|A, p_C|B$ are both Gaussian, then the fact that the Gaussian input distribution on the $A(\tau)$ with the appropriate covariance matrix maximizes $h(B(\tau)|C(\tau))$ in step $(a)$, combined with the convexity of entropy implies that i.i.d Gaussian distribution on the input maximizes $h(B^T) - h(C^T)$.

**Appendix III**

**Proof of Theorem 1**

Achievability of the required rate follows trivially from typical set decoding arguments. We prove the converse here. We prove the converse for the case the differential entropies of $Y_k$ and $Y_k|X_k$ exist. The proof for the case the corresponding entropy terms exist, rather their differential entropies, is essentially identical, with the mutual information expressed in terms of the entropy of the variables. Consider any coding scheme of length $T$ achieving
rates $R_i$ for user $i$. Then, from Fano’s inequality, for any $\epsilon > 0$, we can write

$$T(R_1 - \epsilon/\log K) \leq I(Y_1^T;X_1^T)$$

$$\leq h(Y_1^T) - h(Y_1^T|X_1^T)$$

$$= h(Y_1^T) - I(V^T)$$

$$\leq \sum_{\tau=1}^T h(Y_1(\tau)) - I(V^T), \quad (21)$$

where we have used (2) above. Now, note that since the capacity of the channel only depends on the marginal distributions $p_{Y_i|X_{K_1}}, p_{Y_i|X_i}, i \in K_1$, we can make $V$ a physically degraded version of $Y_{K_1}$ so that $X_i^T \rightarrow Y_i^T \rightarrow V^T$ without changing the capacity. We can now write, for $i \in K_1$,

$$T(R_i - \epsilon/\log K) \leq I(Y_i^T;X_i^T)$$

$$= h(Y_i^T) - h(Y_i^T|X_i^T)$$

$$= h(Y_i^T) - \sum_{\tau=1}^T h(Y_i(\tau)|X_i(\tau))$$

$$\Rightarrow T \left( \sum_{i \in K_1} R_i - (K-1)\epsilon/\log K \right) \leq \sum_{i \in K_1} \left( h(Y_i^T) - \sum_{\tau=1}^T h(Y_i(\tau)|X_i(\tau)) \right)$$

$$\leq h(Y_i^T) - \sum_{i \in K_1} \sum_{\tau=1}^T h(Y_i(\tau)|X_i(\tau)). \quad (22)$$

where the final equation follows from the fact that $Y_i^T$ is independent of $Y_j^T$ for all $i \neq j$. Adding (21) and (22), we get

$$T \left( \sum_{i \in K} R_i - \epsilon \right) \leq \sum_{i \in K} h(Y_i(\tau)) - I(V^T) + h(Y_i^T) - \sum_{i \in K_1} \sum_{\tau=1}^T h(Y_i(\tau)|X_i(\tau))$$

$$\leq \sum_{\tau=1}^T h(Y_1(\tau)) + \sum_{i \in K_1, \tau=1}^T (h(Y_i(\tau)) - h(V(\tau))) - \sum_{i \in K_1} \sum_{\tau=1}^T h(Y_i(\tau)|X_i(\tau))$$

$$= \sum_{\tau=1}^T \left( h(Y_i(\tau)) - h(V(\tau)) + \sum_{i \in K_1} (h(Y_i(\tau)) - h(Y_i(\tau)|X_i(\tau))) \right)$$

$$= \sum_{\tau=1}^T \left( I(Y_i(\tau);X_i(\tau)) + \sum_{i \in K_1} I(X_i(\tau);Y_i(\tau)) \right) \quad (27)$$

$$= \sum_{\tau=1}^T \sum_{i \in K_1} I(X_i(\tau);Y_i(\tau)) \quad (28)$$

$$= \sum_{\tau=1}^T \sum_{i \in K} I(X_i(\tau);Y_i) \quad (29)$$

$$\leq T \max_{\tau} \sum_{i \in K} I(X_i(\tau);Y_i) \quad (30)$$

where (25) follows from the fact that $X_{K_1}^T \rightarrow Y_{K_1}^T \rightarrow V^T$ combined with the result of Lemma 1. The outer bound hence follows.
APPENDIX IV
PROOF OF COROLLARY 1

The condition that \( \sum_{i \in \mathcal{K}_1} A_i \preceq I_{N_i} \) implies that a Gaussian random \( N_1 \) dimensional vector \( Z \) can be found so that \( V = \sum_{i \in \mathcal{K}_1} U_i + Z \) and \( Y_i = X_i + V \), which means that \( V \) is a degraded version of \( U_{\mathcal{K}_1} \). This fact, combined with the condition that \( U_i \) is degraded version of \( Y_i \) implies that \( V \) is a degraded version of \( Y_{\mathcal{K}_1} \) as required by Theorem 1. The optimality of random coding and treating interference as noise hence follows from the theorem. Here, we only need to show that Gaussian inputs are optimal, when \( \mathcal{K}_i = \mathbb{C}^{M_i} \), and there is a power constraint on the inputs. Consider any achievable scheme where \( E \left( \frac{1}{T} \sum_{t=1}^{T} X_i(t) X_i(t)^\dagger \right) = \Gamma_i \). Then, following the steps of the proof of Theorem 1, we can derive equation (25), which is reproduced below.

\[
T \left( \sum_{i \in \mathcal{K}} R_i - \epsilon \right) \leq \sum_{\tau=1}^{T} h(Y_1(\tau)) + \sum_{\tau=1}^{T} (h(Y_{\mathcal{K}_1}(\tau)) - h(V(\tau))) - \sum_{i \in \mathcal{K}_1} \sum_{\tau=1}^{T} h(Y_i(\tau)|X_i(\tau))
\]

Here, we can use the Gaussian distribution to evaluate each of the entropy terms above. To see this, we invoke the convexity of entropy, and the fact that under a covariance matrix constraint, the Gaussian distribution maximizes entropy in the first term above. We can use Lemma 1 which shows that the use of the Gaussian distribution to evaluate the second and third terms above, outer bounds the terms. The final entropy term is evaluated using Gaussian because of the definition of the channel. Thus, if we have a power constraint, we are restricted to the set of all Gaussian distributions on the input at Transmitter \( i \) with covariance \( \Gamma_i \), where \( tr(\Gamma_i) \leq P_i \).

APPENDIX V
PROOF OF LEMMA 2

Proof: The proof is similar to the proof of Lemma 2. We only highlight the differences here. We have \( \tilde{A}^T \rightarrow \tilde{B}^T \rightarrow \tilde{C}^T \)

\[
I(\tilde{A}^T; \tilde{B}^T) = I(\tilde{A}^T; \tilde{B}^T|\tilde{C}^T) = h(\tilde{B}^T|\tilde{C}^T) + h(\tilde{B}^T|\tilde{A}^T) - h(\tilde{C}^T|\tilde{A}^T) - h(\tilde{B}^T|\tilde{C}^T, \tilde{A}^T)
\]

\[
\leq \sum_{k=1}^{F} \sum_{\tau=1}^{T} h\left( B^{(k)}(\tau)|C^{(k)}(\tau) \right) + h\left( B^{(k)}(\tau)|A^{(k)}(\tau) \right) - h\left( C^{(k)}(\tau)|A^{(k)}(\tau) \right) - h\left( C^{(k)}(\tau) \right)
\]

where in (c), we have used the chain rule and the fact that conditioning cannot increase differential entropy in the first two terms, and the definition of the channels \( p_B^{(k)} p_{C|B}^{(k)} \) in the final two terms above, similar to Lemma 1. The arguments for the derivation of the final step from (c), and the optimality of Gaussian inputs if the channels are Gaussian, are identical to the proof of Lemma 1 and hence, omitted here.

APPENDIX VI
PROOF OF THEOREM 2

The bounds (7) and (8) are trivial. We only need to show (9). Consider any achievable scheme. Let \( T \) be the length of the code. Consider any \( S \subseteq \mathcal{K}_i \). Then, for \( i \in S \), from Fano’s inequality we can write for any \( \epsilon > 0 \),

\[
T(R_i - \epsilon) \leq H(Y_i^T), \quad (31)
\]

\[
\Rightarrow T \left( \sum_{i \in S} R_i - |S|\epsilon \right) \leq \sum_{i \in S} H(Y_i^T) = H(Y_S^T | U_{S^c}), \quad (32)
\]

where, in the final two equations, we have used the fact that \( (Y_i, U_i) \) is independent of \( (Y_j, U_j) \) for \( j \neq i \), \( (i, j) \in \mathcal{K}_i \). Now, using Fano’s inequality for \( W_1 \), we get

\[
T(R_1 - \epsilon) \leq I(Y_1^T; X_1^T), \quad (33)
\]
transmitter also generates $p$ picks a $p$ channel. Without loss of generality, let us assume that $W$ is deterministic and the pair $(U_i, Y_i)$ is found. Receiver $m$ generates the $m$ sequence in the $m$th bin, such that it is jointly typical with $(X_i, U_i, Y_i)$ from which marginal distributions $p_{U_i}(u_i), p_{Y_i}(y_i)$ and $p_{U_i, Y_i}(u_i, y_i)$ can be calculated. The transmitter generates $2^{TR_1}$ sequences of $X_i^T$, each sequence generated independently and i.i.d according to $p_{X_i}(x_i)$. Let the generated sequences be denoted by $X_i^T(m), m \in \{1, 2, \ldots, 2^{TR_1}\}$. Then the message $W_1 = m$ is encoded using $X_i^T(m)$. Now consider Transmitter $i \in K$. Note that $p_{X_i}(x_i)$ along with the channel induces $p_{X_i, U_i, Y_i}(x_i, u_i, y_i)$ from which marginal distributions $p_{U_i}(u_i), p_{Y_i}(y_i)$ and $p_{U_i, Y_i}(u_i, y_i)$ can be calculated. The transmitter generates $2^{TH_i}$ sequences of $Y_i^T$, each sequence generated independently and i.i.d according to $p_{Y_i}(y_i)$. These sequences of $Y_i^T$ are distributed uniformly into $2^{TH_i}$ bins. To encode the $m$th message, the transmitter picks a $Y_i^T$ sequence in the $m$th bin, such that it is jointly typical with $U_i^T(m_i)$ for some $m_i \in \{1, 2, \ldots, 2^{TH_i}\}$. If no such sequence is found, then an error is declared. Otherwise, the message is encoded using the $X_i^T$ which generates the $(U_i^T, Y_i^T)$ sequence picked. The existence of such a $X_i^T$ sequence is guaranteed, because the channel is deterministic and the pair $(U_i^T, Y_i^T)$, by virtue of being jointly typical, has a non-zero probability of occurrence.

**Remark:** The encoding strategy at Transmitter 2 is similar to the optimal coding strategy over the deterministic broadcast channel [31].

**Decoding Strategy:** Receiver 1, on receiving $Y_i^T$, chooses the unique index $W_1 = m$, such that

$$Y_i^T, U_2^T(m_2), U_3^T(m_3), \ldots, U_K^T(m_K), X_i^T(m)$$

is jointly typical, for some $m_i \in \{1, 2, \ldots, 2^{K}\}$ for $i = 2, 3, \ldots, K$. An error is declared if no such unique index $m$ is found. Receiver $i \in K_1$ can decode $W_i$ using the bin-index of the received $Y_i^T$ sequence. Note that since the channel is deterministic, there are no errors at Receiver $i \neq 1$, if the encoding at Transmitter $i$ is successful.

**Error Analysis and Achieved rate:** Since the coding scheme is symmetric over all messages, we will analyze the probability of error assuming $W_1 = 1$ is encoded at transmitter $i \in K$; because of symmetry in the coding scheme, the probability of error of encoding this set of messages gives the probability of error averaged over all messages. Now, we divide the possible set of errors into two types: errors at Transmitter $i \in K_1$ and errors at Receiver 1.

For $i \in K_1$, let

$$E_i(m_j) = \begin{cases} \text{At Transmitter } i, \ Y_i^T \text{ belongs to the } m_i\text{th bin, } U_i^T = U_i^T(m) & \text{for some } m \in \{1, 2, \ldots, 2^{TH_i}\} \Rightarrow (U_i^T, Y_i^T) \notin A_e(U_i, Y_i) \\ \end{cases}$$

where $A_e(U_i, Y_i)$ represents the $e$-jointly typical set of $(U_i^T, Y_i^T)$ pairs. Note that $E_i(m_j)$ corresponds to the event that no jointly typical pair $Y_i^T, U_i^T$ was found in the $m_i\text{th bin at Transmitter } i$ when encoding $W_i = m_j$. The overall probability of error can now be expressed as

$$P_e^T = \sum_{i \in K_1} \Pr(E_i(1)) + \Pr(\text{Decoding error at receiver } 1|E_2^c(1), E_3^c(1), \ldots, E_K^c(1)).$$
Now, consider Receiver 1. Note that, at this receiver, the decoding procedure and hence the error events are very similar, in nature, to the errors that can occur over a multiple access channel (MAC), when the asymptotically optimal typical set coding procedure is used. The only difference is that, in this particular case, the receiver is only interested in one message, i.e., $W_1$, which reduces the number of possible error events as compared to the classical MAC. Now, given that the message $W_i = 1$ is encoded at Transmitter $i$ for $i \in K$, and no errors occurred at the transmitter, a sequence $U_i^T$ is found at the transmitter such that it is jointly typical with a $Y_i^T$ sequence in the first bin. Let us assume, without loss of generality, that this sequence found is $U_i^T (1)$, i.e., $U_i^T (1)$ is used in the encoding procedure at Transmitter $i$, for $i \neq 1$. If $E_i^e (1)$ occurs for $i \neq 1$, then, because of the deterministic nature of the channel, the received sequence $Y_i^T$ is jointly typical with $U_i^T (1), U_2^T (1), \ldots, U_K^T (1), X_i^T (1)$. Errors can occur if $Y_i^T (1)$ is jointly typical with $U_2^T (m_2), U_3^T (m_3), \ldots, U_K^T (m_K), X_i^T (m_1)$ for some $m_1 \neq 1, m_i \in \{1, 2, \ldots, 2^T\}$ for $i \in K$. We wish to evaluate the probability of occurrence of this event. Let us define

\[ E_1(m_1, m_2, m_3, \ldots, m_K) = \{ Y_1^T, U_2^T (m_2), U_3^T (m_3), \ldots, U_K^T (m_K), X_i^T (m_1) \text{ is jointly typical} \} \]

By the union bound, we can bound the error at Receiver 1 as

\[
\Pr(\text{Error at Receiver 1}|E_2^e (1), E_3^e (1), \ldots, E_K^e (1)) \\
\leq \sum_{m_1=2}^{2^T \Omega_1} \sum_{m_2 \in \{1, 2, \ldots, 2^T\}, i \in K_1} \Pr(E_1(m_1, m_2, \ldots, m_K)) \\
= \sum_{m_1=2}^{2^T \Omega_1} \sum_{S \subseteq K_1} \sum_{m_i \neq 1, i \in S} \sum_{m_i=1, i \in S^c} \Pr(E_1(m_1, m_2, \ldots, m_K)) \\
= \sum_{S \subseteq K_1} \sum_{m_1=2}^{2^T \Omega_1} \sum_{m_i \neq 1, i \in S} \Pr(E_1(m_1, m_2, \ldots, m_K))
\]

Now, the overall error probability can be bounded as

\[
P_e^T \leq \sum_{S \subseteq K_1} \sum_{m_1=2}^{2^T \Omega_1} \sum_{m_i \neq 1, i \in S} \Pr(E_1(m_1, m_2, \ldots, m_K)) + \sum_{i \in K_1} \Pr(E_i(1)). \tag{40}
\]

It has been shown in [32] that if, for $i \neq 1$,

\[
R_i \leq H(Y_i) \tag{41}
\]

\[
R_i \leq \Omega_i + H(Y_i) - I(Y_i; U_i) \tag{42}
\]

then, asymptotically as $T \to \infty$, the probability of $E_i(1)$ occurring vanishes. Now, we estimate the remaining term in (40) below. Let $A_e$ denote the set of all $e$-jointly typical sequences of $(Y_1^T, U_2^T, \ldots, U_K^T, X_i^T)$. Then, for any $S \subseteq K_1$, we can write

\[
\sum_{m_1=2}^{2^T \Omega_1} \sum_{m_i \neq 1, i \in S} \Pr(E_1(m_1, m_2, \ldots, m_K)) \\
\leq \sum_{m_1=2}^{2^T \Omega_1} \sum_{m_i \neq 1, i \in S} \Pr \left( (Y_1^T, U_2^T (m_2), U_3^T (m_3), \ldots, U_K^T (m_K), X_i^T (m_1)) \in A_e \right) \\
\leq \sum_{m_1=2}^{2^T \Omega_1} \sum_{m_i \neq 1, i \in S} \sum_{Y_i^T, U_1^T (m_1) = U_i^T} \Pr (Y_i^T | U_i^T) \Pr (X_i^T (m_1) = X_i^T) \prod_{j \in K_1} \Pr (U_j^T (m_j) = U_j^T)
\]
these inequalities using Fourier-Motzkin elimination, we get the achieved rate region in the desired form.

\[ \sum_{m_1=2}^{2^TR_1} \sum_{m_i \neq 1, i \in S} |A_i| \leq 2^{-TH(Y_1|U_{S^c})} + 2^{-TH(U_{S^c}|X_1)} - TH(X_1) + K\epsilon \]

\[ \sum_{m_1=2}^{2^TR_1} \sum_{m_i \neq 1, i \in S} 2^{-TY(Y_1|U_{S^c},X_1) + (K+2)\epsilon} \]

\[ = 2^TR_1 2T \sum_{i \in S} T\Omega_i 2^{-TH(Y_1|U_{S^c}) + (K+2)\epsilon}, \]

where, in (a), we have used the fact that \( U_j \) is independent \( U_{j'} \) for \( j \neq j' \). From the above equation, it can be concluded that, if,

\[ R_1 + \sum_{i \in S} \Omega_i \leq H(Y_1|U_{S^c}), \forall S \subseteq K_1 \]

(43)

then, asymptotically as \( T \to \infty \), the probability of error at Receiver 1 vanishes. Note that if \( S \) is the null-set, the above equation can be equivalently expressed as \( R_1 \leq H(Y_1|V) \). Thus, if the rates \( R_i, i \in K \) and the parameters \( \Omega_i, i \in K_1 \) satisfy (41), (42) and (43), the rate-tuple \((R_1, R_2, \ldots, R_K)\) is achievable. Eliminating \( \Omega_i, i \in K_1 \) from these inequalities using Fourier-Motzkin elimination, we get the achieved rate region in the desired form.

REFERENCES

[1] M. Maddah-Ali, A. Motahari, and A. Khandani, “Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis,” in *IEEE Trans. on Information Theory*, pp. 3457–3470, 2008.

[2] S. Jafar and S. Shamai, “Degrees of freedom region for the MIMO X channel,” *IEEE Trans. on Information Theory*, vol. 54, pp. 151–170, Jan. 2008.

[3] V. Cadambe and S. Jafar, “Interference alignment and the degrees of freedom of the K user interference channel,” *IEEE Trans. on Information Theory*, vol. 54, pp. 3425–3441, Aug. 2008.

[4] G. Bresler, A. Parekh, and D. Tse, “Approximate capacity of the many-to-one interference channel,” *Allerton Conference*, Sep. 2007.

[5] V. R. Cadambe, S. A. Jafar, and C. Wang, “Interference alignment with asymmetric complex signaling - settling the Host-Madsen-Nostratinia conjecture,” in arXiv:0904.0274, April 2009.

[6] V. R. Cadambe, S. A. Jafar, and S. Shamai, “Interference alignment on the deterministic channel and application to fully connected gaussian interference networks,” *IEEE Transactions on Information Theory*, vol. 55, pp. 269–274, Jan. 2009.

[7] S. Sridharan, A. Jafarian, S. Vishwanath, and S. A. Jafar, “Capacity of symmetric K-user Gaussian very strong interference channels,” *Proceedings of IEEE Globecom*, pp. 1–5, Dec. 2008.

[8] R. Etkin and E. Ordentlich, “The degrees-of-freedom of the K-user Gaussian interference channel is discontinuous at rational channel coefficients,” *IEEE Transactions on Information Theory*, vol. 55, pp. 4932–4946, Nov. 2009.

[9] A. S. Motahari, S. O. Gharan, M. A. Maddah-Ali, and A. K. Khandani, “Forming pseudo-MIMO by embedding infinite rational dimensions along a single real line: Removing barriers in achieving the DOFs of single antenna systems,” arXiv:0904.0274, Feb 2009.

[10] V. Annapureddy and V. Veeravalli, “Gaussian interference networks: Sum capacity in the low interference regime and new outer bounds on the capacity region,” in Submitted to *IEEE Transactions on Information Theory*, arxiv:0802.3495, Feb 2008.

[11] A. Motahari and A. Khandani, “Capacity bounds for the Gaussian interference channel,” *IEEE Transactions on Information Theory*, vol. 55, pp. 620–643, Feb. 2009.

[12] X. Shang, G. Kramer, and B. Chen, “A new outer bound and the noisy-interference sum-rate capacity for Gaussian interference channels,” *IEEE Transactions on Information Theory*, vol. 55, pp. 689–699, Feb. 2009.

[13] V. S. Annapureddy and V. V. Veeravalli, “Sum capacity of MIMO interference channels in the low interference regime,” *Arxiv preprint*, Sep 2009. http://arxiv.org/abs/0909.2074.

[14] X. Shang, B. Chen, G. Kramer, and H. V. Poor, “Capacity regions and sum-rate capacities of vector Gaussian interference channels,” *Arxiv pre-print*, August 2009. http://arxiv.org/abs/0907.0472.

[15] L. Sankar, X. Shang, E. Erkip, and H. Poor, “Ergodic two-user interference channels: Is separability optimal?,” in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pp. 723–729, Sept. 2008.

[16] V. Cadambe and S. Jafar, “Parallel Gaussian interference channels are not always separable,” *IEEE Transactions on Information Theory*, vol. 55, pp. 3983–3990, Sept. 2009.

[17] L. Sankar, X. Shang, E. Erkip, and H. V. Poor, “Ergodic fading interference channels: Sum-capacity and separability,” arxiv.org, vol. abs/0906.0744, 2009. http://arxiv.org/abs/0906.0744.

[18] X. Shang, B. Chen, G. Kramer, and H. V. Poor, “Noisy-interference sum-rate capacity of parallel Gaussian interference channels,” *Arxiv pre-print*, March 2009. http://arxiv.org/abs/0903.0595.

[19] T. Philosof and R. Zamir, “On the loss of single-letter characterization: The dirty multiple access channel,” *IEEE Transactions on Information Theory*, vol. 55, pp. 2442–2454, June 2009.

[20] T. Gou and S. A. Jafar, “Capacity of a class of symmetric SIMO Gaussian interference channels within O(1),” *Arxiv preprint*, May 2009. http://arxiv.org/abs/0905.1745.

[21] E. Telatar and D. Tse, “Bounds on the capacity region of a class of interference channels,” in *IEEE International Symposium on Information Theory*, 2007, pp. 2871–2874, June 2007.

[22] A. E. Gamal and M. Costa, “The capacity region of a class of deterministic interference channels,” *IEEE Trans. Inform. Theory*, vol. 2, pp. 343–346, March 1982.
[23] C. Huang, V. R. Cadambe, and S. A. Jafar, “On the capacity and generalized degrees of freedom of the x channel,” arxiv.org, vol. abs/0810.4741, 2008. http://arxiv.org/abs/0810.4741.
[24] A. S. Avestimehr, S. Diggavi, and D. Tse, “A deterministic approach to wireless relay networks,” Oct 2007, arXiv:cs.IT/0710.3777.
[25] X. Wu and L.-L. Xie, “Asymptotic equipartition property of output when rate is above capacity,” Arxiv pre-print, vol. abs/0908.4445, August 2009. http://arxiv.org/abs/0908.4445.
[26] V. R. Cadambe, S. A. Jafar, and S. Vishwanath, “The capacity region of a class of deterministic Z channels,” Proc. of IEEE International Symposium on Information Theory, June 2009.
[27] A. Jafarian, J. Jose, and S. Vishwanath, “Algebraic lattice alignment for K-user interference channels,” in Proceedings of Allerton Conference on Communication, Controls, and Computing, September 2009.
[28] H. Weingarten, S. Shamai, and G. Kramer, “On the compound MIMO broadcast channel,” in Proceedings of Annual Information Theory and Applications Workshop UCSD, Jan 2007.
[29] T. Gou, S. A. Jafar, and C. Wang, “On the degrees of freedom of finite state compound wireless networks - settling a conjecture by Weingarten et. al,” September 2009. http://arxiv.org/abs/0909.4177.
[30] C. Nair and A. E. Gamal, “The capacity region of a class of three-receiver broadcast channels with degraded message sets,” IEEE Transactions on Information Theory, vol. 55, pp. 4479–4493, Oct. 2009.
[31] T. Cover, “Comments on broadcast channels,” IEEE Transactions on Information Theory, vol. 44, pp. 2524–2530, Oct 1998.
[32] A. El Gamal and E. van der Meulen, “A proof of Marton’s coding theorem for the discrete memoryless broadcast channel,” IEEE Transactions on Information Theory, vol. 27, no. 1, pp. 120–122, 1981.