Blind Index Coding

David T.H. Kao*, Mohammad Ali Maddah-Ali†, and A. Salman Avestimehr*

*University of Southern California, Los Angeles, CA, USA †Bell Labs, Holmdel, NJ, USA

Abstract

We introduce the “blind index coding” (BIC) problem, which generalizes the classic index coding problem by allowing the sender to have some uncertainty about the side information that is available at each receiver. This problem naturally arises in wireless networks, in which users obtain their side information through wireless channels with errors that may be unknown to the sender. For the proposed BIC problem, we develop a new general outer bound by first proving it for the 3-user case, and then generalizing its construction to $K$ users. The proof of the outer bound relies on development of a key lemma that uses a strong data processing inequality to account for the sender’s uncertainty. We also propose a hybrid coding scheme that XORs random combinations of bits from a subset of messages with uncoded bits of other messages in order to blindly exploit side information, and illustrate its gain. We finally generalize the BIC problem to consider a wireless channel from sender to users, and observe that even in the case of two users the solution becomes non-trivial and a natural generalization of the hybrid coding scheme relying on XORing repetitions of uncoded bits strictly outperforms conventional schemes.

I. INTRODUCTION

In many communication scenarios, users may have access to some side information about the messages that are requested by other users. This scenario can arise in, for example, caching networks in which users opportunistically cache content that may be requested by other users in the future. It can also arise in wireless networks in which users can overhear the signals intended by other users over the shared wireless medium [3]. However, due to coordination constraints, each user may only be aware of the amount of side information available at other users in the network, and not its specific content. So, a natural question is: how can we take advantage of the existence of unknown side information to efficiently deliver messages to users?

To understand this problem, and evaluate and isolate the ultimate gain of such side-information, we introduce a basic network communication problem with one sender and several users, depicted in Figure 1. The sender communicates a distinct message, $\vec{w}_i$, to each of $K$ users (labeled $i = 1, \ldots, K$) over a broadcast communication channel, while each user, $i$, has some prior side information ($\vec{\phi}_i$) about other users’ desired messages (e.g., $\vec{w}_j$, where $j \neq i$) that it may use to assist in decoding its own desired message. However, the sender does not know the precise side information given to each user (i.e., the sender is blind), and it must design its transmission strategy using only knowledge of the probability distributions of $\vec{\phi}_i$ for all $i \neq j$.

We refer to this new formulation as the blind index coding (BIC) problem. Our formulation is a generalization of the classic index coding problem [4–6], which is a canonical problem in network communication theory and, despite its simple formulation, remains a powerful tool for analyzing many network communication settings (see e.g., [7–11]). The key difference in BIC problems lies in the sender’s uncertainty regarding side information: In classic index coding, precise knowledge of side information is used by the sender to create transmission strategies that treat message bits differently depending on whether they are within or not within side information at each particular user [12–13]. However, in BIC the sender is unable to distinguish between such message bits, and thus transmission must “blindly” exploit the existence of side information. As we will see, this otherwise minor difference significantly changes the technical challenges of the problem.

The main question that we investigate in this paper is “to what extent and using what techniques can we blindly exploit such side information?” To that end, after formally introducing the BIC problem, our first contribution is the development of an outer bound on the capacity region, which utilizes a strong data processing inequality [13] to capture the inability of the sender to distinguish between bits of a single message known or unknown to a given user prior to transmission. We demonstrate that our converse is tight in two specific cases: namely the two-user and symmetric $K$-user BIC (where all users have the same amount of knowledge about undesired messages). In both cases a simple approach based on random coding suffices to achieve the entire capacity region.

As we move beyond these special cases to the general $K$-user BIC setting, our second contribution is the development of a class of hybrid coding schemes, which XOR random linear combinations of bits of a subset messages with uncoded bits from other messages. In these hybrid codes, the sender XORs uncoded bits in order to probabilistically exploit side information already available at users. We provide an example to show that this approach can outperform random coding, and in fact sometimes meet the new capacity outer bound. For the 3-user BIC problem, we construct an achievable scheme based on this approach and examine its efficacy analytically and numerically relative to existing achievable schemes and the new outer bound.

In our final contribution, we extend the problem to study blind index coding over wireless channels (BICW). Specifically, we consider a model where the sender-to-user links occur over a lossy wireless channel using a binary fading channel model.

The research of A.S. Avestimehr and D.T.H. Kao is supported by NSF Grants CAREER 1408639, CCF-1408755, NETS-1419632, EARS-1411244, ONR award N000141310094, and research grants from Intel and Verizon via the 5G project. Parts of this work will be presented in [1] and [2].
Interestingly, when the reliability of the broadcast channel differs between users, even solving the case of two-users becomes quite challenging.

Therefore, for BICW problems we present the following. We first identify a representative two-user problem that considers lossy sender-to-user transmission links as well as blind side information. For this problem, we demonstrate that in addition to hybrid coding (where XORing uncoded bits of a subset of messages with random combinations of the others played a key role), quite surprisingly XORing the same uncoded bits more than once (i.e., repetition of uncoded bits) can increase the achievable rate. Equipped with this observation, we then proceed to construct a coding scheme that leverages both hybrid codes and repetition of uncoded message bits in order to establish an achievable rate region, and we demonstrate numerically that such a scheme can only result in a larger achievable rate region.

To summarize, the main contributions are as follows:

1) We introduce the Blind Index Coding problem, which generalizes classic Index Coding by considering uncertainty (blindness) at the sender about side information given to users.

2) We derive a novel outer bound on the capacity region of BIC problems which leverages a strong data processing inequality to account for the blindness of the sender.

3) We propose a class of hybrid coding schemes, which XOR random linear combinations from one subset of messages with uncoded bits of another subset.

4) We further generalize the problem to better model wireless settings by studying how lossy sender-to-user links affect the efficacy of the hybrid coding schemes, and we find that repetition coding can enhance the performance of hybrid codes.

This remainder of the paper is organized in the following way. In Section II we formally state the BIC problem first for error-free broadcast and then for broadcast over lossy channels. In Section III we motivate both the notion of hybrid coding and our outer bound using a simple example, for which the inner and outer bounds meet. In Section IV we state and prove the general outer bound for K-users. In Section V we define a hybrid coding scheme and study the achievable symmetric rate for the three-user BIC, and in Section VI we numerically compare achieved rates to the derived outer bounds. In Section VII we consider blind index coding when the sender-to-user links occur over wireless channels. Concluding remarks and open questions are presented in Section VIII.

II. THE BLIND INDEX CODING PROBLEM

In this section, we now formally define the Blind Index Coding problem by stating the network and side information models, and formalizing the notion of capacity.

Network model: As shown in Figure 1 in a BIC problem K users each request a message from a server; i.e., User $i$, for $i = 1, \ldots, K$, desires the $m_i$-bit message $\vec{w}_i$, which is drawn uniformly from a space $\{0, 1\}^{m_i}$. Each user, $i$, has access to side information, $\vec{\phi}_{ij}$, (whose form is described later) about each message $\vec{w}_j$, except the one it desires (i.e., for all $j \neq i$). The sender aims to communicate all messages to respective users via a common error-free channel. The goal of the problem is to design a channel input vector, $\vec{x}$, of minimum length, such that each user can decode its desired message.

Side information model: In a blind index coding problem, each side information signal, $\vec{\phi}_{ij}$, is a random fraction of the bits that make up the message, $\vec{w}_j$. We assume that the sender is “blind” in the sense that it is only aware of the average number of bits in each side information signal.

More specifically, we can model the side information in the following way. Let $\vec{g}_{ij}$ be a length-$m_j$ binary vector drawn i.i.d from a Bernoulli$(1 - \mu_{ij})$ distribution. Side information $\phi_{ij}[\ell] = (\phi_{ij}[1], \phi_{ij}[2], \ldots, \phi_{ij}[m_j])$ is such that, for $\ell = 1, \ldots, m_j$,

$$
\phi_{ij}[\ell] = g_{ij}[\ell] \cdot w_{ij}[\ell].
$$

(1)

User $i$ knows $\vec{g}_{ij}$ for all $j \neq i$, however the sender is only aware of parameters, $\{\mu_{ij}\}$, which govern the probabilistic behavior of the side information. Note that the side information model is equivalent to either 1) randomly sampling bits of a message, or 2) passing a message through a side information channel which is an erasure channel.

**Fig. 1.** A $K$-user blind index coding problem (e.g., $K = 3$) depicted as a server-user networks with user caches. User $i$, for $i = 1, 2, 3$, desires message $\vec{w}_i$ and may use side information about other messages to facilitate decoding; $\vec{\phi}_{ij}$ denotes the side information that User $i$ has about Message $\vec{w}_j$. The sender only has knowledge of the distribution of $\vec{\phi}_{ij}$, and not its precise realization. Notice the amount of side information available may vary across users and messages.

Interestingly, when the reliability of the broadcast channel differs between users, even solving the case of two-users becomes quite challenging.

Therefore, for BICW problems we present the following. We first identify a representative two-user problem that considers lossy sender-to-user transmission links as well as blind side information. For this problem, we demonstrate that in addition to hybrid coding (where XORing uncoded bits of a subset of messages with random combinations of the others played a key role), quite surprisingly XORing the same uncoded bits more than once (i.e., repetition of uncoded bits) can increase the achievable rate. Equipped with this observation, we then proceed to construct a coding scheme that leverages both hybrid codes and repetition of uncoded message bits in order to establish an achievable rate region, and we demonstrate numerically that such a scheme can only result in a larger achievable rate region.

To summarize, the main contributions are as follows:

1) We introduce the Blind Index Coding problem, which generalizes classic Index Coding by considering uncertainty (blindness) at the sender about side information given to users.

2) We derive a novel outer bound on the capacity region of BIC problems which leverages a strong data processing inequality to account for the blindness of the sender.

3) We propose a class of hybrid coding schemes, which XOR random linear combinations from one subset of messages with uncoded bits of another subset.

4) We further generalize the problem to better model wireless settings by studying how lossy sender-to-user links affect the efficacy of the hybrid coding schemes, and we find that repetition coding can enhance the performance of hybrid codes.

This remainder of the paper is organized in the following way. In Section II we formally state the BIC problem first for error-free broadcast and then for broadcast over lossy channels. In Section III we motivate both the notion of hybrid coding and our outer bound using a simple example, for which the inner and outer bounds meet. In Section IV we state and prove the general outer bound for K-users. In Section V we define a hybrid coding scheme and study the achievable symmetric rate for the three-user BIC, and in Section VI we numerically compare achieved rates to the derived outer bounds. In Section VII we consider blind index coding when the sender-to-user links occur over wireless channels. Concluding remarks and open questions are presented in Section VIII.

II. THE BLIND INDEX CODING PROBLEM

In this section, we now formally define the Blind Index Coding problem by stating the network and side information models, and formalizing the notion of capacity.

Network model: As shown in Figure 1 in a BIC problem $K$ users each request a message from a server; i.e., User $i$, for $i = 1, \ldots, K$, desires the $m_i$-bit message $\vec{w}_i$, which is drawn uniformly from a space $\{0, 1\}^{m_i}$. Each user, $i$, has access to side information, $\vec{\phi}_{ij}$, (whose form is described later) about each message $\vec{w}_j$, except the one it desires (i.e., for all $j \neq i$). The sender aims to communicate all messages to respective users via a common error-free channel. The goal of the problem is to design a channel input vector, $\vec{x}$, of minimum length, such that each user can decode its desired message.

Side information model: In a blind index coding problem, each side information signal, $\vec{\phi}_{ij}$, is a random fraction of the bits that make up the message, $\vec{w}_j$. We assume that the sender is “blind” in the sense that it is only aware of the average number of bits in each side information signal.

More specifically, we can model the side information in the following way. Let $\vec{g}_{ij}$ be a length-$m_j$ binary vector drawn i.i.d from a Bernoulli$(1 - \mu_{ij})$ distribution. Side information $\phi_{ij}[\ell] = (\phi_{ij}[1], \phi_{ij}[2], \ldots, \phi_{ij}[m_j])$ is such that, for $\ell = 1, \ldots, m_j$, 

$$
\phi_{ij}[\ell] = g_{ij}[\ell] \cdot w_{ij}[\ell].
$$

(1)

User $i$ knows $\vec{g}_{ij}$ for all $j \neq i$, however the sender is only aware of parameters, $\{\mu_{ij}\}$, which govern the probabilistic behavior of the side information. Note that the side information model is equivalent to either 1) randomly sampling bits of a message, or 2) passing a message through a side information channel which is an erasure channel.
Remark 1. The key difference between our BIC formulation and a classic index coding problem of [4, 6] lies in the uncertainty in message bits given as side information. Notably, if we consider the error-free broadcast scenario and \(\mu_{ij} \in \{0, 1\}\) for all \(i, j\), then side information availability deterministic and known to the sender, and our formulation is identical to [6]. Thus, BIC generalizes classic index coding.

**Capacity Region:** We now consider a BIC problem with \(K\) users and side information parameters \(\{\mu_{ij}\}\) as defined above. For this problem, a \((r_1, r_2, \ldots, r_K)\) scheme with block length \(n\) consists of an encoding function and \(K\) decoding functions.

The encoding function, \(f_{enc}^{(n)}: \{0, 1\}^m \rightarrow \{0, 1\}^n\), uses the knowledge of \(\{\mu_{ij}\}\) for all \(j \neq i\) to map each of \(K\) messages (with message \(\vec{w}_j\) consisting of \(m_j\) bits such that \(\lim_{n \rightarrow \infty} \frac{m_j}{n} = r_j\)) onto a length-\(n\) binary vector, \(\vec{x}\), that is broadcast to all \(K\) users using \(n\) channel uses. We reemphasize that the encoding function relies only on the side information parameters, \(\{\mu_{ij}\}\), and not the side information signals, \(\{\phi_{ij}\}\).

The decoding function applied by User \(i\), \(f_{dec,i}^{(n)}: \{0, 1\}^n \times \prod_{j \neq i} \{0, 1\}^{m_j} \rightarrow \{0, 1\}^{m_i}\), maps the broadcast signal, \(\vec{z}\), as well as \(K - 1\) side information vector pairs, \((\vec{\phi}_{ij}, \vec{g}_{ij})\) for all \(j \neq i\), to an estimate of its desired message, \(\hat{w}_i\).

We say that a rate tuple \((r_1, \ldots, r_K)\) is achievable if there exists a sequence of \((r_1, \ldots, r_K)\) coding schemes with increasing block length, \(n\), such that for every \(i \in \{1, \ldots, K\}\)

\[
\lim_{n \rightarrow \infty} \Pr[\hat{w}_i \neq w_i] = 0. \tag{2}
\]

The capacity region is defined as the closure of the set of all rate tuples \((r_1, \ldots, r_K)\) that are achievable.

The goal of this paper is to study the capacity region of the BIC problem. As we show in Appendix D, the capacity region of a 2-user BIC problem is easy to characterize. Thus, in order to gain a better intuition on BIC problems beyond 2 users, we focus in particular on the 3-user BIC problem.

III. Motivating Example

In this section we motivate both the proposed coding schemes and outer bound using a simple, concrete example. Consider a BIC with three users (i.e., \(K = 3\)) and where Users 2 and 3 have full side information about other users’ messages, while User 1 only knows a third of each of \(\vec{w}_2\) and \(\vec{w}_3\) (i.e., \(\mu_{12} = \mu_{13} = \frac{2}{3}\) and \(\mu_{21} = \mu_{23} = \mu_{31} = \mu_{32} = 0\)). For this particular BIC problem, we will determine the symmetric capacity (i.e., the maximum rate \(r\) such that \(r_1 = r_2 = r_3 = r\) is achievable) by assuming the lengths of all messages are the same (i.e., \(m_1 = m_2 = m_3 = m\) where \(m\) is large), proposing a scheme, and introducing a method to bound the capacity region.

For the sake of comparison, we will first establish a baseline achievable symmetric rate by considering random coding, an often-used approach to coding in the presence of uncertain side information. For example, one natural scheme would be to send random linear combinations of all message bits (i.e., parity bits to supplement side information) over the shared channel, until each user has a sufficient number of linearly independent equations (including side information) to decode all of the messages. For this example, by sending \(m(1 + \mu_{12} + \mu_{13}) + o(m) = \frac{7m}{3} + o(m)\) random parities, each user has at least \(3m + o(m)\) equations for \(3m\) unknowns, meaning that with high probability each user can linearly decode all three messages: random coding achieves \(r_{sym} = \frac{7}{9}\).

We now demonstrate how we can construct a "hybrid coding scheme" by combination of uncoded bits and randomly coded parities to go beyond the rate of \(\frac{7}{9}\). In these hybrid schemes, during each phase of transmission, a subset of messages are randomly coded, and then these are XORed with uncoded bits from a disjoint subset of the messages. For this example, we only require two such phases: In the first phase, each channel input is generated by XORing a random combination of \(\vec{w}_1\) bits, a single uncoded \(\vec{w}_2\) bit, and a single uncoded \(\vec{w}_3\) bit. Each uncoded bit from both \(\vec{w}_2\) and \(\vec{w}_3\) are used only once to generate an input, and thus the first phase consists of exactly \(m\) channel inputs generated in this manner. Formally, for each \(\ell = 1, \ldots, m\), the sender broadcasts \(\vec{c}[\ell] \oplus w_1[\ell] \oplus w_2[\ell] \oplus w_3[\ell]\), where \(\vec{c}[\ell]\) is an length-\(m\) i.i.d. random binary vector. In the second phase, we send \(\frac{8m}{9}\) RLCs of only \(\vec{w}_1\) bits.

Notice that with this scheme, if each user decodes its desired message with error probability vanishing as \(m\) grows large, we achieve rate of \(r_{sym} = \frac{2}{3}\), which is higher than the \(\frac{7}{9}\) achieved through random coding. We now explain why with this scheme such a rate is achievable by explaining how each user decodes its desired message:

User 1:

Notice that during the first phase, for each channel input \(\ell \in \{1, \ldots, m\}\), there is a probability of \((1 - \mu_{12})(1 - \mu_{13}) = \frac{1}{9}\) that User 1 knew both \(w_2[\ell]\) and \(w_3[\ell]\). In such an event, User 1 can cancel \(w_2[\ell] \oplus w_3[\ell]\) and received a “clean” RLC of \(\vec{w}_1\) bits. Therefore, during the first phase User 1 receives (approximately) \(\frac{m}{9}\) such RLCs. In the second phase we supplemented this with an additional \(\frac{8m}{9}\) RLCs of only \(\vec{w}_1\). When combined, at the end of transmission User 1 will be able to identify in total \(m\) linear equations describing the \(m\) desired bits of \(\vec{w}_1\).

User 2:

User 2 already knows all of \(\vec{w}_1\) and \(\vec{w}_3\) and therefore can remove their contributions from each channel input of the first phase. Thus, after canceling the undesired message contributions, User 2 receives exactly the \(m\) bits of \(\vec{w}_2\).

1In the subsequent explanation, we omit the \(o(m)\) to simplify the exposition.
Fig. 2. Illustration of the symmetric-capacity-achieving scheme of the example. The horizontal axis provides scale representation of the number of channel uses dedicated to each phase. Transmission type is illustrated using outlined (uncoded) or shaded (randomly coded) blocks. Notice that, because the sender is blind, parts of messages \( \vec{w}_2 \) and \( \vec{w}_3 \) that are known to User 1 cannot be explicitly aligned as discussed in Remark 2 and thus some parts of \( \vec{w}_1 \) is XORed with \( \vec{w}_2 \) as well as some of \( \vec{w}_2 \) is XORed with \( \vec{w}_3 \). These are displayed as contiguous blocks in the figure for clarity, but in reality would be interleaved throughout the first \( m \) channel uses.

User 3: User 3 already knows all of \( \vec{w}_1 \) and \( \vec{w}_2 \) and therefore can remove their contributions from each channel input of the first phase. Thus, after canceling the undesired message contributions, User 3 receives exactly the \( m \) bits of \( \vec{w}_3 \).

The key intuition on why we XOR uncoded bits of some messages with coded RLCs of others is as follows: Assume our objective is to create an input signal such that User 1 can use side information to cancel "interference" from undesired messages \( \vec{w}_2 \) and \( \vec{w}_3 \). As \( m \) grows large the probability that User 1 can cancel a random combination of \( \vec{w}_2 \) or \( \vec{w}_3 \) vanishes, and thus RLCs of \( \vec{w}_2 \) and \( \vec{w}_3 \) are useless with respect to the objective. However, by XORing uncoded \( \vec{w}_2 \) and \( \vec{w}_3 \) bits, the probability that User 1 can exploit side information to cancel interference remains constant regardless of \( m \). It is worth noting that while Users 2 and 3 eventually know all three messages (through side information and decoding their desired messages), User 1 ends up knowing only parts of messages \( \vec{w}_2 \) and \( \vec{w}_3 \).

**Remark 2.** To obtain further intuition, we can also interpret the proposed scheme as a form of interference alignment. Let \( \vec{w}_i^+ \) and \( \vec{w}_i^- \) for \( i = 2, 3 \) denote subvectors of \( \vec{w}_i \) known and unknown respectively to User 1 via side information. Note that the lengths of \( \vec{w}_i^+ \) and \( \vec{w}_i^- \) are approximately \( \frac{m}{3} \) and \( \frac{2m}{3} \), respectively.

First, consider what strategy the sender could use if it was not blind and could identify these subvectors. It could first send RLCs of \( \vec{w}_2^+ \) and \( \vec{w}_3^- \), and \( \vec{w}_2^- \) bits, knowing that User 1 can cancel the \( \vec{w}_2^+ \) and \( \vec{w}_3^- \) contribution for every such input. Specifically, the non-blind sender aligns the bits User 1 can cancel (\( \vec{w}_2^+ \) and \( \vec{w}_3^- \)), as well as the bits it cannot cancel (\( \vec{w}_2^- \) and \( \vec{w}_3^+ \)). Via an equation counting argument, it is easy to verify that such a scheme achieves a higher rate of \( \frac{2}{3} \).

When the sender is blind, it is unable to distinguish \( \vec{w}_2^+ \) from \( \vec{w}_2^- \) and \( \vec{w}_3^- \) from \( \vec{w}_3^+ \), however we would still like to efficiently send both \( \vec{w}_2 \) and \( \vec{w}_3 \) simultaneously. Therefore our scheme achieves such alignment probabilistically, by using uncoded bits from \( \vec{w}_2^+ \) (\( \vec{w}_3^- \)) to preserve the separation between \( \vec{w}_2^- \) and \( \vec{w}_2^+ \) (\( \vec{w}_3^+ \) and \( \vec{w}_3^- \)). Figure 2 highlights the two desired interference alignment cases, as well as the transmissions where alignment fails due to the sender being blind.

As this example demonstrates, we wish to send as few bits as possible from \( \vec{w}_2 \) and \( \vec{w}_3 \) that are not known as side information at User 1. However, these bits must also be communicated to their respective users. Any converse must somehow capture this tension between communicating all of \( \vec{w}_2 \) and \( \vec{w}_3 \) to their respective users, and minimizing their impact at User 1. In the next section, we will present an outer bound which captures this tension, and for the specific example presented in this section is tight and is expressed as (using the notation of Figure 2):

\[
H(\vec{x}|\vec{w}_1, \vec{w}_2^+, \vec{w}_3^+) \geq \frac{2}{3} H(\vec{x}|\vec{w}_1, \vec{w}_3^+) + \frac{1}{3} H(\vec{x}|\vec{w}_1, \vec{w}_2, \vec{w}_3^+).
\]

The above inequality lower bounds the interference at User 1 (i.e., the contribution of unknown parts of \( \vec{w}_2 \) and \( \vec{w}_3 \)) with a convex combination of terms that either represent providing none of \( \vec{w}_2 \) as side information \( H(\vec{x}|\vec{w}_1, \vec{w}_2^+) \) or all of \( \vec{w}_2 \) as side information \( H(\vec{x}|\vec{w}_1, \vec{w}_3^+) \)). The coefficient weights that describe the combination are a function of the side information parameter \( \mu = \frac{2}{3} \). A more general form of this inequality is the key lemma used to construct the outer bound.

Before concluding the section, we point out that this inequality is valid only when the sender is blind. Indeed, if we consider a non-blind sender like in Remark 2 we see that such a sender may use its knowledge of the side information to create a transmission that invalidates the assertion. This is addressed more specifically for the general form of the assertion (i.e., Lemma 2) in Remark 5 of the subsequent section.

\(^2\)We focus on User 1’s ability to cancel contributions of other messages, since by assumption User 2 and 3 have full knowledge of undesired messages and can cancel any such interference perfectly.
IV. Outer Bound

In this section, we present an outer bound on the capacity region of the BIC problem. We will first state and prove the bound for the 3-user setting and remark on its implications. We then introduce a key lemma and prove the 3-user outer bound. Finally we state a general expression for an outer bound on the general $K$-user BIC capacity region. Its proof is relegated to Appendix C.

A. 3-user Outer Bound

We begin by stating the following result:

**Theorem 1.** Consider a 3-user BIC problem. Rates $(r_1, r_2, r_3)$ are achievable only if,

$$r_i + \mu_{ij} r_j + \left( \frac{\mu_{ik} - [\mu_{ij} - \mu_{jk}][\mu_{ik} - \mu_{kj}]}{1 - \mu_{kj}} \right) r_k \leq 1,$$

for any $i \neq j \neq k \in \{1, 2, 3\}$ and $[a]^+ \triangleq \max\{a, 0\}$.

**Remark 3.** If the sender is not blind (i.e., the side information is known), our BIC problem can be converted to an analogous classic index coding problem with each user, $i$, desiring four different messages whose rates sum to by $r_i$ and whose proportion are determined by $\mu_{ij}$ and $\mu_{ki}$ for $i \neq j \neq k$.

For this resulting classic index coding problem, using the coding techniques of [12], it can be shown that rate tuples $(r_1, r_2, r_3)$ satisfying, for all $i \neq j \neq k \in \{1, 2, 3\}$,

$$r_i + \mu_{ij} r_j + \mu_{ik} \mu_{jk} r_k \leq 1$$

are achievable. Notice that for some side information parameters (e.g., when $\mu_{23} = \mu_{32} = 0$ and $\mu_{1j} > 0$ for $j = 2, 3$), the rates achieved by a non-blind sender can be greater than the BIC outer bound, (3). The key difference in expressions is the third term on the left side of (3), which captures (at least partially) the capacity loss due to sender blindness.

B. Proof of Theorem 1

To prove Theorem 1, we start by stating and proving a key lemma:

**Lemma 2.** Consider a BIC problem with side information parameters $\{\mu_{ij}\}$. Then, for any $(r_1, \ldots, r_K)$ scheme with block length $n$ and any random variable $V$ that is independent of $\tilde{w}_j$ and $\tilde{g}_j$ with $i \neq j$ (but may depend on other messages and channel parameters), we have

$$H \left( \tilde{x} \bigg| \phi_{ij}, \tilde{g}_j, V \right) \geq \mu_{ij} H \left( \tilde{x} | V \right) + (1 - \mu_{ij}) H \left( \tilde{x} | \tilde{w}_j, V \right).$$

(4)

Additionally, if $\mu_{kj} \leq \mu_{ij}$ where $i \neq j \neq k$, then

$$H \left( \tilde{x} \bigg| \phi_{ij}, \tilde{g}_j, V \right) \geq \frac{\mu_{ij} - \mu_{kj}}{1 - \mu_{kj}} H \left( \tilde{x} | V \right) + \frac{1 - \mu_{ij}}{1 - \mu_{kj}} H \left( \tilde{x} \bigg| \phi_{kj}, \tilde{g}_j, V \right).$$

(5)

**Remark 4.** Inequality (4) captures an intuition that can be illustrated through the following toy problem. Consider a scenario where the sender has 4 bits $b_1, b_2, c_1, c_2$. It knows that User 2 knows $c_2$ already and User 3 knows $b_1$ and $b_2$. On the other hand, the sender only knows that User 1 knows either $b_1$ or $b_2$ (but not both) and either $c_1$ or $c_2$ (but not both) and that both of these uncertainties are the result of a (fair) coin flip. If the sender sends a single transmission such that both User 2 and User 3 learn something new about $b_1, b_2, c_1, c_2$, what is the minimum probability that User 1 also learns something new? One possible transmission would be to send $b_1 \oplus c_1$. In this case, User 2 learns $c_1$ and User 3 learns $b_1$, and there is a 75% chance that User 1 learns either $b_1$, $c_1$, or $b_1 \oplus c_1$. In comparison, we can evaluate (4) for the porposed transmission by letting $i = 1$, $j = 2$, $k = 3$, $\mu_{12} = \mu_{13} = \frac{1}{2}$, and assuming $\tilde{w}_2 = [b_1, b_2]$, $\tilde{w}_3 = [c_1, c_2]$, and $V = (\phi_{13}, \tilde{g}_{13})$. In doing so, we see that the right hand side of (4) evaluates to $\mu_{12}(1) + (1 - \mu_{12})(\mu_{13}) = \frac{3}{4}$, signifying that the 75% chance of “leaking” information to User 1 is the lowest possible.

Notice that, as states, Lemma 2 does not assume decodability of any message. Moreover, it applies regardless of the number of channel uses, whereas the toy example assumed only a single channel use. Consequently, Lemma 2 can be viewed as a powerful extension of the intuition from the toy example to vector (i.e., coded) representations of message bits.

**Remark 5.** One can note that if the transmitter is not blind, the sender can construct a signal that invalidates 4. For example, consider the scenario in Remark 4 but now assume that the sender is aware that User 1 knows $b_1$ and $c_1$. The sender can now use this knowledge to send a single transmission $b_1 \oplus c_1$. One can easily verify that, for this one transmission, the left hand side now evaluates to 0, while the right hand side evaluates to $\frac{1}{2}$ which violates the claim. Therefore, the inequality specifically captures the impact of a blind sender.
Remark 6. Inequality [4] can more generally be interpreted as follows. Note that $H(x|\bar{x})$ corresponds to the case that there is no side information about $\vec{w}_j$ provided in the conditioning, and $H(x|\bar{x}, \vec{w}_j)$ corresponds to the case that all of $\vec{w}_j$ is provided as the side information in the conditioning. Therefore inequality [4] to lower bounds $H(\bar{x} | \vec{\phi}_{ij}, \vec{g}_{ij}, V)$ with a weighted average of two extreme cases, where either none or all of $\vec{w}_j$ is provided as side information. A similar interpretation holds for (5), where $\vec{w}_j$ is replaced with $(\vec{\phi}_{kj}, \vec{g}_{kj})$.

Proof: To prove Lemma 2, we first define a virtual side information signal, $\vec{\phi}$, such that $\vec{\phi}_{ij}$ is a physically degraded version of $\vec{\phi}$. To do so, we also specify two channel state sequences, $\vec{g}$ and $\vec{g}'$ drawn i.i.d from two different Bernoulli distributions that take a value of zero with probabilities $\mu'$ and $\delta = \frac{\mu - \mu'}{1 - \mu'}$, respectively. The side information signals are constructed such that for $\ell \in \{1, \ldots, m_j\}$,

$$H(\bar{x}|\vec{\phi}_{ij}, \vec{g}_{ij}, V) = - I(\vec{\phi}_{ij}, \vec{g}_{ij} : \bar{x}|V) + H(\bar{x}|V)$$

$$\geq - s^*(\vec{\phi}' : (\vec{\phi}_{ij}, \vec{g}_{ij})) I(\vec{\phi}', \vec{g}': \bar{x}|V) + H(\bar{x}|V)$$

$$\geq (a) \frac{1 - \mu_{ij}}{1 - \mu'} H(\bar{x}|V) + \frac{1 - \mu_{ij}}{1 - \mu'} H(\bar{x}|V, \vec{\phi}', \vec{g}')$$.

(7)

Step (a), where we evaluated $s^*(\vec{\phi} : (\vec{\phi}_{ij}, \vec{g}_{ij}))$, is proven in Appendix A. Recall that we only require $\mu' \leq \mu_{ij}$ in order for the virtual signal to be properly defined, and we notice the following to complete the proof:

- If $\mu' \geq 0$, then $(\vec{\phi}', \vec{g}') = (\vec{w}_j, \bar{1})$ and we prove (4).
- If $\mu' \geq \mu_{ij}$, then $(\vec{\phi}', \vec{g}')$ is statistically equivalent to $(\vec{\phi}_{ij}, \vec{g}_{ij})$ and we prove (5).

We now use Lemma 2 to prove Theorem 1. First, we note that two side info parameter relationships affect the form of (5): The term $[\frac{\mu_{ij} - \mu_{ij}'}{1 - \mu_{ij}}]$ is nonzero only if both $\mu_{ij} < \mu_{ij}$ and $\mu_{jk} < \mu_{ik}$. In this case,

$$r_i + \mu_{ij} r_j + \frac{(\mu_{ik} - \mu_{ij})(\mu_{ik} - \mu_{jk})}{1 - \mu_{kj}} r_k \leq 1.$$  

(8)

Otherwise, if either $\mu_{kj} \geq \mu_{ij}$ or $\mu_{jk} \geq \mu_{ik}$, then

$$r_i + \mu_{ij} r_j + \mu_{ik} r_k \leq 1.$$  

(9)

We prove these two cases separately, and only address the first case, (8), here. The proof of (9) will use similar techniques, and may be found in Appendix B. We therefore assume $\mu_{kj} < \mu_{ij}$ and $\mu_{jk} < \mu_{ik}$, and start with Fano’s inequality at User $i$:

$$nr_i \leq I(\bar{x}, \vec{\phi}_{ij}, \vec{g}_{ij}, \vec{\phi}_{ik}, \vec{g}_{ik}; \vec{w}_j) + o(n)$$

$$= H(\bar{x} | \vec{\phi}_{ij}, \vec{g}_{ij}, \vec{\phi}_{ik}, \vec{g}_{ik}) - H(\bar{x} | \vec{w}_i, \vec{\phi}_{ij}, \vec{g}_{ij}, \vec{\phi}_{ik}, \vec{g}_{ik}) + o(n)$$

$$\leq n - H(\bar{x} | \vec{w}_i, \vec{\phi}_{ij}, \vec{g}_{ij}, \vec{\phi}_{ik}, \vec{g}_{ik}) + o(n)$$

(10)

$$\leq n - \frac{\mu_{ij} - \mu_{kj}}{1 - \mu_{kj}} H(\bar{x} | \vec{w}_i, \vec{\phi}_{kj}, \vec{g}_{kj}, \vec{\phi}_{ik}, \vec{g}_{ik}) - \frac{1 - \mu_{ij}}{1 - \mu_{kj}} H(\bar{x} | \vec{w}_i, \vec{\phi}_{kj}, \vec{g}_{kj}, \vec{\phi}_{ik}, \vec{g}_{ik})$$.

(11)
where in step (a) we applied (4) from Lemma 2 by letting \( V = (\tilde{w}_i, \tilde{\phi}_{ik}, \tilde{g}_{ik}) \). Notice there are two negative entropy terms, \( A \) and \( B \), to account for. To address the quantity \( A \), we enhance side information at User \( j \) from \( (\tilde{\phi}_{ji}, \tilde{g}_{ji}) \) to \( \tilde{w}_i \), and observe:

\[
nr_j \leq I(\tilde{x}, \tilde{\phi}_j, \tilde{g}_{jk}, \tilde{\phi}_{jk}; \tilde{w}_j) + o(n)
\]

\[
\leq I(\tilde{x}, \tilde{w}_i, \tilde{\phi}_{jk}, \tilde{g}_{jk}; \tilde{w}_j) + o(n)
\]

\[
= H(\tilde{x}|\tilde{w}_i, \tilde{\phi}_{jk}, \tilde{g}_{jk}) - H(\tilde{x}|\tilde{w}_i, \tilde{\phi}_{jk}, \tilde{g}_{jk}) + o(n)
\]

\[
\leq H(\tilde{x}|\tilde{w}_i, \tilde{\phi}_{jk}, \tilde{g}_{jk}) - \mu_{jk}H(\tilde{x}|\tilde{w}_i; \tilde{w}_j) + o(n),
\]

in (b) we used (5). Also, like (13), we find

\[
\ell, i \in \{1, \ldots, K\}
\]

\[
\text{levels has 2 children and each node in the } K - 1 \text{-th level has one child. The label of the } i \text{-th node in level } \ell \text{ is denoted as } v[\ell], i \in \{1, \ldots, K\}, \text{ where if } \ell < K \text{ then } i \in \{1, \ldots, 2^{\ell - 1}\} \text{ and if } \ell = K \text{ then } i \in \{1, \ldots, 2^{\ell - 2}\}. \text{ The index } i \text{ specifies the precise location in the level: nodes } i = 2j - 1 \text{ and } i = 2j \text{ in level } \ell < K \text{ are the left and right children, respectively, of a node } j \text{ in level } \ell - 1. \text{ Node } i \text{ in level } K \text{ is the sole child of node } i \text{ in level } K - 1. \text{ Finally, the labels of an OBT must satisfy the following:}
\]

1) For any path from the root node of the tree to any leaf node, no labels are repeated.
2) Any two nodes with the same parent cannot have the same label.

The first requirement is equivalent to saying that the sequence of labels along any path from root to leaf is a permutation of \( \{1, \ldots, K\} \). This is demonstrated in Figure 3 where we provide an example of an 4-user OBT.

\[
\text{To account for the quantity } B, \text{ we observe}
\]

\[
nr_k = nr_k - n(1 - \mu_{ik})r_k
\]

\[
\leq I(\tilde{x}, \tilde{\phi}_{ki}, \tilde{g}_{ki}, \tilde{\phi}_{kj}, \tilde{g}_{kj}; \tilde{w}_k) - n(1 - \mu_{ik})r_k + o(n)
\]

\[
\leq I(\tilde{x}, \tilde{w}_i, \tilde{\phi}_{kj}, \tilde{g}_{kj}, \tilde{\phi}_{jk}, \tilde{g}_{jk}; \tilde{w}_k) - n(1 - \mu_{ik})r_k + o(n)
\]

\[
= H(\tilde{x}|\tilde{w}_i, \tilde{\phi}_{kj}, \tilde{g}_{kj}, \tilde{\phi}_{jk}, \tilde{g}_{jk}) - H(\tilde{x}|\tilde{w}_i, \tilde{\phi}_{kj}, \tilde{g}_{kj}, \tilde{\phi}_{jk}, \tilde{g}_{jk}) + I(\tilde{\phi}_{ik}, \tilde{g}_{ik}; \tilde{w}_k)
\]

\[
= H(\tilde{x}|\tilde{w}_i, \tilde{\phi}_{kj}, \tilde{g}_{kj}, \tilde{\phi}_{jk}, \tilde{g}_{jk}) - H(\tilde{x}|\tilde{w}_i, \tilde{\phi}_{kj}, \tilde{g}_{kj}, \tilde{\phi}_{jk}, \tilde{g}_{jk}) + o(n)
\]

\[
\leq H(\tilde{x}|\tilde{w}_i, \tilde{\phi}_{kj}, \tilde{g}_{kj}, \tilde{\phi}_{jk}, \tilde{g}_{jk}) - \mu_{kj}H(\tilde{x}|\tilde{w}_i, \tilde{w}_k) + o(n),
\]

in (d) we used (5). Also, like (13), we find

\[
\ell, i \in \{1, \ldots, K\}
\]

\[
nr_j \leq H(\tilde{x}|\tilde{w}_i, \tilde{w}_j) + o(n).
\]
We now state the following outer bound for the $K$-user BIC:

**Theorem 3.** Consider a $K$-user BIC with $K \geq 3$, defined by parameters $\{\mu_{ij}\}$. The rate tuple $(r_1, \ldots, r_K)$ is achievable only if it satisfies,

$$
\Gamma_A[1, 1] \leq 1,
$$

for any $K$-user OBT, where

$$
\Gamma_A[\ell, i] = \begin{cases} 
\frac{r_v[\ell, i] + \zeta[\ell, i] \Gamma_A[\ell + 1, 2i - 1] + (1 - \zeta[\ell, i]) \Gamma_B[\ell + 1, 2i]}{\Gamma_A[\ell + 1, 2i - 1] + (1 - \zeta[\ell, i]) \Gamma_B[\ell + 1, 2i]} & \text{if } \ell < K - 1 \\
0 & \text{if } \ell = K - 1,
\end{cases}
$$

$$
\Gamma_B[\ell, i] = \begin{cases} 
\frac{r_v[\ell, i] + \zeta[\ell, i] \Gamma_A[\ell + 1, 2i - 1] + (1 - \zeta[\ell, i]) \Gamma_B[\ell + 1, 2i]}{\Gamma_A[\ell + 1, 2i - 1] + (1 - \zeta[\ell, i]) \Gamma_B[\ell + 1, 2i]} & \text{if } \ell < K - 1 \\
0 & \text{if } \ell = K - 1,
\end{cases}
$$

$$
\zeta[\ell, i] = \begin{cases} 
\eta_v[\ell + 1, 2i - 1] & \text{if } \ell < K - 1 \\
0 & \text{if } \ell = K - 1,
\end{cases}
$$

where we have, if $\ell > 1$,

$$
\eta_j[\ell, i] = \begin{cases} 
1 & \text{if } j = v[\ell, i], \text{ } i \text{ is odd} \\
0 & \text{if } j = v[\ell - 1, \left\lceil \frac{i}{2} \right\rceil] \\
\min \{\mu_{v[\ell, i], j}, \eta_j[\ell - 1, \left\lceil \frac{i}{2} \right\rceil]\} & \text{otherwise}
\end{cases}
$$

and if $\ell = 1$

$$
\eta_j[1, 1] = \begin{cases} 
1 & \text{if } j = v[1, 1] \\
\mu_{v[1, 1], j} & \text{otherwise}
\end{cases}
$$

The proof of Theorem 3 may be found in Appendix C. Here we remark on how the intuitions from Theorem 1 are extended to $K$ users.

**Remark 7.** Consider the 3-user bound with respect to the more general statement of Theorem 3. Figure 4 depicts the exact assignment of labels for a 3-user OBT that results in Theorem 1.

Recall that the construction of the outer bound in Theorem 1 began with applying Fano’s inequality at User $i$ (i.e., the root node label of the OBT), and then applying (5) of Lemma 2. Applying (5) resulted in two terms $A$ and $B$ in (11), each of which was canceled by analysis of a different user with enhanced side information. This is reflected in the first case of (17), where in addition to the rate of the user associated with the node label, we have the quantities $\Gamma_A[\cdot]$ and $\Gamma_B[\cdot]$ associated with the expressions that will cancel $A$ and $B$, respectively. The scaling terms $\zeta[\ell, i]$ reflect the appropriate scaling terms needed for the cancellation; e.g., consider the final step in the proof of Theorem 1 where we took a weighted sum of (11)–(15). The last quantity, $\eta_j[\ell, i]$, tracks the side information enhancement through each level of recursion.

**Remark 8.** It is worth noting that the terms associated with the $K - 1$-th layer of the OBT are special: This layer represents the “base case” of the recursion, and in the 3-user scenario, we reached this base case after only one application of (5). At the $K - 1$-th layer, instead of (5) we apply (4) which is reflected by the associated value of $\zeta[\ell, i]$ in (19).
Remark 9. Consider \( r \) is larger, then (24) and (25) simplify to

\[
r_i + \sum_{j \neq i} \mu_{ij} r_j \leq 1.
\]

In some cases, this suffices to achieve the full capacity region. For example, the rate region achievable by conventional random coding achievable rate region matches the outer bounds given by Theorem 3 in the following two scenarios (these are formally stated and proven as Propositions 9 and 10 in Appendix D):

- 2-user BICs, for any value of \( \mu_{12} \) and \( \mu_{21} \).
- Symmetric \( K \)-user BICs, where \( \mu_{ij} = \mu \) for all \( i \neq j \).

However, conventional random coding is not optimal in general, and in the rest of this section we propose a new hybrid encoding strategy that XORs random linear combinations of all bits from some messages with uncoded bits from others. This hybrid between random coding and uncoded transmission is the key mechanism to blindly exploit side information. For simplicity, we focus on symmetric rates achievable in an arbitrary 3-user BIC. We will first state the achievable symmetric rate in a theorem, then describe the encoding and decoding strategies before finally proving that the symmetric rate claimed in the theorem is indeed achievable.

### A. 3-user BIC Hybrid Coding

We now state symmetric rate achievable using hybrid coding. We then define a hybrid encoding scheme for three-user BIC problems, using the key points from the motivating example.

**Theorem 4.** Consider a 3-user BIC problem, defined by parameters \( \{\mu_{ij}\} \), where WLOG user indices are such that

\[
\mu_{32} \leq \mu_{23} \leq \max(\mu_{11}, \mu_{13}),
\]

for either \( i \in \{2, 3\} \). Any \( r_{\text{sym}} \) satisfying the following is achievable:

\[
r_{\text{sym}} \leq \min \left\{ \frac{1}{1 + \mu_{21} + \mu_{23}}, \frac{1}{1 + \mu_{31} + \mu_{32}} \right\},
\]

\[
r_{\text{sym}} \leq \max \left\{ \frac{1}{1 + \mu_{23} + \mu_{12} + \mu_{13}(1 - \mu_{23} + \mu_{32})(1 - \mu_{12})}, \frac{1}{1 + \mu_{12} + \mu_{13}} \right\}.
\]

**Remark 9.** Consider \( r_{\text{sym}} \) satisfying (24) and (25). In the right hand side of (25), if the second term within the maximization is larger, then (24) and (25) simplify to \( r_{\text{sym}} \leq \min \left\{ \frac{1}{1 + \mu_{21} + \mu_{23}}, \frac{1}{1 + \mu_{31} + \mu_{32}} \right\} \). In this case, from (22) it is clear that conventional random coding suffices to achieve the desired rate. Hence, our hybrid coding scheme increases the symmetric rate only if the first term in the max of (25) is larger. Additionally, since conventional random coding suffices in the first case, to prove Theorem 4 we need only to describe a scheme and proves achievability of \( r_{\text{sym}} \) satisfying

\[
r_{\text{sym}} \leq \min \left\{ \frac{1}{1 + \mu_{21} + \mu_{23}}, \frac{1}{1 + \mu_{31} + \mu_{32}} \right\}.
\]

We now define our hybrid coding scheme where, for any \( r_{\text{sym}} \) satisfying (26), the sender will communicate \( m = nr_{\text{sym}} - \delta_n \) bits where \( \delta_n \) is chosen such that \( \delta_n = o(n) \) to each receiver in \( n \) channel uses, such that probability of error vanishes as \( n \) grows to infinity.\(^3\)

\(^3\)For any three users, such a condition must hold for at least one permutation of indices.

\(^4\)The \( o(n) \) term 1 accounts for the fact that \( m \) must be integer, and 2) as we shall see, ensures that decoding error will vanish as \( n \) grows large.
Fig. 5. Hybrid coding scheme for 3-user BIC, where \( N \) generate boxes represent RLCs of a single message.

**Encoding:** The hybrid coding scheme is characterized by three parameters, \( N_1, N_2, \) and \( N_3 \). For each \( i \in \{1, 2, 3\} \), we generate \( N_i \) random linear combinations (RLC) of the bits only in \( \vec{w}_i \), denoted by vector \( \vec{J}_i \). The precise values of \( N_1, N_2, \) and \( N_3 \) are specified later, however we point out as depicted in Figure 5, that \( N_1 - m \geq N_2 \geq N_3 \).

As shown in the figure, the sender combines RLCs and uncoded bits of messages in five phases. During Phase 1, each input is the XOR of one bit from each of \( \vec{J}_1, \vec{J}_2, \) and \( \vec{J}_3 \), where we take bits from each vector sequentially. Phase 1 ends and Phase 2 begins when the bits in \( \vec{J}_3 \) are exhausted (i.e., after \( N_3 \) channel uses). Similarly, the number of channel uses allocated to each phase of transmission are dictated by when we exhaust the bits of a certain type: Phase 2 inputs consist of an XOR of \( \vec{J}_1, \vec{J}_2, \) and \( \vec{w}_3 \) bits, and ends when we have no more bits from \( \vec{J}_2 \). Phase 3 inputs consist of an XOR of \( \vec{J}_1, \vec{w}_2, \) and \( \vec{w}_3 \) bits, and ends when we have no more bits from \( \vec{w}_3 \). Phase 4 inputs consist of an XOR of \( \vec{J}_1 \) and \( \vec{w}_2 \) bits, and ends when we have no more bits from \( \vec{w}_2 \); and Phase 5 inputs consist of only a \( \vec{J}_1 \) bit.

**Decoding:** We now describe the decoding scheme of each user. Users 2 and 3 each decodes all 3 messages. As in conventional random coding, this requires that User 2 and 3 each receive a sufficient number of independent linear combinations of messages bits, either via side information or the shared channel.

A key point in our coding scheme lies in how User 1 exploits the hybrid coding structure to decode \( \vec{w}_1 \). As in the example, User 1 uses side information to cancel out the combinations of known \( \vec{w}_2 \) and \( \vec{w}_3 \) bits from symbols received in Phases 3 and 4. It uses these “clean” RLC of only \( \vec{w}_1 \) bits along with those RLC received during Phase 5 to linearly decode only \( \vec{w}_1 \).

For the scheme to achieve \( r_{\text{sym}} \) (i.e., in order for decoding error probability to vanish as \( n \) grows large), we claim that choosing \( N_1, N_2, \) and \( N_3 \) as

\[
N_1 = n \quad \text{and} \quad N_2 = nr_{\text{sym}} \mu_{23} \quad \text{and} \quad N_3 = nr_{\text{sym}} \mu_{32},
\]

results in a probability of decoding error that vanishes as \( n \to \infty \). We prove this formally in the following subsection.

**Remark 10.** Recall from the illustrative example, we wanted to maximize the chance that User 1 can clean \( \vec{w}_2 \) and \( \vec{w}_3 \) content from a transmission, and we assume that both User 2 and 3 decode all three messages. Thus the phases of transmission in Figure 5 have the following roles: Phase 5 provides RLCs about \( \vec{w}_1 \) to User 1. Phase 5 also provides enough \( \vec{w}_1 \) RLCs for each of User 2 and User 3 to decode \( \vec{w}_1 \) (with the help of side information). A fraction, \( (1 - \mu_{12}) \), of Phase 4 is useful to User 1 after using side information to clean the \( \vec{w}_2 \) component, to obtain a clean RLC of only \( \vec{w}_1 \). Similarly, a fraction, \( (1 - \mu_{12})(1 - \mu_{13}) \), of Phase 3 is useful to User 1 by cleaning both the \( \vec{w}_2 \) and \( \vec{w}_3 \) components, to obtain a clean RLC of \( \vec{w}_1 \). Note that User 1 only uses clean RLCs from Phases 3-5 to decode \( \vec{w}_1 \). Because Users 2 and 3 each decoded \( \vec{w}_1 \) from Phase 5, each cancels out the \( \vec{w}_1 \) component from Phases 1-4, and then each uses the remaining residual symbols to decode both messages \( \vec{w}_2 \) and \( \vec{w}_3 \).

**B. Proof of Achievability**

We now address the achievability of rate \( r_{\text{sym}} \) satisfying (26), using the hybrid network codes we just defined. Before proceeding we recall that the message size \( m \) is such that \( m = nr - \delta_n \), where \( \delta_n \) is positive and \( \delta_n = o(n) \).

To prove that the rate is achievable, we must show that the probability that any user does not decode its desired message vanishes as \( n \to \infty \) (i.e., \( Pr[\vec{w}_i \neq \vec{w}_1] \to 0 \)). Moreover, since our decoding strategy requires that User 2 and User 3 decode all three messages, we also show that the probability of decoding error of all messages at Users 2 and 3 vanishes as \( n \to \infty \). Specifically, we have the following possible error events, each of which must approach 0 as \( n \to \infty \):

\[
\begin{align*}
\mathcal{E}_1 & \quad \text{User 1 fails to decode } \vec{w}_1, \\
\mathcal{E}_2 & \quad \text{User 2 fails to decode } \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}. \\
\mathcal{E}_3 & \quad \text{User 3 fails to decode } \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}.
\end{align*}
\]

For each event, we will separate the error analysis into different sources of error, and for each source of error we will use one of two analysis techniques to prove that the probability of such an event occurring vanishes with large \( n \). In order to provide clarity, and since User 1’s decoding strategy was the primary difference between hybrid coding and conventional random coding, we will revisit these two techniques after first applying them in the context of analyzing the probability of event \( \mathcal{E}_1 \) occurring.

Recall that User 1 first uses its side information to “clean” transmissions from Phases 3 and 4 resulting in random linear combinations (RLCs) of only bits from \( \vec{w}_1 \). It then combines clean RLCs with those received during Phase 5 (recall from...
Figure [5] that Phase 5 only has $\vec{w}_1$ content) and attempts to linearly decode $\vec{w}_1$. Therefore, we express User 1’s decoding error as the union of two events, $\mathcal{E}_1 = \mathcal{E}_{1a} \cup \mathcal{E}_{1b}$, defined as:

$\mathcal{E}_{1a}$: The total number of random linear combinations (RLCs) cleaned from Phases 3 and 4 and received in Phase 5 is less than $m + \delta_n$, where $\delta_n$ grows with $n$ and $0 < \delta_n < \delta_n$.

$\mathcal{E}_{1b}$: The random matrix that describes the transformation of $\vec{w}_1$ to received (clean) RLCs is rank deficient.

We may thus represent receiving a clean RLC in the $\ell$-th channel use of Phase 3 as a Bernoulli$(1-\mu_{12} - \mu_{13} + \mu_{12}\mu_{13})$ random variable $\lambda_3[\ell]$ which is i.i.d. across $\ell = 1, \ldots, m - N_2 + N_3$ and receiving a clean equation in the $\ell$-th channel use of Phase 4 as a Bernoulli$(1-\mu_{12})$ random variable $\lambda_4[\ell]$ which is i.i.d. across $\ell = 1, \ldots, N_2 - N_3$. We now note that the duration of Phases 3 and 4 ($D_3$ and $D_4$) are by construction,

$$D_3 = m - N_2 + N_3 = nr_{sym}(1 - \mu_{23} + \mu_{32}) - \delta_n,$$

$$D_4 = N_2 - N_3 = nr_{sym}(\mu_{23} - \mu_{32}),$$

and that the duration of Phase 5 may be bounded as

$$D_5 = N_1 - N_2 - m$$

$$= n - nr_{sym}(1 + \mu_{23}) + \delta_n$$

$$(a) \geq nr_{sym}(1 + \mu_{23} + \mu_{12} + \mu_{13}(1 - \mu_{23} + \mu_{32})(1 - \mu_{12})) - nr_{sym}(1 + \mu_{23}) + \delta_n$$

$$= nr_{sym}(\mu_{12} + \mu_{13}(1 - \mu_{23} + \mu_{32})(1 - \mu_{12})) + \delta_n$$

$$= nr_{sym}(1 - (1 - \mu_{23} + \mu_{32})(1 - \mu_{12}) - (\mu_{23} - \mu_{32})(1 - \mu_{12}) + \delta_n$$

$$\geq nr_{sym} - D_3(1 - \mu_{12})(1 - \mu_{13}) - D_4(1 - \mu_{12}),$$

(28)

where step (a) results directly from (26). From this, the probability of $\mathcal{E}_{1a}$ occurring is, in the limit,

$$\lim_{n \to \infty} \Pr[\mathcal{E}_{1a}] = \lim_{n \to \infty} \Pr \left( \frac{D_3}{n} \sum_{\ell=1}^{D_3} \lambda_3[\ell] + \frac{D_4}{n} \sum_{\ell'=1}^{D_4} \lambda_4[\ell'] + D_5 \right) < m + \delta_n \right]$$

$$\leq \lim_{n \to \infty} \Pr \left( \frac{D_3}{n} \sum_{\ell=1}^{D_3} \lambda_3[\ell] + \frac{D_4}{n} \sum_{\ell'=1}^{D_4} \lambda_4[\ell'] + nr_{sym} - D_3(1 - \mu_{12})(1 - \mu_{13}) - D_4(1 - \mu_{12}) \right) < m + \delta_n \right]$$

$$= \lim_{n \to \infty} \Pr \left( \frac{D_3}{n} \sum_{\ell=1}^{D_3} \lambda_3[\ell] - E \left[ \frac{D_3}{n} \sum_{\ell=1}^{D_3} \lambda_3[\ell] \right] + \frac{D_4}{n} \sum_{\ell'=1}^{D_4} \lambda_4[\ell'] - E \left[ \frac{D_4}{n} \sum_{\ell'=1}^{D_4} \lambda_4[\ell'] \right] + nr_{sym} \right) < m + \delta_n \right]$$

$$= \lim_{n \to \infty} \Pr \left( \frac{D_3}{n} \sum_{\ell=1}^{D_3} \lambda_3[\ell] - E \left[ \frac{D_3}{n} \sum_{\ell=1}^{D_3} \lambda_3[\ell] \right] + \frac{D_4}{n} \sum_{\ell'=1}^{D_4} \lambda_4[\ell'] - E \left[ \frac{D_4}{n} \sum_{\ell'=1}^{D_4} \lambda_4[\ell'] \right] + nr_{sym} \right) < \delta_n - \delta_n \right]$$

$$(c) = 0,$$

where in (b) we applied the bound (28) while noting that if $a \geq b$ then $\Pr[a < c] \leq \Pr[b < c]$, and in (c) we invoked the law of large numbers while noting that $\delta_n - \delta_n$ is positive by construction.

Now consider the event $\mathcal{E}^{c}_{1a} \cap \mathcal{E}_{1b}$. This describes the case where User 1 receives enough (i.e., $m + \delta_n$) clean equations but the randomly generated matrix that maps $\vec{w}_1$ to clean equations has rank less than $m$. This type of error is well studied throughout the network coding literature. For instance, from expression (3) of [13], the probability of a $m \times m + \delta_n$ random binary matrix having rank less than $m$ can be bounded as

$$\Pr(\mathcal{E}^{c}_{1a} \cap \mathcal{E}_{1b}) \leq 2^{-\delta_n},$$
which implies, as desired,
\[
\lim_{n \to \infty} \Pr(E_{1a}^c \cup E_{1b}) = 0.
\]

We now revisit the analysis and note that \( E_{1a} \) may be thought of as the event where the actual amount of randomly available side information was “not enough” because it deviated significantly from the mean. On the other hand \( E_{1a} \cap E_{1b} \) describes the case where there was a sufficient amount of side information, but the randomly generated coding scheme failed to communicate the remaining desired message bits. For each type of error we applied a different analysis technique. To address the first, we applied a concentration inequality to show that the probability that the amount of randomly available side information deviates significantly from the mean vanishes as \( n \) grows large. To address the second, we applied existing analysis on the properties of randomly generated matrices to show that the probability of a rank-deficient encoding matrix vanishes as \( n \) grows large.

To prove that the probabilities of error events \( E_2 \) and \( E_3 \) also vanish as \( n \) grows large, we must systematically break down these error events into subevents of these two types. Since these two users apply the same decoding process, we now focus on User 2, and we identify such subevents.

Recall the User 2 first uses its side information and Phase 5 transmissions (i.e., RLCs with only \( \vec{w}_3 \) content) to decode \( \vec{w}_3 \). Let \( E_{2,1} \) be the event where User 2 fails to decode \( \vec{w}_1 \) which we further breakdown into the following subevents, \( E_{2,1a} \) and \( E_{2,1b} \):
\[
E_{2,1a}: \text{User 2 does not receive enough side information, i.e., } 1^T \vec{g}_{21} \leq \mu_{21} n_{\text{sym}} + \delta_n \text{, where } \delta_n \text{ grows with } n \text{ and } 0 < \delta_n < \delta_n.
\]
\[
E_{2,1b}: \text{The random matrix that describes the transformation of } \vec{w}_1 \text{ to Phase 5 RLCs is rank deficient.}
\]

One can verify using the same methods as in the analysis of \( E_1 \) that the probability of either \( E_{2,1a} \) or \( E_{2,1a} \cap E_{2,1b} \) occurring vanishes with large \( n \) as long as the rate \( r_{\text{sym}} \) satisfies (26).

Next, recall that after decoding \( \vec{w}_1 \), User 2 removes \( \vec{w}_1 \) content from Phases 1–4, and proceeds to decode both \( \vec{w}_2 \) and \( \vec{w}_3 \). Let \( E_{2,(2,3)} \) denote the event where User 2 fails to decode \( \{\vec{w}_2, \vec{w}_3\} \), and we now study specifically \( E_{2,1a} \cap E_{2,(2,3)} \). Notice that by construction, after the \( \vec{w}_1 \) content has been removed, User 2 will receive some uncoded bits of \( \vec{w}_2 \) from Phase 4. Furthermore, User 2 can also use its side information to clean the \( \vec{w}_3 \) component from some transmissions during Phase 3 to recover these bits from the \( \vec{w}_3 \). Noting this observation, we can now specify the final two error events for \( E_2 \) analysis:
\[
E_{2,1a} \cap E_{2,(2,3)}^a: \text{After decoding } \vec{w}_1 \text{ and removing its content from Phases 1–4, the number of bits about } \vec{w}_3 \text{ User 2 learns from side information and the number of clean uncoded bits about } \vec{w}_2 \text{ User 2 learns from Phases 3 and 4 is significantly less than the mean.}
\]
\[
E_{2,1a} \cap E_{2,(2,3)}^b: \text{After decoding } \vec{w}_1 \text{ and removing its content from Phases 1–4, the random matrix that describes the transformation of } \vec{w}_2 \text{ and } \vec{w}_3 \text{ to transmissions in Phases 1 and 2 is rank deficient.}
\]

Again, one can verify using the same methods as in the analyses of \( E_{1a} \) and \( E_{1b} \) that the probabilities of these events occurring vanish with large \( n \) as long as the rate \( r_{\text{sym}} \) satisfies (26). Using the analyses of these subevents, we have
\[
\lim_{n \to \infty} \Pr[E_2] = \lim_{n \to \infty} \Pr[E_{2,1}] + \Pr[E_{2,1} \cap E_{2,2,(2,3)}] = \lim_{n \to \infty} \Pr[E_{2,1a}] + \Pr[E_{2,1a} \cap E_{2,2,(2,3)}] + \Pr[E_{2,1a} \cap E_{2,2,(2,3)}, E_{2,2,(2,3)}^a] + \Pr[E_{2,1a} \cap E_{2,2,(2,3)}^a, E_{2,2,(2,3)}] = 0.
\]

Through similarly identifying subevents of \( E_3 \) we can also establish that \( \Pr[E_3] \to 0 \) as \( n \to \infty \). Therefore, the probability of decoding error at each user vanishes as \( n \) grows large.

VI.Numerical Results

In this section we perform numerical analysis of inner and outer bounds to illustrate 1) the gain in achievable rate of hybrid coding over conventional random coding, and 2) the gap between our derived inner and outer bounds.

To limit the scope of possible configurations (parameterized by \( \mu_{ij} \) terms), we focus on two symmetric scenarios for a representative set of parameters. In the first scenario, we consider side information that is “one-sided symmetric” (i.e., network parameters such that \( \mu_{ij} = \mu_{ik} \) for all \( i \neq j \neq k \)) while in the second, we consider side information that is “pairwise symmetric” (i.e., network parameters such that \( \mu_{ij} = \mu_{ji} \) for all \( i \neq j \)). For both scenarios, we will assume that, the size of side information at User 1 is the least and at User 3 is the most.

In Figures 6(a) and 6(b), we demonstrate the gap between our BIC inner and outer bounds while focusing on varying the amount of information at the user with the least side information. Figure 6(a) demonstrates the gap between inner and outer bounds on symmetric capacity for a one-sided symmetric BIC problem. In particular, we fix \( \mu_{21} = \mu_{23} = \frac{1}{2} \) and \( \mu_{13} = \mu_{31} = \frac{1}{2} \) and consider the impact of varying \( \mu_{12} = \mu_{13} = \alpha \) across the range from 0.5 to 1. Figure 6(b) demonstrates the gap between inner and outer bounds on symmetric capacity for a pairwise symmetric BIC problem. In particular, we fix \( \mu_{21} = \mu_{32} = \frac{1}{2} \) and \( \mu_{12} = \mu_{13} = \alpha \) across the range from 0.5 to 1.
In Figures 6(c) and 6(d) we demonstrate the gap between our BIC inner and outer bounds while focusing on varying the amount of information at the user with the most side information. Specifically, in Figure 6(c) we look at a one-sided symmetric scenario and fix $\mu_{12} = \mu_{13} = \frac{1}{2}$ and $\mu_{21} = \mu_{23} = \frac{1}{3}$, while varying $\mu_{31} = \mu_{32} = \frac{1}{3}$ across a range from 0 to $\frac{1}{2}$, while in Figure 6(d) we look at the pairwise symmetric scenario and fix $\mu_{12} = \mu_{21} = \frac{1}{3}$ and $\mu_{13} = \mu_{31} = \frac{1}{2}$, while varying $\mu_{23} = \mu_{32} = \frac{1}{3}$ across a range from 0 to $\frac{1}{2}$.

In the two BIC problems depicted in Figures 6(a) and 6(b) we point out that as the user with the least amount of side information loses even more side information (increasing $a$ or $b$), the rate achievable by conventional random codes decreases. On the other hand, in the BIC problems depicted in Figures 6(c) and 6(d) since amount of side information of the least knowledgeable user remains constant (i.e., $\mu_{12}$ and $\mu_{13}$ are fixed), the rate achieved by conventional random coding is constant across the range.

![Graphs showing inner and outer bounds](image)

**Fig. 6.** Inner and outer bounds on the symmetric capacity of example 3-user BIC problems: (a) One-sided side information symmetry, and (b) pairwise side information symmetry, while varying the least knowledgeable user’s side information under; and (c) one-sided side information symmetry, and (d) pairwise side information symmetry, while varying the most knowledgeable user’s side information.

With the figures, we highlight the following observations about our inner and outer bounds:

1) There exists a threshold for side information parameters where below this threshold, in the best hybrid coding strategy all three users decode all messages and thus the achieved rate is the same as conventional random codes. In particular, this is true for small $a$ and $b$ in Figures 6(a) and 6(b) and larger $c$ and $d$ in Figures 6(c) and 6(d), respectively. However, beyond this threshold (larger $a$ and $b$ and smaller $c$ and $d$), we observe a clear potential for increased rate from hybrid codes. It is worth noting that the regimes where hybrid codes offer a rate increase regimes are those further from the fully symmetric BIC (where all network parameters, $\mu_{ij}$, are the same), wherein the entire capacity region is achievable using conventional random coding (see Appendix D, Proposition 10).

2) Although there exists a gap between our inner and outer bounds, we highlight a specific case where our new hybrid coding scheme both provides strictly positive rate gain over conventional random coding and meets the new upper bound: in Figure 6(d) when $d = 0$. This scenario is related to the one considered in the motivating example of Section III in the sense that Users 2 and 3 know each other’s complete message as side information.
VII. BLIND INDEX CODING OVER WIRELESS CHANNELS

In this section, we generalize the BIC problem model further to consider the impact of uncertainty not only within the side information given to users, but also in the sender-to-user broadcast channel (recall that in the BIC problem this channel was error free). In particular, we emulate loss of packetized transmissions due fading in wireless channels using a binary fading model for the sender-to-user broadcast. Consequently, the problem considered here will be referred to as blind index coding over wireless channels (BICW).

As we will see, considering wireless transmissions adds new challenges to the problem, and surprisingly repetition of uncoded bits (within the hybrid coding framework) will become a powerful technique for increasing achievable rate. Unlike the BIC problem considered in the previous sections, even the two-user BICW problem is nontrivial. Hence, in this section we focus on a a two-user problem representative of general BICW problems. After formally defining the representative problem, we demonstrate numerically the resulting gain in achievable rate that our scheme provides over conventional methods.

A. Wireless Broadcast Channel Model

In the BICW scenario the channel output received by User $i$, $y^n_i$, is governed by a binary fading process. Specifically, let $\gamma_i$ be a binary vector with the same length as the channel input vector $\bar{x}$ and drawn i.i.d from a Bernoulli$(1-\epsilon_i)$ distribution. The channel output for User $i$ is given by the input-output relationship

$$y_i[n] = \gamma_i[n] x[n].$$

User $i$ knows $\gamma_i$, however the sender is only aware of parameters $\{\epsilon_i\}$, which govern the probabilistic behavior of the sender-to-user broadcast channel.

In this section, we assume the model depicted in Figure 9 containing only two users where $\epsilon_1 < \epsilon_2$, $\mu_{32} = 1$, and $\mu_{21} = \mu$ (i.e., User 1 has a better channel than User 2 but no side-information).

![Fig. 7. Two-user instance of the BICW problem.](image)

Remark 11. We assume that $\epsilon_1 < \epsilon_2$ and that side-information was only given to User 2 (i.e., $\mu_{21} = 1$) for ease of exposition. In all other two-receiver settings (i.e., arbitrary $\epsilon_1$ and $\epsilon_2$ and side-information at either receiver), either there is no index coding gain even if the server knows the side-information or the natural generalization of our proposed scheme recovers some index coding gain to outperform conventional approaches.

Our main result for this setting is as follows.

**Theorem 5.** For the two-user BICW problem defined above, the rate region $\mathcal{R}$ is achievable, where $\mathcal{R}$ is the set of all non-negative rate pairs $(r_1, r_2)$ satisfying

$$r_1 + r_2 \leq 1 - \epsilon_1, \quad (30)$$

$$\omega_1(L) r_1 + \omega_2(L) r_2 \leq 1 - \epsilon_2, \quad L = 1, \ldots, L_{\text{max}}$$

where

$$\omega_1(L) = \frac{1 - \epsilon_2}{1 - \epsilon_1} \epsilon_1 + \mu (1 - \epsilon_2^2) \omega_2(L) + L (1 - \epsilon_2) (1 - \omega_2(L)), \quad (32)$$

$$\omega_2(L) = \min \left\{ \frac{1 - \epsilon_i^2}{1 - \mu \epsilon_i^2}, 1 \right\}, \quad (33)$$

$$L_{\text{max}} \triangleq 1 + \left\lfloor \frac{\log(\mu)}{\log(\epsilon_1/\epsilon_2)} \right\rfloor. \quad (34)$$

Remark 12. Notice that as $\epsilon_2 \to 0$ (and by the assumption $\epsilon_2 > \epsilon_1$, as $\epsilon_1 \to 0$), the BICW problem reverts to a BIC problem. Moreover as $\epsilon_2 \to 0$, $\omega_1(L) \to \mu$ and $\omega_2(L) \to 1$, resulting in the achievable region of rate pairs satisfying:

$$r_1 + r_2 \leq 1, \quad \mu r_1 + r_2 \leq 1,$$
which is equivalent (given assumptions on $\mu_{12}$ and $\mu_{21}$) to the two-user BIC capacity region (formally stated in Appendix D Proposition 5).

B. Proof of Theorem 5

This section is organized as follows. We first define the hybrid coding scheme by specifying a class of generator matrices which map length-$m$ message vectors to length-$n$ codewords, and which are parametrized by three quantities: $\rho$, $L$, and $\alpha$. For each $n$, the transmitter maps two messages, $\vec{w}_1$ and $\vec{w}_2$ to codewords using corresponding generator matrices (with different parameters), and XORs the two codewords to produce the channel input vector. We then specify the method of decoding and establish the achievable rate region for our coding scheme when fixing the generator matrix parameters for all $n$. By doing so, we show that for any $(r_1, r_2) \in \mathcal{R}$ (as defined in Theorem 5) there exists a choice of parameters such that $(r_1, r_2)$ is achievable, thus proving Theorem 5.

1) Encoding: Our hybrid coding scheme encodes $\vec{w}_1$ and $\vec{w}_2$ separately and linearly, before combining the resulting codewords through bit-wise XOR. The codeword for each message is constructed in a manner similar to the component of BIC hybrid codes from the previous section specific to a single message component: uncoded repetitions of message bits are supplemented by a random linear combinations. The specific mapping from message to codeword is formalized in the following definition, parametrized for a given $n$ by three quantities $\rho$, $L$ and $\alpha$:

Definition 2 (Repetition plus Random Parity (RRP) Matrix). An $n \times m$ RRP matrix with parameters $\rho \in [0, 1]$, $L \in \mathbb{N}$, and $\alpha \in [0, 1]$ is a binary matrix, $\mathbf{U}$, with the form:

$$
\mathbf{U} = \begin{bmatrix} \mathbf{B}^T & \mathbf{A}_1^T & \ldots & \mathbf{A}_{L+1}^T & \mathbf{0}^T \end{bmatrix}^T,
$$

where

$$
\mathbf{A}_\ell = \begin{cases} 
\mathbf{I}_m & \text{if } \ell \leq L \\
\mathbf{0}_{m \times m} & \text{else}
\end{cases}
$$

and $\mathbf{B}$ is a $\rho m \times m$ matrix with entries drawn i.i.d. from Bernoulli $\left(\frac{1}{2}\right)$. For feasibility, we require that $\alpha m$ is an integer, and

$$(L + \alpha \frac{m}{n}) + \rho \leq 1. \quad (36)$$

Remark 13. Simply stated, an RRP matrix maps a length-$m$ message vector to a length-$n$ codeword by repeating each uncoded message bit either $L$ or $L + 1$ times. The parameter $\alpha$ specifies the fraction of bits repeated $L + 1$ times, while $\rho$ specifies the proportion of length-$n$ codeword reserved for random linear coded parity. Inequality (36) ensures that $\mathbf{U}$ is an $n \times m$ matrix.

It is worth noting that in the hybrid encoding scheme described for three-user (non-wireless) BIC, the mapping of message $\vec{w}_1$, $\vec{w}_2$, and $\vec{w}_3$ to sequences before XOR (i.e., the individually colored bars in Figure 5) could be interpreted as RRP matrices. For $\vec{w}_1$, we chose $L = \alpha = 0$ and for messages $\vec{w}_2$ and $\vec{w}_3$ we chose $L = 1$ and $\alpha = 0$. The use of RRP matrices with $L > 1$ and $\alpha > 0$ (i.e., the repetition of uncoded message bits) is the key innovation to hybrid coding that enables higher rate in the wireless setting.

Using the defined RRP matrices, we now describe the encoding scheme that maps messages $\vec{w}_1$ and $\vec{w}_2$ to a length-$n$ channel input vector. Let $n$, $m_{1}^{(n)}$, and $m_{2}^{(n)}$ be given. For each $n$, let $\mathbf{U}_1$ be a $n \times m_{1}^{(n)}$ RRP matrix with parameters $\left(\rho_{1}, L_{1}, \alpha_{1}\right)$ and $\mathbf{U}_2$ be a $n \times m_{2}^{(n)}$ RRP matrix with parameters $\left(\rho_{2}, L_{2}, \alpha_{2}\right)$. The channel input vector $\vec{x}^n$ is given by (assuming modulo-2 addition):

$$
\vec{x}^n = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \vec{w}_1 \\ \vec{w}_2 \end{bmatrix} = \mathbf{U}_1 \vec{w}_1 + \mathbf{U}_2 \vec{w}_2.
$$

Figure 8 depicts an example hybrid encoding with repetitions for the two-user BICW setting. In this particular example, $L = 2$ and $\alpha = 0.5$. 

![Fig. 8. An example hybrid coding scheme for the two-user BICW setting, where $(\rho_1, L_1, \alpha_1) = (0.25, 2, 0.5)$, and $(\rho_2, L_2, \alpha_2) = (1, 0, 0)$. Outlined boxes represent uncoded bits, shaded boxes represent RLCs of a single message.](image-url)
2) Decoding: We now specify the decoding strategy and then characterize the achievable rates for our scheme with fixed parameters $\rho_i$, $L_i$, and $\alpha_i$, $i = 1, 2$. In what follows, we choose $(\rho_2, L_2, \alpha_2) = (1, 0, 0)$ (i.e., User 2’s generator matrix, $U_2$, is a random matrix). Choosing parameters $(\rho_1, L_1, \alpha_1)$ is more nuanced and will be addressed within the analysis. For brevity, we will not explicitly analyze the error rates of our scheme for given $n$, but instead provide a sketch of the achievability proof using existing results for random linear codes over point-to-point erasure channels.

In our decoding strategy, User 1 first decodes $\vec{w}_2$ and peels its interfering contribution from its received signal, and then decodes its desired message, $\vec{w}_1$. Receiver 2 only decodes $\vec{w}_2$. We first describe decoding $\vec{w}_2$ at each receiver.

Recall that the channel input at any time, $t$, is given by $x[t] = U_i(t,:)[\vec{w}_1 + U_2(t,:)\vec{w}_2]$, where $U_i(t,:) = \{t\}$ is the $t$-th row of generator matrix $U_i$. The decoding strategy for $\vec{w}_2$ used by both receivers is based on the following observation. If $t$ and $t' \neq t$ both correspond to a repetition of the same message bit from $\vec{w}_1$, then the module-2 sum of these yields $x[t] + x[t'] = (U_2(t,:) + U_2(t',:))\vec{w}_2$, which is a random linear combination of only $\vec{w}_2$ bits (since $\rho_2 = 1$). By this method we “clean” equations of $\vec{w}_1$. User 2 has the additional option of using its side-information to clean equations, which has the same essence.

The cleaned random linear equations are used by each receiver in conjunction with those that by construction were only functions of $\vec{w}_2$ (i.e., for those $t$ where $U_1(t,:) = 0$) to decode $\vec{w}_2$. After decoding $\vec{w}_2$, User 1 removes the contribution of $\vec{w}_2$ from its received signal before decoding $\vec{w}_1$. If any of these decodings fail, then an error occurs. We now claim that the decoding scheme yields the following achievable rates, proven in Appendix E.

**Lemma 6.** Consider the two-receiver BWIC problem defined by network parameters $\epsilon_1$, $\epsilon_2$, and $\mu$, and let $\rho_1 \in [0, 1]$, $L_1 \in \mathbb{N}$, and $\alpha_1 \in [0, 1)$ be fixed. A rate pair $(r_1, r_2)$ is achievable if it satisfies,

\begin{align}
r_1 &\leq \frac{1 - \rho_1}{L_1 + \alpha_1}, \quad (37) \\
(1 - \epsilon_1) r_1 &\leq \frac{1 - \epsilon_1}{\epsilon_1} (1 - \epsilon_1) r_1 \leq 1 - L r_1, \\
\frac{[1 - \epsilon_1 L_1 + \alpha_1 (\epsilon_1 L_1 - \epsilon_1 L_1 + 1)]}{1 - \epsilon_1} r_1 + r_2 &\leq (1 - \epsilon_1)(1 - \rho_1), \quad (39) \\
\mu (1 - \epsilon_2 L_2 + \alpha_1 (\epsilon_2 L_2 - \epsilon_2 L_1 + 1)) r_1 + r_2 &\leq (1 - \epsilon_2)(1 - \rho_1). \quad (40)
\end{align}

From Lemma 6, it is clear that by considering the union or achievable rate pairs over all $(\rho_1, L_1, \alpha_1)$ we arrive at the rate region achievable by our schemes. Specifically, let $\mathcal{R}(\rho_1, L_1, \alpha_1)$ for $\rho_1 \in [0, 1]$, $L_1 \in \mathbb{N}$, and $\alpha_1 \in [0, 1)$ be defined as the set of all pairs $(r_1, r_2)$ satisfying (37)-(40), and we define a rate region:

$$
\mathcal{R} \triangleq \bigcup_{\rho_1, L_1, \alpha_1} \mathcal{R}(\rho_1, L_1, \alpha_1).
$$

To complete the proof of Theorem 5, we now demonstrate that the region $\mathcal{R}$ (as defined in Theorem 5) is contained within $\mathcal{R}$ (given in (41)), and thus is achievable. To do so, we need only show that for every rate pair $(r_1, r_2) \in \mathcal{R}$, there exists parameters $(\rho_1, L_1, \alpha_1)$ such that (37)-(40) are satisfied. We therefore fix $r_1$ to any value in the interval $[0, 1 - \epsilon_1]$, and choose parameters $\rho_1^*, L_1^*$, and $\alpha_1^*$ as

\begin{align}
L_1^* &= \text{maximize } \min \{L, L_{\text{max}} \} \\
\text{subject to } \quad L \in \mathbb{N} \\
\frac{\epsilon_1}{1 - \epsilon_1} r_1 &\leq 1 - L r_1 \\
\alpha_1^* &= \begin{cases} 
0 & \text{if } L_1^* = L_{\text{max}} \\
1 - r_1 \left( \frac{\epsilon_1 L_1^* + L_1^*}{1 - \epsilon_1} \right) & \text{if } L_1^* < L_{\text{max}} 
\end{cases} \\
\rho_1^* &= \frac{\epsilon_1 L_1^* - \alpha_1^*(\epsilon_1 L_1^* - \epsilon_1 L_1^* + 1)}{1 - \epsilon_1} r_1,
\end{align}

where $L_{\text{max}}$ as defined in (34). Notice that given $r_1$, we first determine the appropriate $L_1^*$, then $\alpha_1^*$, then finally $\rho_1^*$ and that both (37) and (38) are satisfied by the chosen parameters.

Substituting these into (39) and (40), we see that $r_2$ is achievable if it satisfies both of the following inequalities:

\begin{align}
r_2 &\leq (1 - \epsilon_1) - \rho_1^*(1 - \epsilon_1) - r_1 \left[ 1 - \epsilon_1 L_1^* + \alpha_1^*(\epsilon_1 L_1^* - \epsilon_1 L_1^* + 1) \right] \\
&= 1 - \epsilon_1 - r_1, \\
r_2 &\leq (1 - \epsilon_2) - \rho_1^*(1 - \epsilon_2) - r_1 \mu \left[ 1 - \epsilon_2 L_1^* + \alpha_1^*(\epsilon_2 L_1^* - \epsilon_2 L_1^* + 1) \right] \\
&= (1 - \epsilon_2) \left[ 1 - r_1 \left( \frac{\epsilon_1 L_1^*}{1 - \epsilon_1} + \mu \frac{1 - \epsilon_1}{1 - \epsilon_2} - \frac{\alpha_1^*(\epsilon_1 - \mu \epsilon_2)}{1 - \epsilon_1} \right) \right].
\end{align}
We now point out that (45) is equivalent to (30) and (46) is equivalent to (31) evaluated at
where in (a) we compared the evaluated expression with
\( \omega \)
Denote the subvector of \( \vec{w} \) respectively:
and the right hand side of inequality represents the tightest version of (31) for fixed \( r \)
C. Numerical Results

For blind index coding over wireless channels, we recall that the key difference was the usefulness of repeating uncoded bits within the hybrid coding scheme. Therefore, we now provide numerical results for three BICW scenarios, characterized by \( \epsilon_1, \epsilon_2, \) and \( \mu \). In each, we plot \( \mathcal{R} \) and highlight regimes (along x-axes) wherein the number of repetitions used in our scheme increases. For each scenario, we point out the gain in \( r_2 \) offered by repetition-based hybrid codes over conventional schemes, and for further comparison we also depict rate regions (along y-axes) wherein the number of repetitions used in our scheme is indeed achievable, thus completing the proof of Theorem 5.

Proposition 7. For the two-user BICW problem setting considered in Theorem 5 an achievable rate pair \( (r_1, r_2) \) must satisfy

\[
\max \left\{ r_1 + r_2, \mu r_1 + \frac{1 - \epsilon_1}{1 - \epsilon_2} r_2 \right\} \leq 1 - \epsilon_1.
\]

Proof: The bound may be separated into two outer bounds that correspond to the first and second terms within the \( \max \), respectively:

- \( r_1 + r_2 \leq 1 - \epsilon_1 \)
- \( \mu \frac{1 - \epsilon_2}{1 - \epsilon_1} r_1 + r_2 \leq 1 - \epsilon_2 \)

Denote the subvector of \( \vec{w}_1 \) given as side information as \( \vec{w}_1^+ \) and the complementary subvector as \( \vec{w}_1^- \). We prove the first bound by applying Fano’s inequality at each user to observe:

\[
nr_1 \leq I \left( \bar{y}_1, \gamma_1; \vec{w}_1 \right) + o(n)
= I \left( \bar{y}_1, \gamma_1; \vec{w}_1 \right) + o(n)
= H \left( \bar{y}_1 | \gamma_1 \right) - H \left( \bar{y}_1 | \gamma_1, \vec{w}_1 \right) + o(n)
\leq H \left( \bar{y}_1 | \gamma_1 \right) - H \left( \bar{y}_2 | \gamma_2, \vec{w}_1 \right) + o(n)
\leq n(1 - \epsilon_1) - H \left( \bar{y}_2 | \gamma_2, \vec{w}_1^+ \right) + o(n),
\]

\[
nr_2 \leq I \left( \bar{y}_2, \gamma_2, \vec{w}_2 \right) + o(n)
= H \left( \bar{y}_2 | \gamma_2, \vec{w}_2 \right) - H \left( \bar{y}_2 | \gamma_2, \vec{w}_2, \vec{w}_1 \right) + o(n)
\leq H \left( \bar{y}_2 | \gamma_2, \vec{w}_2 \right) + o(n)
= H \left( \bar{y}_2 | \gamma_2, \vec{w}_1^+ \right) + o(n),
\]

where in step (a) we observed that because the sender does not know the fading channel state of the sender-to-user channel. We complete the proof of the first bound by combining (49) and (50) and normalizing by \( n \) as \( n \) grows large.

To prove the second bound, we consider a genie which provides the sender of knowledge regarding which bits of \( \vec{w}_1 \) are given as side information to User 2. We again applying Fano’s inequality at each user, but in a different way, to observe

\[
nr_1 \leq I \left( \bar{y}_1, \gamma_1; \vec{w}_1^- \right) + o(n)
\leq I \left( \bar{y}_1, \gamma_1, \vec{w}_1^+, \vec{w}_2; \vec{w}_1^- \right) + o(n)
\leq H \left( \bar{y}_1 | \gamma_1, \vec{w}_1^+, \vec{w}_2 \right) + o(n),
\]

\[
nr_2 \leq I \left( \bar{y}_2, \gamma_2, \vec{w}_2 \right) + o(n)
= H \left( \bar{y}_2 | \gamma_2, \vec{w}_2 \right) - H \left( \bar{y}_2 | \gamma_2, \vec{w}_2, \vec{w}_1 \right) + o(n)
\leq n(1 - \epsilon_2) - H \left( \bar{y}_2 | \gamma_2, \vec{w}_1^+, \vec{w}_2 \right) + o(n),
\]
where in step (b) we applied Lemma 1 of [16] which when applied to our problem states that (because the sender does not know the binary fading channel states \( \{ \gamma_i \} \)),

\[
H ( \hat{y}_1 | \gamma_1^+, \hat{w}_1^+, \hat{w}_2 ) + o(n),
\]

To complete the proof of the second outer bound, we scale \( \{ \hat{y}_i \} \) by \( \frac{1-\epsilon_2}{1-\epsilon_1} \) and combine with (52). ■

In Figure 9(a), \( \epsilon_1 = \frac{1}{2} \), \( \epsilon_2 = \frac{3}{4} \), \( \mu = \frac{1}{4} \) notice that when \( r_1 \) is near the point-to-point capacity of 0.5, hybrid coding recovers all of the available index coding gain. This is because when \( r_1 \) is near 0.5, the primary challenge is not blindly exploiting side-information, but rather accounting for interference incurred at User 1. For this set of network parameters, we point out that for any fixed value of \( r_1 \), hybrid coding offers at least 62% of the available index coding gain.

In Figure 9(b), \( \epsilon_1 = \frac{1}{2} \), \( \epsilon_2 = \frac{9}{10} \), \( \mu = \frac{1}{10} \) we consider a BICW setting where side-information is plentiful (User 2 knows 90% of \( \hat{w}_1 \)). In this case, \( L_{max} = 4 \) and the piece-wise linear boundary of the hybrid coding achievable rate region has more linear segments, with segments corresponding to the number of repetitions used. For this setting and for any fixed \( r_1 \), HRC always achieves at least 68% of the available index coding gain.

Finally, in Figure 9(c), \( \epsilon_1 = \frac{1}{2} \), \( \epsilon_2 = \frac{3}{4} \), \( \mu = \frac{9}{10} \) we consider a BICW setting with very little side-information (User 2 knows 10% of \( \hat{w}_1 \)). In this case, \( L_{max} = 1 \) and from the figure, it is apparent that although any index coding gain is modest, it is still strictly positive for all \( r_1 \in \{ 0, 1-\epsilon_1 \} \).

From the scenarios depicted in Figure 9, we make the following unifying conclusions:

1) Regardless of the network parameters (\( \epsilon_1, \epsilon_2, \) and \( \mu \)) hybrid coding always increases the achievable rate region.

2) If we consider a fixed \( r_1 \), the number of repetitions used in the hybrid encoding scheme increases when User 2 has a weaker channel and more side information (i.e., \( \epsilon_2 \) grows larger and \( \mu \) grows smaller).

3) Hybrid coding can be capacity achieving, as seen on the boundary of the rate regions in all three figures when \( r_1 \) is close to it’s maximum.

VIII. CONCLUDING REMARKS

In this paper, we introduced a generalization of index coding called blind index coding, which captures key issues in distributed and wireless settings. We demonstrated that the BIC problem introduces new interesting challenges that require new analytical tools through three main contributions: 1) we presented new outer bounds that leveraged a lemma based on a notion of strong data processing to capture the lack of knowledge at the sender, 2) we proposed a class of hybrid coding schemes which mix uncoded bits of a subset of messages with randomly linear combinations of other messages, and 3) we demonstrated that in scenarios where the sender-to-user channel is not error-free (specifically, a wireless binary fading channel) that repetition of uncoded bits within hybrid codes can further increase the achievable rate.

The blind index coding problem in the general setting remains an open problem. Therefore to conclude the paper we revisit one class of interesting symmetric side information BIC problems (from Section VII) that remains unsolved and yet offers a simple and concrete enough case for progress to be made, potentially revealing new insights. Consider the following 3-user BIC scenario when side information parameters are pairwise symmetric: \( \mu_{12} = \mu_{21} = a \), \( \mu_{13} = \mu_{31} = b \), \( \mu_{23} = \mu_{32} = c \) with \( a \geq b \geq c \) (for a concrete example we refer the reader to Figure 6(d)). From Theorem 4, we find the symmetric achievable rate:

\[
R_{sym} = \max \left\{ \frac{1}{1 + a + b + c - ab}, \frac{1}{1 + a + b} \right\},
\]

and from Theorem 3 we have the capacity bound:

\[
R_{sym} \leq \frac{1}{1 + a + b - \frac{(a-c)(b-c)}{1-c}}.
\]

Notice first that, as in the numerical example of Figure 6(d) if \( c = 0 \) or \( c = b \) the upper bound is tight and capacity is achieved. However, within the interval \( c \in (0, b) \) there exists a gap between achievability and converse.

Additionally, recall that the first quantity in the max of (53) is the rate achieved by hybrid coding and the second is by conventional random coding. Clearly, hybrid coding provides a rate gain when \( c < ab \). This regime is one where the side information Users 2 and 3 have about each others’ messages is large and thus Phases 1 and 2 in Figure 3 are small. Our hybrid coding assumes that User 1 ignores these phases, but when they are larger (i.e., as \( c \) grows) these transmissions may be used by User 1 to decode messages \( \hat{w}_2 \) and \( \hat{w}_3 \). In particular, the case where \( c = ab \) (a point notably within the interval \( (0, b) \)) represents a threshold where the structure of our hybrid code can no longer expect to hide linear subspaces of \( \hat{w}_2 \) and \( \hat{w}_3 \) from User 1.

We conjecture that at this threshold, any method of encoding \( \hat{w}_2 \) and \( \hat{w}_3 \) that satisfies the decodability condition at Users 2 and 3 also allows User 1 to decode \( \hat{w}_2 \) and \( \hat{w}_3 \) (i.e., at this threshold it is the converse and not achievable scheme that may be tightened).
Fig. 9. Rate regions achieved by different schemes — Conventional random codes (blue), time-division between separate random codes (green), hybrid coding (red), and genie-aided (non-blind) index coding (white) — for three different 2-user BICW problems. The number of repetitions used in the hybrid coding scheme is stated along the x-axis. (a) For this setting, $L_{\max} = 2$; (b) For this setting, $L_{\max} = 4$ and we have emphasized using dashed lines bounds (30) and (31) for all $L$ that comprise the boundary of $R$; (c) For this setting, $L_{\max} = 1$ and notice even with very little side information, our hybrid coding scheme strictly outperforms conventional schemes.

REFERENCES

[1] D. T. H. Kao, M. A. Maddah-Ali, and A. S. Avestimehr, “Blind index coding,” to appear in Information Theory (ISIT), 2015 IEEE International Symposium on.

[2] ———, “Blind index coding over wireless channels: The value of repetition coding,” to appear in Communications (ISIT), 2015 IEEE International Conference on.

[3] ———, “Align-and-forward relaying for two-hop erasure broadcast channels,” in Information Theory (ISIT), 2014 IEEE International Symposium on, June 2014.

[4] Y. Birk and T. Kol, “Coding on demand by an informed source (iscod) for efficient broadcast of different supplemental data to caching clients,” Information Theory, IEEE Transactions on, vol. 52, no. 6, pp. 2825–2830, June 2006.
We now prove that $s^\ast((\phi', g')\{\phi_{ij}[1], g_{ij}[1]\}) = s^\ast((\phi', g')\{\phi_{ij}[1], g_{ij}[1]\}) = 1 - \mu^\ast$ by showing that it may be bounded both from above and below by the same value. We first address the upper bound:

$$s^\ast((\phi', g')\{\phi_{ij}[1], g_{ij}[1]\}) = \max_{\mathcal{I} \subseteq \{1, \ldots, m\}} s^\ast((\phi_{ij}[\mathcal{I}], g_{ij}[\mathcal{I}])))$$

$$\leq s^\ast((\phi'[1], g'[1])\{\phi_{ij}[1], g_{ij}[1]\}).$$

(55)

In step (a) we apply the tensorization property of $s^\ast(\cdot)$ [14], and in (b) we observed that all variables are i.i.d. across $\ell$. To simplify exposition, we now use the following notation: Let $P_\alpha(\cdot)$ and $P_\beta(\cdot)$ denote probability mass functions for $(\phi'[1], g'[1])$ and $(\phi_{ij}[1], g_{ij}[1])$ respectively, and let $Q_\alpha(\cdot)$ and $Q_\beta(\cdot)$ be arbitrary probability mass functions for $(\phi'[1], g'[1])$ and $(\phi_{ij}[1], g_{ij}[1])$ respectively. Note that the support of both $(\phi'[1], g'[1])$ and $(\phi_{ij}[1], g_{ij}[1])$ is $\{(0,0), (0,1), (1,1)\}$. Using this notation, we now observe

$$s^\ast((\phi'[\mathcal{I}], g'[\mathcal{I}])), (\phi_{ij}[\mathcal{I}], g_{ij}[\mathcal{I}])) = \sup_{Q_\alpha \neq P_\alpha} D(Q_\beta||P_\beta)$$

$$= \sup_{Q_\alpha \neq P_\alpha} \left[ P_\alpha(0,0) \log \left( \frac{P_\beta(0,0)}{Q_\beta(0,0)} \right) + P_\beta(0,1) \log \left( \frac{P_\beta(0,1)}{Q_\beta(0,1)} \right) \right]$$

$$+ P_\beta(1,1) \log \left( \frac{P_\beta(1,1)}{Q_\beta(1,1)} \right) / D(Q_\alpha||P_\alpha)$$

$$= \sup_{Q_\alpha \neq P_\alpha} \left[ (\delta + (1-\delta)P_\alpha(0,0)) \log \left( \frac{\delta + (1-\delta)P_\alpha(0,0)}{\delta + (1-\delta)Q_\alpha(0,0)} \right) + (1-\delta)P_\alpha(0,1) \log \left( \frac{(1-\delta)P_\alpha(0,1)}{(1-\delta)Q_\alpha(0,1)} \right) \right]$$

$$+ (1-\delta)P_\alpha(1,1) \log \left( \frac{(1-\delta)P_\alpha(1,1)}{(1-\delta)Q_\alpha(1,1)} \right) / D(Q_\alpha||P_\alpha)$$

(56)
Remark 14. Note that the validity of Lemma 2 only requires the upper bound (57). However, by evaluating the lower bound (59) as well, we may confirm the exact value of \( s^*(⟨(\phi', \vec{g}); (\phi_{ij}, \vec{g}_{ij})⟩) \). This value has an intuitive interpretation as the success probability of the channel that takes each bit of the virtual signal \( \phi' \) as input and gives \( \vec{\phi}_{ij} \) as output.

B. Proof of (9)

If \( \mu_{kj} \geq \mu_{ij} \), we observe that \( \vec{w}_i \) and \( (\vec{\phi}_{ij}, \vec{g}_{ij}) \) are statistically enhanced versions of \( (\vec{\phi}_{ki}, \vec{g}_{ki}) \) and \( (\vec{\phi}_{kj}, \vec{g}_{kj}) \) respectively. We may further enhance the side information of User \( k \) by also providing \( (\vec{\phi}_{ik}, \vec{g}_{ik}) \). Applying Fano’s inequality at User \( k \) with side information enhancement, we find

\[
\begin{align*}
\text{nr}_k & \leq \min I(\vec{x}, \vec{\phi}_{ki}, \vec{g}_{ki}; \vec{\phi}_{kj}, \vec{g}_{kj}; \vec{w}_k) + o(n) \\
& \leq I(\vec{x}, \vec{w}_i, \vec{\phi}_{ij}, \vec{g}_{ij}; \vec{\phi}_{ik}, \vec{g}_{ik}; \vec{w}_k) + o(n) \\
& = I(\vec{\phi}_{ik}, \vec{g}_{ik}; \vec{w}_k) + H(\vec{x}|\vec{w}_i, \vec{\phi}_{ij}, \vec{g}_{ij}; \vec{\phi}_{ik}, \vec{g}_{ik}) - H(\vec{x}|\vec{w}_i, \vec{\phi}_{ij}, \vec{g}_{ij}, \vec{w}_k) + o(n) \\
& = n(1 - \mu_{ik})r_k + H(\vec{x}|\vec{w}_i, \vec{\phi}_{ij}, \vec{g}_{ij}, \vec{\phi}_{ik}, \vec{g}_{ik}) - H(\vec{x}|\vec{w}_i, \vec{\phi}_{ij}, \vec{g}_{ij}, \vec{w}_k) + o(n) \\
& \overset{(a)}{=} n(1 - \mu_{ik})r_k + H(\vec{x}|\vec{w}_i, \vec{\phi}_{ij}, \vec{g}_{ij}, \vec{\phi}_{ik}, \vec{g}_{ik}) - \mu_{ij}H(\vec{x}|\vec{w}_i, \vec{w}_k) + o(n),
\end{align*}
\]

where in step (a) we applied (4) from Lemma 2 by letting \( V = (\vec{w}_i, \vec{w}_k) \).

By combining (10) and scaled versions of (60) and (15), and taking the limit as \( n \) grows large, we arrive at (9). Similarly, if \( \mu_{jk} \geq \mu_{ik} \) we may switch the roles of Users \( j \) and \( k \) in the above analyses to arrive at a similar conclusion.
C. Proof of Theorem 3

We now formally prove Theorem 3. The following notation and claim will simplify exposition of the proof. Let $\bar{g}_j^{(n)}$ be a length-$m_j$ vector of i.i.d. Bernoulli random variables that take a value of 0 with probability $\eta$, and let

$$\phi_j^{(n)}[\ell] = g_j^{(n)}[\ell] w_j[\ell].$$

Notice, for instance, that $\phi_j^{(n)}[\ell]$ is statistically equivalent to $(\phi_{i,j}, g_{i,j})$, and that $\psi_j^{(n)}$ and $\bar{\psi}_j^{(n)}$ are equal to $(\bar{w}_j, \bar{1})$ and $(\bar{0}, \bar{0})$ respectively. Additionally, we define $\bar{\psi}_j^{(n)} \triangleq (\bar{\phi}_j^{(n)}, \bar{g}_j^{(n)})$. We now formalize the notion of statistically enhanced side information, with the following claim, which is consequence of the sender being blind to the precise side-information:

**Claim 8.** Let $\eta_1 > \eta_2$ be given. For any $k \in \{1, \ldots, K\}$ and $V$ independent of $\bar{w}_j, g_j^{(n)}$, and $\bar{g}_j^{(n)}$, we have

$$I(\bar{x}, \psi_j^{(n)}; V, \bar{w}_k) \leq I(\bar{x}, \bar{\psi}_j^{(n)}; V, \bar{w}_k),$$

and

$$H(\bar{x}|\psi_j^{(n)}; V) \geq H(\bar{x}|\bar{\psi}_j^{(n)}; V).$$

**Proof:** The proof is an immediate consequence of the sender being blind to the side-information channels: since each quantity is a function only of the marginal distribution of $\bar{g}_j^{(n)}$ we may in fact define $\bar{\psi}_j^{(n)}$ as a physically degraded version of $\bar{\psi}_j^{(n)}$. Hence, the right side of (61) can be seen as the mutual information between a message and an enhanced channel output, and the right side of (62) can be seen as an enhance signal adding conditioning.

The proof now proceeds as follows. At each node $(\ell, i)$ in the OBT with $\ell < K$, we will apply Fano’s inequality to a virtual user that desires to receive message $\bar{w}_i^{(\ell)}$. This virtual user is given side information signals that are statistically enhanced versions of $\{\phi_{v[\ell,i],j}\}$ (i.e. the actual user in the BIC problem). The statistical properties of the side information channels governed by $\{\eta_{v[\ell,i],j}\}$ as defined by the OBT structure and (20). We will see that by applying Lemma 2 to the expansion of Fano’s inequality at each node, and scaling the resulting scaling expressions according, that terms on the right hand side of each expression will cancel and we will arrive at the stated bound.

We denote the complete collection of side information and channel state information given to a virtual user represented by the $i$-th node in level $\ell$ of the OBT as $\bar{\Psi}[\ell, i] = (\bar{\psi}_j^{(\ell)[\ell, i]}, \ldots, \bar{\psi}_j^{(K)[\ell, i]}).$ Similarly, we denote the collection of side information/channel state given to the actual User $v[\ell, i]$ as $\bar{\Psi}[v[\ell, i]].$

If $i$ is odd, recall from (20) that $\eta_{v[\ell,i],j} = 1$ which implies that none of the virtual user’s desired message is provided as (enhanced) side information. Thus, for off $i$ and $\ell < K - 1$ we have

$$nr_{v[\ell, i]} \leq I(\bar{x}, \bar{\Psi}_v[\ell, i]; \bar{w}_{v[\ell, i]}) + o(n)$$

$$\leq I(\bar{x}, \bar{\Psi}_v[\ell, i]; \bar{w}_{v[\ell, i]}),$$

$$= H(\bar{x}|\bar{\Psi}_v[\ell, i]) - H(\bar{x}|\bar{\Psi}_v[\ell, i], \bar{w}_{v[\ell, i]}) + o(n)$$

$$\leq (a) H(\bar{x}|\bar{\Psi}_v[\ell, i]) - \zeta(\ell, i) H(\bar{x}|\bar{\Psi}_v[\ell, i], 2i - 1)) - (1 - \zeta(\ell, i)) H(\bar{x}|\bar{\Psi}_v[\ell, i] + 2i)) + o(n),$$

where in step (a) we applied a combination of Lemma 2 and observing from (20) that $\bar{\Psi}[\ell + 1, 2i - 1]$ or $\bar{\Psi}[\ell + 1, 2i]$ can only increase conditioning relative to $\bar{\Psi}[\ell, i], \bar{w}_{v[\ell, i]}.$

If $i$ is even, some of the virtual user’s desired message may have been provided as side information. Thus, for even $i$ and $\ell < K - 1$ we have

$$nr_{v[\ell, i]} \leq I(\bar{x}, \bar{\Psi}_v[\ell, i]; \bar{w}_{v[\ell, i]}) + o(n)$$

$$\leq I(\bar{x}, \bar{\Psi}_v[\ell, i]; \bar{w}_{v[\ell, i]}),$$

$$= I(\bar{\psi}_j^{(\ell)[\ell, i]}; \bar{w}_{v[\ell, i]}),$$

$$= H(\bar{x}|\bar{\Psi}_v[\ell, i]) - H(\bar{x}|\bar{\Psi}_v[\ell, i], \bar{w}_{v[\ell, i]}) + o(n)$$

$$= n(1 - \eta_{v[\ell, i],j})[\ell, i]) r_{v[\ell, i]} + H(\bar{x}|\bar{\Psi}_v[\ell, i], \bar{w}_{v[\ell, i]})) + o(n)$$

$$\leq (b) n \left[1 - \eta_{v[\ell, i],j} \left(\ell - 1, \left(\frac{i}{2}\right)\right) r_{v[\ell, i]} + H(\bar{x}|\bar{\Psi}_v[\ell, i], \bar{w}_{v[\ell, i]})) + o(n)$$

$$\leq (c) n \left[1 - \eta_{v[\ell, i],j} \left(\ell - 1, \left(\frac{i}{2}\right)\right) r_{v[\ell, i]} + H(\bar{x}|\bar{\Psi}_v[\ell, i], \bar{w}_{v[\ell, i]})) + o(n)$$

$$- \zeta(\ell, i) H(\bar{x}|\bar{\Psi}_v[\ell, i], 2i - 1)) - (1 - \zeta(\ell, i)) H(\bar{x}|\bar{\Psi}_v[\ell, i] + 2i)) + o(n).$$

In (b) we applied (20) and in (c) (a) we applied a combination of Lemma 2 and observed from (20) that $\bar{\Psi}[\ell + 1, 2i - 1]$ or $\bar{\Psi}[\ell + 1, 2i]$ can only increase conditioning relative to $\bar{\Psi}[\ell, i], \bar{w}_{v[\ell, i]}.$
We now address the base cases (i.e., when \( \ell = K - 1 \)). We first observe that at the \( K - 1 \) level, because all paths from root to leaf are permutation of all user indices, if \( j \notin \{ v[K - 1, i], v[K, i] \} \) then \( \eta_{v[K - 1, i]} = 1 \). Equivalently, if \( j \notin \{ v[K - 1, i], v[K, i] \} \) then \( \eta_{v[K - 1, i]} = (i_j', 1) \).

Now recall that if \( i \) is odd, then \( \eta_{v[K - 1, i]}(K - 1, i) = 1 \), and from Fano we have

\[
\begin{align*}
nr_{v[K - 1, i]} &\leq I(\hat{x}, \hat{\Psi}_{v[K - 1, i]}; \hat{w}_{v[K - 1, i]}) + o(n) \\
&\leq I(\hat{x}, \hat{\Psi}_{K - 1, i}; \hat{w}_{v[K - 1, i]}) + o(n) \\
&= \sum_{j \neq v[K - 1, i]} H(\hat{x}) - H(\hat{x}|\hat{\Psi}_{v[K], i}, \{ w_{j} \}_{j \neq v[K, i]}) + o(n) \\
&\leq H(\hat{x}) - H(\hat{x}|\hat{\Psi}_{v[K], i}, \{ w_{j} \}_{j \neq v[K, i]}) + o(n),
\end{align*}
\]

where in step (d) we applied (4). If \( i \) is even, then \( \eta_{v[K - 1, i]}(K - 1, i) \) can be less than 1, and from Fano we have

\[
\begin{align*}
nr_{v[K - 1, i]} &\leq I(\hat{x}, \hat{\Psi}_{v[K - 1, i]}; \hat{w}_{v[K - 1, i]}) + o(n) \\
&\leq I(\hat{x}, \hat{\Psi}_{K - 1, i}; \hat{w}_{v[K - 1, i]}) + o(n) \\
&= \sum_{j \neq v[K - 1, i]} H(\hat{x}) - H(\hat{x}|\hat{\Psi}_{v[K], i}, \{ w_{j} \}_{j \neq v[K, i]}) + o(n) \\
&\leq H(\hat{x}) - H(\hat{x}|\hat{\Psi}_{v[K], i}, \{ w_{j} \}_{j \neq v[K, i]}) + o(n),
\end{align*}
\]

Finally, at the \( K \)-th level of the OBT we have trivially

\[
\begin{align*}
nr_{v[K], i} &\leq I(\hat{x}, \hat{\Psi}_{v[K], i}; \hat{w}_{v[K], i}) + o(n) \\
&\leq I(\hat{x}, \hat{w}_{j} \neq v[K, i]; \hat{w}_{v[K], i}) + o(n) \\
&\leq H(\hat{x}|\hat{w}_{j} \neq v[K, i]) + o(n).
\end{align*}
\]

Using (63)–(67), and scaling expressions according the the coefficients \( \zeta[\ell, i] \) at each node, we recursively arrive at

\[
\begin{align*}
nr_{v[K], i} &\leq I(\hat{x}, \hat{\Psi}_{v[K], i}; \hat{w}_{v[K], i}) + o(n) \\
&\leq I(\hat{x}, \hat{w}_{j} \neq v[K, i]; \hat{w}_{v[K], i}) + o(n) \\
&\leq H(\hat{x}|\hat{w}_{j} \neq v[K, i]) + o(n),
\end{align*}
\]

D. Capacity Regions Achievable using Conventional Random Codes

Random linear codes are well studied in many network coding problems and a proof of achievability under the same principles of the scheme in the BIC context (i.e., all users decoding all messages) may be found, for instance, in [17].

We apply the achievable rate condition (22) to arrive at the following two propositions:

**Proposition 9.** Consider a two-user BIC defined by parameters \( \mu_{12} \) and \( \mu_{21} \). The capacity region is the set of all rate pairs \( (r_1, r_2) \) satisfying

\[
\begin{align*}
r_1 + \mu_{12}r_2 &\leq 1, \\
\mu_{21}r_1 + r_2 &\leq 1.
\end{align*}
\]

*Proof:* The converse results from Theorem 1 by letting \( i, j \in \{1, 2\} \), \( k = 3 \) and fixing \( r_3 = 0 \), while achievability is a result of evaluation of (22). \( \square \)

**Proposition 10.** Consider a \( K \)-user BIC where \( \mu_{ij} = \mu \) for all \( i \neq j \). The capacity region is the set of all rate tuples \( (r_1, \ldots, r_K) \) satisfying for every \( i \in \{1, \ldots, K\} \)

\[
r_1 + \mu \sum_{j \neq i} r_j \leq 1.
\]

*Proof:* Achievability results directly from evaluation of (22). To prove the converse, we observe that when \( \mu_{ij} = \mu \) for all \( j \neq i \), for all \( \ell \)

\[
\zeta[\ell, i] = 0,
\]
Evaluating recursively through the OBT yields
\[ Γ_A[1, 1] = r_{v[1,1]} + μr_{v[2,2]} + \ldots μr_{v[K,2^K]} ≤ 1. \]

Since the path from root to leaf is a permutation of user indices (i.e., all user indices are represented and there exist no repeats), we arrive at (70).

E. Proof of Lemma 6

Consider \((ρ_1, L_1, α_1)\) fixed and rate pair \((r_1, r_2)\) such that (37)-(40) are satisfied. To prove the lemma, we demonstrate that there exists a scheme and that using this scheme \((r_1, r_2)\) is achievable according to Definition II. To do so, we first show that if (37) is satisfied, then a sequence of \(U_1\) matrices (and by proxy encoding functions) exists. We then argue that using the described encoding and decoding strategies with \((ρ_2, L_2, α_2) = (1, 0, 0)\), the usual equation counting argument applied to conventional random codes in erasure channels suffices to prove that error probability vanishes as \(n → ∞\).

If (37) is satisfied, then for every \(n\), we may choose \(m_1^{(n)} = \lfloor nr_1 \rfloor\), thereby satisfying (36) and guaranteeing the existence of a RRP matrix for \(w_1\). Furthermore, we see that \(\lim_{n → ∞} \frac{nr_1}{n} = r_1\). Bearing this in mind, we let \(m_2^{(n)} = \lfloor nr_2 \rfloor\) for each \(n\) and consider the scheme that uses RRP matrices with parameters \((ρ_1, L_1, α_1)\) and \((ρ_2, L_2, α_2)\) to encode \(w_1\) and \(w_2\) respectively.

Recall that in our decoding strategy, \(w_2\) is decoded at each receiver by first extracting clean equations of \(w_2\) (where by clean, we mean the contribution of \(w_2\) can be canceled out as explained previously). Then, the next step is decoding \(w_2\) from the clean equations. We now calculate the probability that at time \(t\) Receiver \(i\), where \(i = 1, 2\), can extract a clean equation of \(w_2\) for three cases:

- If \(t = (ℓ - 1)m_1^{(n)} + k\), where \(ℓ\) is a positive integer with \(ℓ ≤ L_1\) and \(1 ≤ k ≤ m_1^{(n)}\), then \(U_1(t, :)w_1 = \ell\)-th repetition of \(w_1[k]\). The probability of receiving a clean equation at User 1 is \(η_1(t) = (1 - ϵ_1)(1 - ϵ_1^{λ-1})\), and at User 2 is \(η_2(t) = (1 - ϵ_2)(1 - ϵ_2^{λ-1})\), which are the products of probabilities that User \(i\) at time \(t\) receives an unerased transmission and has previously received a transmission (or side-information in at User 2) containing the same \(w_1[k]\).

- If \(L_1 m_1^{(n)} < t ≤ L_1 m_1^{(n)} + ρ_1 n\), then \(U_1(t, :)w_1\) is random combination of \(w_1\) bits, and \(η_1(t) = η_2(t) = 0\) (i.e., to decode \(w_2\), each receiver ignores these transmissions).

- If \(L_1 m_1^{(n)} + ρ_1 n < t\), then \(U_1(t, :)w_1 = 0\), and the probability of a clean equation is the probability of an unerased transmission: \(η_1(t) = 1 - ϵ_1\) and \(η_2(t) = 1 - ϵ_2\).

From capacity analysis of point-to-point erasure channels, we note that a random linear coded message, \(m_2\), with rate \(r_2\) is decodable at User 1 with arbitrarily low probability of error as \(n → ∞\) if the number of received random linear equations of \(w_2\) is sufficiently large. Specifically, by counting the number of clean random linear equations of \(w_2\) received by User 1, we see that \(r_2\) must satisfy:

\[
\lim_{n → ∞} \frac{1}{n} \sum_{t=1}^{n} η[t] ≥ \frac{1}{n} \sum_{k=1}^{α_1 m_1^{(n)}} \sum_{ℓ=1}^{L_1} (1 - ϵ_1)(1 - ϵ_1^{λ-1}) + \sum_{k=α_1 m_1^{(n)} + 1}^{m_1^{(n)}} \sum_{ℓ=1}^{L_1} (1 - ϵ_1)(1 - ϵ_1^{λ-1}) + (n - L_1 m_1^{(n)}) - α_1 m_1^{(n)} - ρ_1 n(1 - ϵ_1) = (1 - ϵ_1)(1 - ρ_1) - r_1 [1 - ϵ_1^{λ_1} + α_1(ϵ_1^{λ_1} - ϵ_1^{λ_1+1})].
\]

Through analogous analysis, we find that communicating of \(w_2\) to User 2 is possible with arbitrarily low error probability if \(r_2 ≤ (1 - ϵ_2)(1 - ρ_2) - r_1 μ [1 - ϵ_2^{λ_2} + α_1(ϵ_2^{λ_2} - ϵ_2^{λ_2+1})]\).

Notice that (72), (73) are equivalent to (39), (40) and thus if both expressions are satisfied, then \(w_2\) is decodable at each receiver with high probability.

We now address achievability of \(r_1\) assuming that User 1 has already successfully decoded and canceled \(w_2\) from its received signal. It is sufficient to show that \(r_1\) satisfying any one of (37)-(40) is achievable, and we do so by proving achievability of \(r_1\) satisfying (38). Observe that the repetition portion of the RRP matrix supplies through the erasure channel a subset of message bits to User 1. Even if the repetition-code-supplied bits are removed, note that the random linear code portion of the RRP matrix still represents a random linear code applied to bits unknown after repetition. To decode, the total number of bits received though repetition and linearly independent equations received must be equal to \(m_1^{(n)}\).
Using Hoeffding’s inequality [18], one can show that with high probability as \( n \to \infty \) the number of bits received through repetition coding is concentrated around its mean,

\[
\sum_{k=1}^{\alpha_1 m_1^{(n)}} \epsilon_1^{L_1+1} + \sum_{k=\alpha_1 m_1^{(n)}+1}^{m_1^{(n)}} \epsilon_1^{L_1} = m_1^{(n)} \left( \epsilon_1^{L_1} - \alpha_1 (\epsilon_1^{L_1} - \epsilon_1^{L_1+1}) \right).
\]

Therefore, after the repetition phase, approximately \( m_1^{(n)} \left( 1 - \epsilon_1^{L_1} + \alpha_1 (\epsilon_1^{L_1} - \epsilon_1^{L_1+1}) \right) \) bits remain to be communicated using the random linear coding phase.

Through the usual argument that random linear combinations are independent w.h.p. as \( n \to \infty \), the random coding portion of the scheme supplies approximately \( \rho_1 n (1-\epsilon_1) \) equations. Therefore, as \( n \to \infty \), we may expect the random linear coding portion to resolve all message bits of \( \vec{w}_1 \) that were not received during the repetition phase if

\[
m_1^{(n)} \left( 1 - \epsilon_1^{L_1} + \alpha_1 (\epsilon_1^{L_1} - \epsilon_1^{L_1+1}) \right) \leq \rho_1 n (1-\epsilon_1),
\]

or simply a rate \( r_1 \) is achievable if it satisfies

\[
r_1 = \lim_{n \to \infty} \frac{m_1^{(n)}}{n} \leq \rho_1 \frac{1-\epsilon_1}{\epsilon_1^{L_1} - \alpha_1 (\epsilon_1^{L_1} - \epsilon_1^{L_1+1})}.
\]

\(\square\)