Align-and-Forward Relaying for Two-hop Erasure Broadcast Channels

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

[*]University of Southern California, Los Angeles, CA, USA

[†]Bell Labs, Alcatel-Lucent, Holmdel, NJ, USA

Abstract—We consider the problem of broadcast over wireless erasure networks. To understand the challenges and opportunities of these setups, we study a two-hop erasure broadcast channel consisting of a single source, two relays, and two destinations desiring independent messages. In our network, no transmitter has channel state knowledge of erasures on outgoing links (i.e., no CSIT): The source has no knowledge of any channel state, each relay only has knowledge of the channel states of its incoming link, and destinations are provided with full channel knowledge.

We propose a scheme, referred to as *Align-and-Forward*, that exploits the (unknown) common subspace of received signals at the relays, which results from the source-to-relay broadcast, in order to minimize the dimension of the interference subspace at each destination. We show that Align-and-Forward outperforms available alternative schemes in terms of sum-rate. We also present new outer-bounds and demonstrate the optimality of Align-and-Forward in certain regimes.

I. INTRODUCTION

Capacity analysis of erasure networks can provide valuable insight into protocol design for packetized wireless networks and thus has been the subject of much research. For instance, in [1] it was shown that the capacity of unicast or multicast in wireless erasure networks is achievable through random linear network coding at all nodes, implying that it is optimal for all intermediate relays to act as "dumb" mixers to enable all destinations to decode the packets. However, as one goes beyond multicast to multiple unicast or even broadcast scenarios, the capacity of packetized wireless networks and their optimal communication protocols remain, in general, an open problem.

In order to gain broader insight into the challenges posed by general wireless erasure networks, we consider broadcast in a two-hop erasure network, containing a single source, two relays, and two destinations (shown in Figure 1). We assume all transmitting nodes (source and relays) are provided with no knowledge of erasures affecting their own transmissions; thus, our model is one of no channel state information at transmitters (i.e., no CSIT) and no feedback. Relays have knowledge only of erasures on their incoming link, whereas destinations have full knowledge of all channel states. This problem remains one of the simplest unsolved cases in wireless erasure networks (for one-hop broadcast, the full capacity region is achievable through time-division and random linear

codes [2]) and captures communication challenges posed by: 1) multi-hop communication paths, 2) multiple simultaneous communication sessions, and 3) distributed relaying.

For the two-hop erasure broadcast channel described above, we propose a transmission scheme which relies on a novel relaying strategy referred to as Align-and-Forward. In all network regimes, Align-and-Forward either matches or outperforms the best known scheme proposed in [2]. Align-and-Forward relies on a form of intersession coding that exploits the (unknown) overlap in bits transmitted by the source and received by both relays in the first hop of the network. Relays use these overlapping bits to form and broadcast bits of common interest to both destinations. It is important to note that relays are able to exploit the overlap despite not knowing which bits are overlapping at both relays. Simply stated, Align-and-Forward implements a probabilistic form of interference alignment at the relays without explicit knowledge of the targeted alignment subspace. Although it is known that intersession coding can provide capacity gain in erasure broadcast networks, all prior works have relied on *feedback* to achieve the gains (e.g., [3]–[5] and references therein). Quite interestingly, Align-and-Forward demonstrates a gain from intersession coding without requiring feedback.

We also develop new upper bounds on the sum-capacity of the two-hop erasure broadcast channel. Our bounds leverage a recent result of [6] which captures the entropy "leakage" of information from a transmitter to an unintended receiver in erasure networks with no CSIT. We use the lemma, along with a genie-aided construction, to develop complementary inequalities, that when summed, yield two novel upper bounds.

II. NETWORK MODEL

We study the wireless erasure network depicted in Figure 1, which consists of a source, two relays, and two destinations. Each node's transmission is broadcast to all connected receiving nodes, while each received sequence is corrupted by random symbol (packet) erasures. At a receiving node, sequences from different transmitters are received orthogonally (i.e, without interference). Channel input symbols are binary and erasures occur independently on each transmit-receive link. As shown in Figure 1, in this paper, we consider a symmetric network, parametrized by the erasure probability tuple, $(\epsilon_1, \epsilon_2, \epsilon_3)$. Without loss of generality, we assume $\epsilon_2 \leq \epsilon_3$.

Formally, at time t, let $X_A[t] \in \{1, 2\}$ denote the channel input at node A and $Y_{BA}[t] \in \{0, 1, 2\}$ denote the channel output on the link from node A to node B. Let $G_{BA}[t] \in \{0, 1\}$ denote a channel state, indicating whether an erasure occurs (i.e., $G_{BA}[t] = 0$) on the link from node A to node B at time t. Channel input-output relationships are given by

$$Y_{BA}[t] = G_{BA}[t] X_A[t]. \tag{1}$$

We assume codes of length n, and we use the notation X_A^n to refer to the vector of channel inputs $X_A^n = [X_A[1] \quad \ldots \quad X_A[n]]$. Accordingly, we define Y_{BA}^n and G_{BA}^n. For each pair of nodes A and B, G_{BA}^n is an i.i.d. Bernoulli process with parameter $1 - \epsilon_k$, where ϵ_k is specified in Figure 1. Finally, we denote the tuple of all channel state variables as $\overline{G}^n = (G_{R_1S}^n, G_{R_2S}^n, G_{D_1R_1}^n, G_{D_1R_2}^n, G_{D_2R_1}^n, G_{D_2R_2}^n)$.
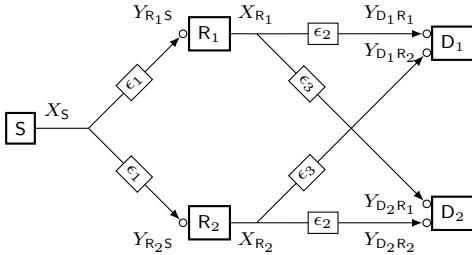

Fig. 1. The symmetric two-hop erasure broadcast channel.

We assume destinations have full channel state information (CSI) for the entire the network ($\overline{G}^n$), and relays have CSI only of their own incoming link, i.e., Relay 1 knows only $G_{R_1S}^n$ and Relay 2 knows only $G_{R_2S}^n$. This means that each relay is unaware of the erasures within the other's received signal. The source has no CSI whatsoever.[1]

We are interested in the sum-capacity of the broadcast messaging scenario, where the messages desired by Destination 1 and Destination 2, M_1 and M_2 respectively, are independent.

III. ACHIEVABLE SCHEME

In this section, we present a new scheme for the two-hop erasure broadcast channel. The innovative aspect of our approach is a relaying strategy referred to as Align-and-Forward.

For comparison, we first summarize the best known scheme of [2], which we refer to as the time-division (TD) approach since it is based on time-division between messages and random linear coding.

Time-Division Scheme: In the TD scheme, each node allocates a portion of transmissions to either M_1 or M_2. We use the variable α_A to denote the proportion of time dedicated to communicating M_1 at a transmitting node A, and consequently $1 - \alpha_A$ represents the time allocated to communicating M_2. The result is a separation of the two-hop erasure broadcast

channel into two virtual unicast erasure networks, one for each message. Figure 2 depicts the virtual network for M_1, with new erasure probabilities:

$$\epsilon_0^{\mathsf{TD}} = 1 - \alpha_{\mathsf{S}}$$
$$\epsilon_1^{\mathsf{TD}} = \epsilon_1$$
$$\epsilon_2^{\mathsf{TD}} = \alpha_{\mathsf{R}_1} \epsilon_2 + (1 - \alpha_{\mathsf{R}_1})$$
$$\epsilon_3^{\mathsf{TD}} = \alpha_{\mathsf{R}_2} \epsilon_3 + (1 - \alpha_{\mathsf{R}_2}).$$
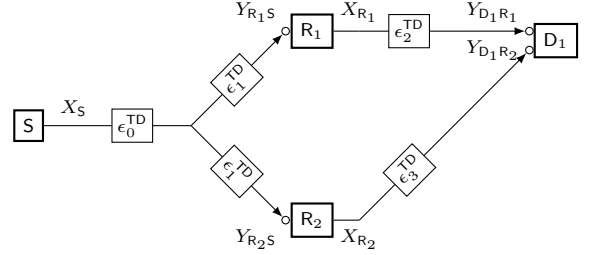

Fig. 2. The virtual S-D_1 network emulated using the TD scheme.

Within the virtual network allocated to a message, the source and relays utilize the random linear codes to construct channel inputs, and the maximum rate is given by the min-cut of the virtual network. Given network parameters ϵ_1, ϵ_2, and ϵ_3, optimization of the time-divisions to maximize sum-rate in the two-hop erasure broadcast channel yields the rate [2]:

$$r_\Sigma^{\mathsf{TD}} = \min \left\{ 1 - \epsilon_1^2, (1 - \epsilon_1) + (1 - \epsilon_3) \left[2 - \frac{1 - \epsilon_1}{1 - \epsilon_2} \right]_+, 2 - 2\epsilon_2 \right\}. \tag{2}$$

As an example, when $\epsilon_1 = \epsilon_2 = \frac{1}{2}$ and $\epsilon_3 = \frac{4}{5}$, the TD approach achieves a rate of $r_\Sigma^{\mathsf{TD}} = \frac{7}{10}$ bits by choosing $\alpha_{\mathsf{S}} = \alpha_{\mathsf{R}_1} = \alpha_{\mathsf{R}_2} = \frac{1}{2}$.

Our scheme also utilizes random linear encoding and time-division at the source, but achieves a sum-rate gain by improving the approach taken at the relays: Intersession encoding is used to create transmissions simultaneously useful to both destinations. For these "bits of common interest" from Relay 1, the component that is undesirable to Destination 2 (i.e., bits about M_1) is *aligned* with what Destination 2 receives from Relay 2. Similarly, Relay 2 aligns parts of M_2 with what Destination 1 receives from Relay 1. We note that because each relay is unaware of what the other has received, the exact subspace that should be targeted for alignment is unknown to each relay. Even more surprisingly, our scheme demonstrates that alignment is possible without relays even knowing which of their transmissions are aligned.

Because the general Align-and-Forward scheme involves many steps, we first present its construction and sum-rate gain with an illustrative example. For simplicity, we present the example assuming large n and invoking law of large numbers. Additionally, we use calligraphic typesetting to distinguish the vectors of inputs (e.g. $\mathcal{X}$), unerased outputs (e.g. $\mathcal{Y}$), and encoded bits (e.g. $\mathcal{Q}$ or $\mathcal{W}$) of our scheme versus the general vector input-output variables of the channel.
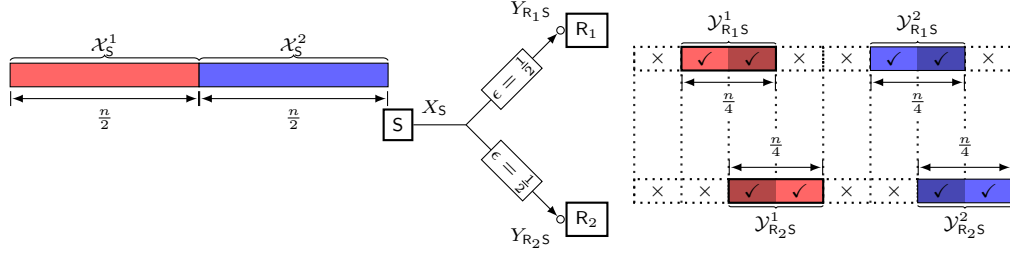
Fig. 3. First-hop transmission scheme: Although erasures occur randomly, bits received and bits lost to erasures at both relays are grouped into solid and dotted blocks respectively. Colors used to denote message content: red for M_1 and blue for M_2. Bits that overlap (received by both relays) are depicted as shaded blocks, but overlap is not known to either relay.
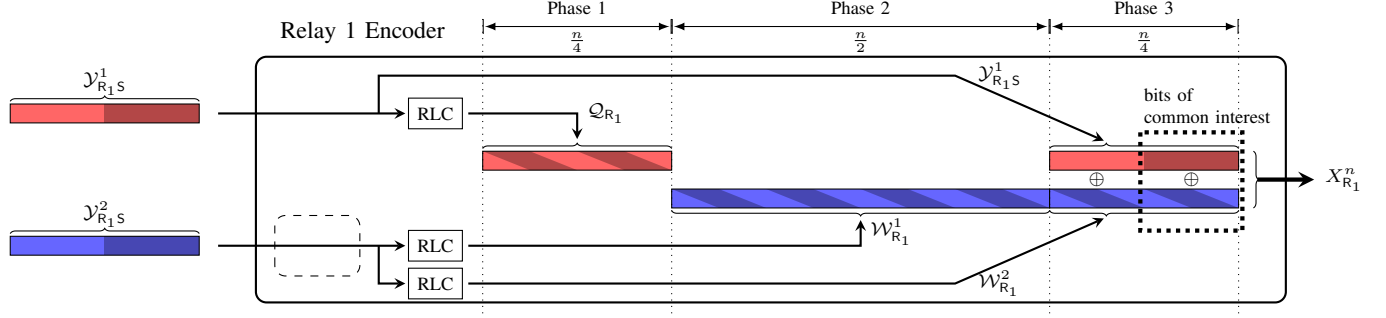

Fig. 4. Relaying scheme employed at Relay 1: Message M_1 content in red, M_2 content in blue. Shaded areas represent RLC bits overlapping at the relays. The lower left dashed box represents a processing block (shown in Figure 6) not needed for the example but necessary for the general A&F scheme.
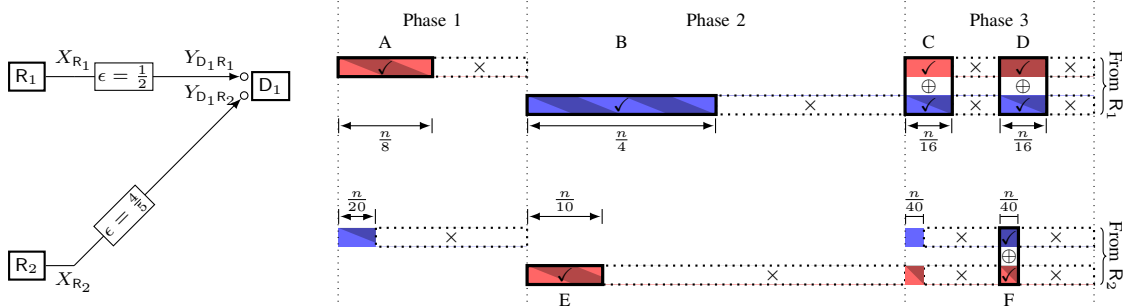

Fig. 5. Second-hop transmission and decoding at Destination 1: Message M_1 content in red and M_2 content in blue. Shaded areas represent RLC bits overlapping at the relays. Bold-outlined blocks with alphabetical labels are used to decode M_1.

A. Illustrative Example

Let $\epsilon_1 = \epsilon_2 = \frac{1}{2}$ and $\epsilon_3 = \frac{4}{5}$. Recall that for this network, the TD scheme achieves $r_\Sigma^{TD} = \frac{7}{10}$ bits. We show that our scheme improves upon this, achieving a sum rate of $r_\Sigma^{A\&F} = \frac{3}{4}$ bits, which is indeed the sum-capacity for this example.

Let messages M_1 and M_2 both consist of $\frac{3n}{8}$ bits, where n is a large integer. For ease of exposition, we ignore terms of order $o(n)$, which vanish in the calculation of rate per channel use. The source applies random linear encoding on each message to create $\frac{n}{2}$ random linear combination (RLC) bits. Denote with $\mathcal{X}_S^i$ the vector of RLC bits created for message M_i. Elements of $\mathcal{X}_S^1$ and $\mathcal{X}_S^2$ are broadcast sequentially by the source, as shown in Figure 3. Each relay receives approximately $\frac{n}{4}$ unerased bits per message, M_i, denoted with the vectors $\mathcal{Y}_{AS}^i$ for $A \in \{R_1, R_2\}$. Because erasures occur independently, approximately $\frac{n}{8}$ RLC bits are received at both relays (overlapping) for each message. However, due to lack of knowledge, each relay is unaware of which RLC bits have been received by the other relay.

Relay 1 re-encodes received RLC bits using the following three-phase transmission scheme (depicted in Figure 4):

Phase 1 ($\frac{n}{4}$ **channel uses**): Relay 1 applies a random linear code upon $\mathcal{Y}_{R_1S}^1$, and creates a vector, $\mathcal{Q}_{R_1}$ depicted as the red striped block, of $\frac{n}{4}$ RLC bits describing M_1, and broadcasts these sequentially.

Phase 2 ($\frac{n}{2}$ **channel uses**): Relay 1 applies a random linear code upon $\mathcal{Y}_{R_1S}^1$, and creates a vector, $\mathcal{W}_{R_1}^1$ depicted as the blue striped block, of $\frac{n}{2}$ RLC bits describing M_2, and broadcasts these sequentially.

Phase 3 ($\frac{n}{4}$ **channel uses**): Relay 1 applies another random linear code upon $\mathcal{Y}_{R_1S}^2$, and creates a vector, $\mathcal{W}_{R_1}^2$, of $\frac{n}{4}$ RLC bits describing M_2. Each of these is XORed with one (*uncoded*) element of $\mathcal{Y}_{R_1S}^1$ and broadcasted.

As we will see later, Phase 3 increases the efficiency of the scheme, versus TD approach, by creating messages which are interesting to both receivers. However, there are two subtle points in Phase 3. The first one is that not all bits created in this phase are of common interest. Indeed all of these bits created at Relay 1 will be useful to Destination 1 (as will be explained later), however *only part of them* are also useful for

Destination 2. This part is shown in Fig. 4 in bold-dotted box. Remember that one part of $\mathcal{Y}_{R_1 S}^1$, colored by shaded red, is also available at Relay 2. Only the bits that are created by XORing this shaded-red part of $\mathcal{Y}_{R_1 S}^1$ and RLC of $\mathcal{W}_{R_1}^1$ are of common interest. The reason is that some of these shaded-red bits have been previously communicated by Relay 2 in Phase 2, and overheard by Destination 2. Thus, Destination 2 can use them to clean some of the bits received from Relay 1 in Phase 3, and resolve its own (blue) message.

The second subtle point is that Relay 1 cannot identify the shaded-red part of $\mathcal{Y}_{R_1 S}^1$, due to its local knowledge of the channel state. Otherwise, in Phase 3, Relay 1 could use just the shaded-red part of $\mathcal{Y}_{R_1 S}^1$ and create messages that are all of common interest. This point justifies why we use *uncoded* versions of $\mathcal{Y}_{R_1 S}^1$ in Phase 3: Applying RLC over $\mathcal{Y}_{R_1 S}^1$ mixes the overlapping (shaded) part with the non-overlapping (unshaded) part. Since the unshaded part of $\mathcal{Y}_{R_1 S}^1$ has not been overheard by Destination 2, it cannot be canceled.

Relay 2 (not depicted) uses the analogous three-phase scheme, but switches the roles of messages: a random linear code is applied to $\mathcal{Y}_{R_2 S}^2$ during Phase 1 to create $\mathcal{Q}_{R_2}$, two other random linear codes are applied to $\mathcal{Y}_{R_2 S}^2$ during Phases 2 and 3 to create $\mathcal{W}_{R_2}^1$ and $\mathcal{W}_{R_2}^2$ respectively, and elements of $\mathcal{W}_{R_2}^2$ are XORed with uncoded elements of $\mathcal{Y}_{R_2 S}^2$ during Phase 3.

Focusing on Destination 1, we now describe the procedure to decode M_1 by clarifying the nature of bits received during each phase from each relay, while referring to Figure 5.

Relay 1, Phase 1 (Block A of Fig. 5): Destination 1 receives approximately $\frac{n}{8}$ bits from $\mathcal{Q}_{R_1}$, which describe M_1.

Relay 1, Phase 2 (Block B of Fig. 5): Destination 1 receives approximately $\frac{n}{4}$ bits from $\mathcal{W}_{R_1}^1$, which carry information about M_2. With high probability, Destination 1 can recreate the original $\frac{n}{4}$ bits in $\mathcal{Y}_{R_1 S}^2$, *entirely*. Moreover, Destination 1 can use knowledge of all erasures in the network $(\overline{G}^n)$ to identify approximately half of these $(\frac{n}{8})$ as overlapping bits (colored shaded-blue).

Relay 1, Phase 3 (Blocks C & D of Fig. 5): Recall that, from the previous phase, Destination 1 already knows $\mathcal{Y}_{R_1 S}^2$. It uses this knowledge to remove the $\mathcal{W}_{R_1}^2$ component from Phase 3 transmissions and receives approximately $\frac{n}{8}$ bits from $\mathcal{Y}_{R_1 S}^1$, which all carry information about M_1. Therefore, as we mentioned before, all the bits sent by Relay 1 in this phase, and received by Destination 1, are useful to Destination 1.

Relay 2, Phase 1: Destination 1 ignores these packets.

Relay 2, Phase 2 (Block E of Fig. 5): Destination 1 receives approximately $\frac{n}{10}$ bits from $\mathcal{W}_{R_2}^1$, which describe M_1.

Relay 2, Phase 3 (Block F of Fig. 5): The bits received from Relay 2 during this phase, can be split into two parts. The part which is useful at Destination 1 is denoted by F. This part has been formed at Relay 2 by XORing RLC of $\mathcal{Y}_{R_2 S}^1$ with uncoded shaded-blue part of $\mathcal{Y}_{R_2 S}^2$. The reason is as follows. Recall that Block B was enough for Destination 1 to resolve all of $\mathcal{Y}_{R_1 S}^2$. Due its full access to channel state information, Destination 1 can recognize the shaded-blue part of $\mathcal{Y}_{R_1 S}^2$ (M_2 bits overlapping at both relays). Therefore, such knowledge is used by Destination 1

to clean interference from Block F. The number that can be recovered is $(1 - \epsilon_3)\frac{1}{2} \times \frac{n}{4} = \frac{n}{40}$. The coefficient $\frac{1}{2}$ comes from the fact only half of the messages received in this phase from Relay 2 falls in Block F.

In order to determine the rate achieved for M_1 we now determine the maximum number of linearly independent RLC bits received by Destination 1. First we observe that Destination 1 can decode all $\frac{n}{4}$ elements of $\mathcal{Y}_{R_1 S}^1$ from what it receives in Phase 1 and Phase 3 from Relay 1. Also it receives $\frac{n}{10} + \frac{n}{40} = \frac{n}{8}$ RLC bits describing M_1 from Relay 2 in Blocks E and F. Therefore, in total, it has $\frac{3n}{8}$ RLC which can be shown are independent and thus are enough to recover M_1. The same argument is valid for Destination 2. Therefore, we achieve the sum-rate of $r_\Sigma = 2 \times \frac{1}{n}\frac{3n}{8} = \frac{3}{4}$.

B. General Scheme

We now describe the end-to-end general scheme, which requires an additional step within the Align-and-Forward relaying scheme. This step was unnecessary for the example, but it allows us to guarantee a rate either matching or outperforming the TD scheme for all network instances.

Source Encoding & Transmission: Each message, M_1 and M_2, consists of $n\frac{r_\Sigma^{\text{A\&F}}}{2}$ bits, where $r_\Sigma^{\text{A\&F}}$ is specified later. For each M_i, the source applies a random linear code and creates $\frac{n}{2}$ RLC bits, and broadcasts these sequentially.

Align-and-Forward Relaying: Recall that in the first hop, each relay (i) received a sequence $Y_{R_i S}^n$, which contains approximately $n(1 - \epsilon_1)$ unerased bits. Among these, approximately $\frac{n}{2}(1 - \epsilon_1)$ describe M_1 and $\frac{n}{2}(1 - \epsilon_1)$ describe M_2. Following the notation of the example, we refer to the unerased bits at Relay i as vectors $\mathcal{Y}_{R_i S}^1$ and $\mathcal{Y}_{R_i S}^2$, where $\mathcal{Y}_{R_i S}^1$ carries information for M_1 and $\mathcal{Y}_{R_i S}^2$ carries information for M_2. Due to independence of erasure events, overlap in received transmissions occurs: of the approximately $n(1 - \epsilon_1)$ bits received by Relay 1, approximately $n(1 - \epsilon_1)^2$ have also been received by Relay 2, and vice versa. Relays, however, do not know which bits overlap (i.e., shaded bits in Figure 4).

Align-and-Forward consists of three transmission phases, with the proportion of channel uses allocated to Phases 1 denoted as τ_1, Phase 2 denoted as τ_2, and Phase 3 denoted as τ_3, where τ_1, τ_2, and τ_3 are given by

$$\tau_1 \triangleq \min\left\{\frac{1 + \epsilon_1}{2}, \frac{(1 - \epsilon_1)\epsilon_2}{2(1 - \epsilon_2)}\right\}, \quad (3)$$

$$\tau_2 \triangleq \left[1 - \frac{1 - \epsilon_1}{2(1 - \epsilon_2)}\right]_+, \quad (4)$$

$$\tau_3 \triangleq \frac{1 - \epsilon_1}{2}. \quad (5)$$

Note that although $\tau_1 + \tau_2 + \tau_3 = 1$, the exact number of channel uses per phase must be an integer. Therefore, we allot $\lfloor n\tau_1 \rfloor$, $\lfloor n\tau_2 \rfloor$, and $\lfloor n\tau_3 \rfloor$ channel uses to Phases 1, 2, and 3, and stipulate that during any channel uses remaining from the total of n, the relays remain silent.

We first summarize objectives of each transmission phase before describing the Align-and-Forward relaying scheme in detail:

- During Phase 1 the transmissions of Relay i carry only information about message M_i. These are useful Destination i, but will be ignored by Destination i' ($i' = 3 - i$).
- During Phase 2 the transmissions of Relay i carry only information about message $M_{i'}$. Clearly these transmissions are useful to Destination i'. An important detail is that even though these transmissions are not immediately useful for Destination i, still, Destination i overhears and saves and saves them. These bits will be used in Phase 3 as side information, which allows relays to create transmission bits which are useful for both Destinations, thus increasing efficiency.
- Finally, during Phase 3, each relay applies an intersession coding technique to mix RLC bits from M_1 and M_2 such that:
  1) All transmissions from Relay i in this Phase are useful to Destination i. This is because any information that Relay i communicates about $M_{i'}$, it has already delivered to Destination i in Phase 2.
  2) Part of transmissions of Relay i in this phase are useful for Destination i'. This only includes the transmissions that combine RLC bits of M_i that are already available at Destination i'. Destination i' has overheard these components of M_i in Phase 2 from Relay i', because these were also received by Relay i' from the source in the first hop (shaded RLCs in Figure 5).

During Phase 1 a random linear code is applied to $\mathcal{Y}_{R_i}^i$ to create a vector $\mathcal{Q}_{R_i}$ at Relay i. Elements of $\mathcal{Q}_{R_i}$ are then broadcast sequentially. Note that the sequence of elements from $\mathcal{Q}_{R_i}$ are a codeword, generated by the random linear code. The codebook used by Relay i is a function of the erasures that have occured in the first-hop link, $G_{R_iS}^n$. Since destinations have full CSI, they will know which codebook has been used.

Phase 2, in general, does not proceed as in the example. We require an additional processing block (shown in Figure 6) which addresses the dual role of Phase 2 transmissions. Recall that Phase 2 transmissions from Relay i provide message information to Destination i' and side information to Destination i. In the example, this was easily accomplished, since $\epsilon_2 = \epsilon_1$ and we claimed that Destination i was able to decode all of $\mathcal{Y}_{R_iS}^{i'}$ during Phase 2. However, if $\epsilon_2 > \epsilon_1$, the time alloted to Phase 2 in (4) is insufficient to communicate the entire vector $\mathcal{Y}_{R_iS}^{i'}$. Therefore, in our scheme, each Relay i reduces the number of bits about $\mathcal{Y}_{R_iS}^{i'}$ communicated to Destination i. Before specifying the method of reduction, we point out a tradeoff between parts of $\mathcal{Y}_{R_iS}^{i'}$ useful to Destination i and parts useful to Destination i'.

Destination i desires the overlapping bits from $\mathcal{Y}_{R_iS}^{i'}$ to use as side information to cancel interference terms in Phase 3. On the other hand, Destination i' desires the non-overlapping bits from $\mathcal{Y}_{R_iS}^{i'}$ because these are bits that are not received by, and therefore cannot be communicated by, Relay i'. Recall that Relay i is unable to distinguish between overlapping and non-overlapping bits, and consequently applying a random linear

code would mix overlapping and non-overlapping bits. It is impossible for Relay i' to align with such a mixture, since it did not receive the non-overlapping bits, so coding elements of $\mathcal{Y}_{R_iS}^{i'}$ reduces the number of alignment opportunities in Phase 3 transmissions. On the other hand, coded (mixed) RLC bits of $\mathcal{Y}_{R_iS}^{i'}$ are more likely to be useful to Destination i' during Phase 2.

Our method of reduction balances the effect on each phase to create a vector of bits at Relay i, $\mathcal{V}_{R_i}$ which is meant to be fully decoded by Destination i. The block diagram for the approach is shown in Figure 6 which is inserted (for the general scheme) into the lower left dashed region of Figure 4. We define two reduction parameters ν_u and ν_c, which represent the amount of information about $\mathcal{Y}_{R_iS}^{i'}$ that is uncoded and coded, respectively. Notice that ν_u and ν_c are parameters used at both relays, but are applied to opposite message RLCs.

Two operations are applied at Relay i to $\mathcal{Y}_{R_iS}^{i'}$. In the first, $\lfloor n\nu_u \rfloor$ RLC bits are randomly *selected* from $\mathcal{Y}_{R_iS}^{i'}$ to create a subvector $\mathcal{V}_{R_i}^u$. In the second, a random linear code is applied to $\mathcal{Y}_{R_iS}^{i'}$ to create a vector of $\lfloor n\nu_c \rfloor$ bits, denoted as $\mathcal{V}_{R_i}^c$. A random linear code is then applied to the concatenated vector of bits $\mathcal{V}_{R_i} = (\mathcal{V}_{R_i}^u, \mathcal{V}_{R_i}^u)$ to create the vector $\mathcal{W}_{R_i}^1$, which has cardinality $\lfloor n\tau_2 \rfloor$. Relay i broadcasts elements of $\mathcal{W}_{R_i}^1$ serially during Phase 2.

For our scheme, ν_u and ν_c are such that they are non-negative, satisfy

$$ n(\nu_u + \nu_c) \leq \min\left\{ \lfloor n\tau_2 \rfloor (1 - \epsilon_2), \left| \mathcal{Y}_{R_iS}^{i'} \right| \right\}, \quad (6) $$

and maximize the sum-rate (this is described in greater detail in Section III-C). The constraint (6) is meant to ensure that Destination i can decode all of $\mathcal{V}_{R_i}$, from only Phase 2 transmissions.

Finally, Phase 3 proceeds in the following manner. A random linear code is applied to $\mathcal{V}_{R_i}$ to create a vector $\mathcal{W}_{R_i}^2$ of RLC bits with cardinality $\lfloor n\tau_3 \rfloor$. Each element of $\mathcal{W}_{R_i}^2$ is XORed with a distinct element of $\mathcal{Y}_{R_iS}^i$ and broadcast serially. If there are not enough elements in $\mathcal{Y}_{R_iS}^i$, then the last $\lfloor n\tau_3 \rfloor - |\mathcal{Y}_{R_iS}^i|$ elements of $\mathcal{W}_{R_i}^2$ are broadcast without any intersession coding.
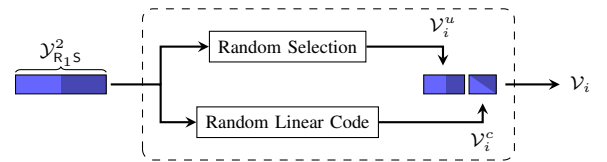

Fig. 6. Reduction processing block used by Relay 1. Notice that processing is applied to bits for M_2.

Decoding: We describe the decoding procedure as a two-stage process. In the first stage, each destination uses overheard, undesired information to "clean" intersession coded (i.e., Phase 3) transmissions. Specifically, Destination i first decodes the vector $\mathcal{V}_{R_i}$ from transmissions sent during Phase 2 from Relay i. All of $\mathcal{V}_{R_i}$ is used to remove the $M_{i'}$ component of broadcasts received from Relay i during Phase 3. The subvector $\mathcal{V}_{R_i}^u \subseteq \mathcal{V}_{R_i}$ is then used to remove the $M_{i'}$ component

from a portion of transmissions received from Relay i' during Phase 3.

Once interference has been canceled, Destination i is left with a vector of RLC bits that are a function of only M_i, the channel states $\overline{G}$, and the codes used at transmitting nodes. Stage 2 of decoding therefore requires that Destination i decode M_i using its full CSI, $\overline{G}^n$, and knowledge of codebooks used at each node.

C. Error Analysis & Achieved Sum-Rate

Recall that decoding occurs in two stages. For each we must clearly define the modes of failure. We define a Stage 1 error as the event where Destination i fails to correctly decode $\mathcal{V}_{R_i}$ from Phase 2 transmissions from Relay i. Let $\widehat{\mathcal{V}}_{R_i}$ denote the estimate of $\mathcal{V}_{R_i}$ at Destination i, and define the probability of Stage 1 error as

$$P_{\mathcal{E}_{i1}}^{(n)} \triangleq \Pr\left(\widehat{\mathcal{V}}_{R_i} \neq \mathcal{V}_{R_i}\right). \tag{7}$$

Communication of $\mathcal{V}_{R_i}$ from Relay i to Destination i during Phase 2, is equivalent to using block codewords of length-$\lfloor n\tau_2 \rfloor$ over a binary erasure channel with erasure probability ϵ_2. The capacity of the point-to-point binary erasure channel is well known [?], and we have that if

$$|\mathcal{V}_{R_i}| \leq \lfloor n\tau_2 \rfloor (1 - \epsilon_2), \tag{8}$$

then there exist random linear codes such that the probability of error $P_{\mathcal{E}_{i1}}^{(n)} \to 0$, as $n \to \infty$. By construction, $\mathcal{V}_{R_i}$ contains $(\lfloor n\nu_u \rfloor + \lfloor n\nu_c \rfloor)$ bits, and ν_u and ν_c satisfy (6). Hence, the condition holds.

We now argue that if Stage 1 proceeds without error, the probability of error in Stage 2 of decoding, which deals with RLC bits for only a single message, is less than or equal to probability of error in an equivalent unicast wireless erasure network. Consider the unicast network shown in Figure 7, and
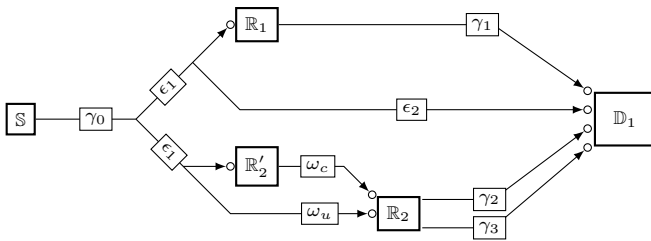

Fig. 7.   The virtual S-D$_1$ network emulated using the A&F scheme.

let erasure probabilities in the network be given by

$$\gamma_0 = \tfrac{1}{2}, \tag{9}$$

$$\gamma_1 = 1 - (1 - \epsilon_2)\tau_1, \tag{10}$$

$$\gamma_2 = 1 - (1 - \epsilon_3)\tau_2, \tag{11}$$

$$\gamma_3 = 1 - (1 - \epsilon_3)(1 - \epsilon_1)\nu_u, \tag{12}$$

$$\omega_u = 1 - \frac{2\nu_u}{1 - \epsilon_1}, \tag{13}$$

$$\omega_c = 1 - \nu_c. \tag{14}$$

We now claim the following:

Claim 1. *Assume that Stage 1 of decoding was successful and that τ_1, τ_2, τ_3, ν_u, and ν_c are fixed and satisfy (3)–(5) and (6) respectively. If there exists a set of length-n random linear block codes (one for each transmitting node) for the unicast wireless erasure network in Figure 7 that achieves a rate r_1 with probability of error $P'_{\mathcal{E}_{12}}$. Then, there exists a set of random linear codes for the two-hop broadcast channel that achieves rate r_1 with probability $P_{\mathcal{E}_{12}} \leq P'_{\mathcal{E}_{12}}$*

Proof: Notice that the network depicted in Figure 7 is constructed such that each incoming edge at $\mathbb{D}_1$ delivers approximately the number of unerased bits that one phase delivers to Destination 1 about M_1 in our scheme: The first (top) edge with erasure probability γ_1 corresponds to Phase 1 from Relay 1, the second edge corresponds to Phase 3 from Relay 1, the third edge corresponds to Phase 2 from Relay 2 and the fourth edge corresponds to Phase 3 from Relay 2. The same statistical equivalence can be seen in what $\mathbb{R}_1$, $\mathbb{R}'_2$, and $\mathbb{R}_2$ receive.

Now consider a length-n random linear code applied at each node in Figure 7, and apply it in our scheme for the two-hop erasure broadcast channel. To construct a code for each part of the two-hop erasure broadcast channel scheme, we puncture the corresponding code from the network of Figure 7.

- At the source in the two-hop erasure broadcast channel, a length $\frac{n}{2}$ code is created for M_1 by puncturing half of the RLC bits from the code used by $\mathbb{S}$ in Figure 7.
- For Phase 1 transmissions from Relay 1, $n - \lfloor n\tau_1 \rfloor$ bits are punctured from the code used by $\mathbb{R}_1$.
- For Phase 1 transmissions from Relay 1, $n - \lfloor n\tau_1 \rfloor$ bits are punctured from the code used by $\mathbb{R}_1$.
- The exact code used at $\mathbb{R}'_2$ is used in the coded portion of the reduction block of Relay 2.
- The code used for the first (top) channel input of $\mathbb{R}_2$ is punctured for use by Relay 2 for Phase 2.
- The code used for the second (bottom) channel input of $\mathbb{R}_2$ is punctured for use by Relay 2 for Phase 3.

Random puncturing results in statistics identical to that of the network in Figure 7, and thus achieves the same probability of error. If there exists a puncturing approach with lower probability of error, than using that puncturing approach can only improve the reliability of the linear code used in the two-hop erasure broadcast channel scheme. ∎

This allows us to leverage a main result of [1], which state that random linear codes are sufficient to achieve the capacity (cut-set outer bound) of the unicast problem (see Theorem 1 and note Remark 2 and Remark 4 of [1]). Evaluating the

capacity of the eight cuts possible in Figure 7 yields

$$C_{\{S\}} = \frac{1 - \epsilon_1^2}{2}, \tag{15}$$

$$C_{\{S,R_1\}} = \frac{1 - \epsilon_1}{2} + \frac{(1 - \epsilon_1)\epsilon_1(1 - \epsilon_2)}{2} + (1 - \epsilon_2)\tau_1, \tag{16}$$

$$C_{\{S,R_2\}} = \frac{1 - \epsilon_1}{2} + \nu_c + \nu_u \epsilon_1, \tag{17}$$

$$C_{\{S,R_2'\}} = \frac{1 - \epsilon_1^2}{2} + (1 - \epsilon_3)(\tau_2 + \nu_u(1 - \epsilon_1)), \tag{18}$$

$$C_{\{S,R_1,R_2\}} = (1 - \epsilon_2)\tau_1 + \frac{(1 - \epsilon_1)(1 - \epsilon_2)}{2}(1 - 2\nu_u) + \nu_c + \nu_u, \tag{19}$$

$$C_{\{S,R_1,R_2'\}} = (1 - \epsilon_2)\tau_1 + \frac{(1 - \epsilon_1)(1 - \epsilon_2)}{2}(\epsilon_1) + (1 - \epsilon_3)(\tau_2 + \nu_u(1 - \epsilon_1)) + \frac{1 - \epsilon_1}{2}, \tag{20}$$

$$C_{\{S,R_2,R_2'\}} = \frac{1 - \epsilon_1}{2} + (1 - \epsilon_3)(\tau_2 + \nu_u(1 - \epsilon_1)), \tag{21}$$

$$C_{\{S,R_1,R_2,R_2'\}} = (1 - \epsilon_2)\tau_1 + \frac{(1 - \epsilon_1)(1 - \epsilon_2)}{2} + (1 - \epsilon_3)(\tau_2 + \nu_u(1 - \epsilon_1)), \tag{22}$$

and taking the minimum of these yields a rate achievable through random linear codes. Some are clearly redundant (e.g., (18)), however, determining the minimum still requires evaluating (3)–(5) and optimizing over ν_u and ν_c. We point out that the suboptimal choice where $\nu_u = 0$ and $\nu_c = \min\left\{\lfloor n\tau_2 \rfloor(1 - \epsilon_2), \left|\mathcal{Y}_{R_i S}^{i'}\right|\right\}$, results in the TD rate; in particular, one out of the cuts (15), (21), or (22) will be active.

To ease exposition, we analyze three network regimes separately. For each each regime, we define ν_u and ν_c that maximize sum rate. The regimes and the resulting sum-rates are summarized in Table I.

Regime I ($\epsilon_2 \leq \epsilon_1$): When $\epsilon_2 \leq \epsilon_1$, we let $\nu_u = \frac{1 - \epsilon_1}{2}$ and $\nu_c = 0$. In this case cuts (15) and (21) are less than all others. The min of these is chosen as $r_i^{\mathsf{A\&F}}$.

Regime II ($2\epsilon_2 - 1 \leq \epsilon_1 < \epsilon_2$): When $2\epsilon_2 - 1 \leq \epsilon_1 < \epsilon_2$, optimization over ν_u and ν_c yields the optimal reduction parameters

$$\nu_u^{(II)} = (1 - \epsilon_2 - \frac{1 - \epsilon_1}{2})\min\{1, \frac{2 - \epsilon_2 - \epsilon_3}{(1 - \epsilon_1)(1 - \epsilon_2)(2 - \epsilon_3)}\} \tag{23}$$

$$\nu_c^{(II)} = (1 - \epsilon_2 - \frac{1 - \epsilon_1}{2}) - \nu_u^{(II)}. \tag{24}$$

In this case, one of the cuts (15), (17), or (21) is least. The min of these is chosen as $r_i^{\mathsf{A\&F}}$.

Regime III ($\epsilon_1 < 2\epsilon_2 - 1$): From (4), in Regime III Align-and-Forward relaying dedicates no time to sending the message of the weaker destination (i.e., $\tau_2 = 0$). Thus, in this regime $\mathcal{V}_{R_i} = \emptyset$, and Stage 1 decoding errors never occur (i.e., $P_{\mathcal{E}_{i1}} = 0$). Cut (22) is the least in this case, and the resulting per-user rate is $r_i^{\mathsf{A\&F}} = r_i^{\mathsf{TD}} = 1 - \epsilon_2$.

IV. UPPER BOUNDS

We now present a novel capacity upper bound for our setup:

Theorem 2 (Upper Bounds). *The sum-capacity, C_Σ, of the 2-hop erasure broadcast channel with $\epsilon_2 \leq \epsilon_3$ satisfies*

$$C_\Sigma \leq 1 - \epsilon_1 + 2\frac{(1 - \epsilon_2)(1 - \epsilon_3)}{2 - \epsilon_2 - \epsilon_3}, \tag{25}$$

$$C_\Sigma \leq \frac{(\epsilon_3 - \epsilon_2)(1 - \epsilon_1) + 4(1 - \epsilon_2)(1 - \epsilon_3)}{2 - \epsilon_2 - \epsilon_3}. \tag{26}$$

Our bound is, in general, not tight with the sum-rate achieved in our scheme. However, in network regimes where Align-and-Forward meets neither the cut-set nor the recent broadcast-cut upper bound of [5], our bound is often tighter than the existing bounds (see Figure 8).
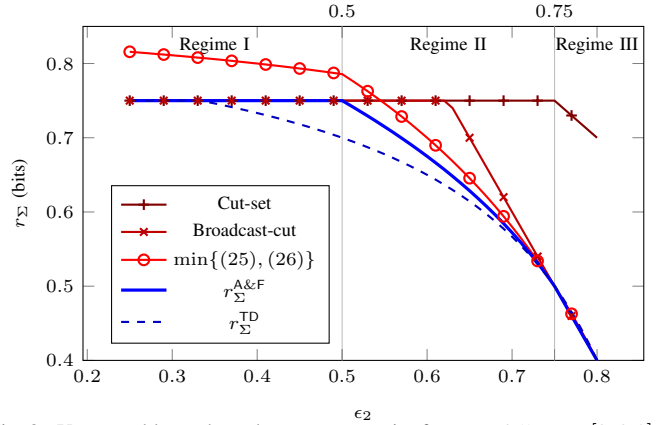

Fig. 8. Upper and lower bounds on sum capacity for $\epsilon_1 = 0.5$, $\epsilon_2 \in [0, 0.8]$, $\epsilon_3 = 0.8$. For this range of values, in Regime I, the cut set bound is tight with our scheme. In Regime II, our new outer bound is significantly closer to the achieved rate. In Regime III, A&F and TD are both capacity achieving.

We will require the following lemma, which is related to Lemma 3 in [6], and proven in Appendix A:

Lemma 3 (Entropy Leakage with no CSIT). *Let Y_A^n and Y_B^n be the channel outputs of a two-user (1-hop) erasure broadcast channel, with independent erasures occurring with probabilities ϵ_a and ϵ_b respectively, with $\epsilon_A \leq \epsilon_B$. If U is a random variable such that the Markov relationships*

$$U \text{ --- } X^n \begin{array}{c} \diagup Y_A^n \\ \diagdown Y_B^n \end{array}$$

hold and $\overline{G}^n$ is knowledge of all erasures in the network, then

$$H(Y_B^n | U, \overline{G}^n) \geq \frac{1 - \epsilon_B}{1 - \epsilon_A} H(Y_A^n | U, \overline{G}^n). \tag{27}$$

Proof of Theorem 2: Assume communication rates of r_1 and r_2 from the source to Destinations 1 and 2 respectively are

Regime	Network Conditions	Reduction Parameters	Sum-Rate ($r_\Sigma^{\mathsf{A\&F}}$)
I	$\epsilon_2 \leq \epsilon_1$	$\nu_u = \frac{1-\epsilon_1}{2}, \nu_c = 0$	$(1-\epsilon_1) + \min\left\{\epsilon_1(1-\epsilon_1), 2(1-\epsilon_3)\left[1 - \frac{1-\epsilon_1}{2(1-\epsilon_2)} + \frac{(1-\epsilon_1)^2}{2}\right]\right\}$
II	$2\epsilon_2 - 1 \leq \epsilon_1 < \epsilon_2,$	$\nu_u = \nu_u^{(II)}, \nu_c = \nu_c^{(II)}$	$(1-\epsilon_1) + \min\left\{\epsilon_1(1-\epsilon_1),\ \nu_c^{(II)} + \epsilon_1\nu_u^{(II)},\ 2(1-\epsilon_3)\left[1 - \frac{1-\epsilon_1}{2(1-\epsilon_2)} + (1-\epsilon_1)\nu_u^{(II)}\right]\right\}$
III	$\epsilon_1 < 2\epsilon_2 - 1$	$\nu_u = 0, \nu_c = 0$	$2 - 2\epsilon_2$

TABLE I

ACHIEVABLE SUM RATES USING ALIGN-AND-FORWARD RELAYING.

achievable. Then, we establish the following two inequalities:

$$nr_1 \overset{\text{(Fano)}}{\leq} I(M_1; Y_{\mathsf{D_1R_1}}^n, Y_{\mathsf{D_1R_2}}^n | \overline{G}^n) + n\varepsilon_n$$
$$= I(M_1; Y_{\mathsf{D_1R_1}}^n | \overline{G}^n) + H(Y_{\mathsf{D_1R_2}}^n | Y_{\mathsf{D_1R_1}}^n, \overline{G}^n)$$
$$- H(Y_{\mathsf{D_1R_2}}^n | Y_{\mathsf{D_1R_1}}^n, \overline{G}^n, M_1) + n\varepsilon_n, \qquad (28)$$

$$nr_2 \overset{\text{(Fano)}}{\leq} I(M_2; Y_{\mathsf{D_2R_1}}^n, Y_{\mathsf{D_2R_2}}^n | \overline{G}^n) + n\varepsilon_n$$
$$\leq I(M_2; Y_{\mathsf{D_2R_1}}^n, Y_{\mathsf{D_2R_2}}^n | \overline{G}^n, M_1) + n\varepsilon_n$$
$$\overset{(a)}{\leq} I(M_2; Y_{\mathsf{D_1R_1}}^n, Y_{\mathsf{D_2R_2}}^n | \overline{G}^n, M_1) + n\varepsilon_n$$
$$= I(M_2; Y_{\mathsf{D_1R_1}}^n | \overline{G}^n, M_1)$$
$$+ I(M_2; Y_{\mathsf{D_2R_2}}^n | Y_{\mathsf{D_1R_1}}^n, \overline{G}^n, M_1) + n\varepsilon_n$$
$$= I(M_2; Y_{\mathsf{D_1R_1}}^n | \overline{G}^n, M_1)$$
$$+ H(Y_{\mathsf{D_2R_2}}^n | Y_{\mathsf{D_1R_1}}^n, \overline{G}^n, M_1) + n\varepsilon_n$$
$$\overset{(b)}{\leq} I(M_2; Y_{\mathsf{D_1R_1}}^n | \overline{G}^n, M_1)$$
$$+ \frac{1-\epsilon_2}{1-\epsilon_3} H(Y_{\mathsf{D_1R_2}}^n | Y_{\mathsf{D_1R_1}}^n, \overline{G}^n, M_1) + n\varepsilon_n. \quad (29)$$

Step (a) is obtained by giving a genie signal, $Z_{\mathsf{D_2R_1}}^n$, to Destination 2 such that the combined tuple $(Y_{\mathsf{D_2R_1}}^n, Z_{\mathsf{D_2R_1}}^n)$ and the random vector $Y_{\mathsf{D_1R_1}}^n$, conditioned on $\overline{G}^n$, are identically distributed random variables. In step (b) we applied Lemma 3 to the term $H(Y_{\mathsf{D_2R_2}}^n | Y_{\mathsf{D_1R_1}}^n, \overline{G}^n, M_1)$. We now drop the term ε_n to avoid confusion with erasure probabilities ϵ_k. Scaling (29) and summing with (28) we find

$$n\left(r_1 + \frac{1-\epsilon_3}{1-\epsilon_2}r_2\right) \leq \frac{\epsilon_3 - \epsilon_2}{1-\epsilon_2} I(M_1; Y_{\mathsf{D_1R_1}}^n | \overline{G}^n)$$
$$+ \frac{1-\epsilon_3}{1-\epsilon_2} \overbrace{\left(I(M_1; Y_{\mathsf{D_1R_1}}^n | \overline{G}^n) + I(M_2; Y_{\mathsf{D_1R_1}}^n | \overline{G}^n, M_1)\right)}^{I(M_1, M_2; Y_{\mathsf{D_1R_1}}^n | \overline{G}^n)}$$
$$+ H(Y_{\mathsf{D_1R_2}}^n | Y_{\mathsf{D_1R_1}}^n, \overline{G}^n)$$
$$\leq \frac{\epsilon_3 - \epsilon_2}{1-\epsilon_2} I(M_1; Y_{\mathsf{D_1R_1}}^n | \overline{G}^n) + \frac{1-\epsilon_3}{1-\epsilon_2} H(Y_{\mathsf{D_1R_1}}^n | \overline{G}^n)$$
$$+ H(Y_{\mathsf{D_1R_2}}^n | Y_{\mathsf{D_1R_1}}^n, \overline{G}^n)$$
$$\leq \frac{\epsilon_3 - \epsilon_2}{1-\epsilon_2} I(M_1; Y_{\mathsf{R_1S}}^n | \overline{G}^n)$$
$$+ n\frac{1-\epsilon_3}{1-\epsilon_2} \min\{(1-\epsilon_2), (1-\epsilon_1)\} + n(1-\epsilon_3). \quad (30)$$

We may also derive a complementary bound for $\frac{1-\epsilon_3}{1-\epsilon_2}r_1 + r_2$:

$$n\left(\frac{1-\epsilon_3}{1-\epsilon_2}r_1 + r_2\right) \leq \frac{\epsilon_3 - \epsilon_2}{1-\epsilon_2} I(M_2; Y_{\mathsf{R_2S}}^n | \overline{G}^n)$$
$$+ n\frac{1-\epsilon_3}{1-\epsilon_2} \min\{(1-\epsilon_2), (1-\epsilon_1)\} + n(1-\epsilon_3). \quad (31)$$

Since $Y_{\mathsf{R_2S}}^n$ and $Y_{\mathsf{R_1S}}^n$ are identically distributed, we also have

$$I(M_1; Y_{\mathsf{R_1S}}^n | \overline{G}^n) + I(M_2; Y_{\mathsf{R_2S}}^n | \overline{G}^n)$$
$$= I(M_1; Y_{\mathsf{R_1S}}^n | \overline{G}^n) + I(M_2; Y_{\mathsf{R_1S}}^n | \overline{G}^n)$$
$$\leq I(M_1; Y_{\mathsf{R_1S}}^n | \overline{G}^n) + I(M_2; Y_{\mathsf{R_1S}}^n | \overline{G}^n, M_1)$$
$$= I(M_1, M_2; Y_{\mathsf{R_1S}}^n | \overline{G}^n) \leq n(1-\epsilon_1). \qquad (32)$$

Scaling and summing (30) and (31) , and noting (32), we find

$$r_1 + r_2 \leq \frac{2(1-\epsilon_3)}{2 - \epsilon_2 - \epsilon_3} \min\{(1-\epsilon_2), (1-\epsilon_1)\}$$
$$+ \frac{\epsilon_3 - \epsilon_2}{2 - \epsilon_2 - \epsilon_3}(1-\epsilon_1) + 2\frac{(1-\epsilon_3)(1-\epsilon_2)}{2 - \epsilon_2 - \epsilon_3}. \quad (33)$$

Evaluating the min operation for the two possibilities and simplifying yields the bounds (25) and (26) respectively.

REFERENCES

[1] A. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 789–804, 2006.

[2] A. Dana, R. Gowaikar, and B. Hassibi, "On the capacity region of the broadcast problems over wireless erasure networks," in *42nd Annual Allerton Conference on Communication, Control, and Computing*, 2004.

[3] C.-C. Wang, "On the capacity of 1-to-k broadcast packet erasure channels with channel output feedback," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 931–956, 2012.

[4] M. Gatzianas, L. Georgiadis, and L. Tassiulas, "Multiuser broadcast erasure channel with feedback: Capacity and algorithms," *Information Theory, IEEE Transactions on*, vol. 59, no. 9, pp. 5779–5804, 2013.

[5] Q. Geng and H. T. Do, "On the capacity region of broadcast packet erasure relay networks with feedback," arXiv preprint arXiv:1312.1727.

[6] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, "Capacity results for binary fading interference channels with delayed csit," submitted to IEEE Transactions on Information Theory, arXiv preprint arXiv:1301.5309.

[7] ——, "Binary fading interference channel with no and delayed local CSIT," submitted to IEEE Transactions on Information Theory.

APPENDIX

A. Proof of Lemma 3

This proof originally appears in [7]. Recall that G_A^n and G_B^n are i.i.d. Bernoulli processes with paramters $1-\epsilon_A$ and $1-\epsilon_B$ respectively, with $\epsilon_B \geq \epsilon_A$. Let $\tilde{G}_A^n$ be another i.i.d. Bernoulli random process with parameter $\frac{1-\epsilon_B}{1-\epsilon_A}$, and define

$$\tilde{Y}_A[t] = \tilde{G}_A[t]Y_A[t] \qquad (34)$$
$$= \tilde{G}_A[t]G_A[t]X[t]. \qquad (35)$$

It is clear that, assuming no CSIT, $\tilde{Y}_A^t$ and Y_B^t are statistically equivalent for $t \in \{1, 2, \ldots, n\}$. Moreover, from

the assumption of Markov relationships, statistical equivalence holds when conditioned on U. Noting this, we see

$$H(Y_B[t]|Y_B^{t-1},U,\overline{G}^n) \stackrel{(a)}{=} H(Y_B[t]|Y_B^{t-1},U,\overline{G}^t)$$

$$= (1-\epsilon_B)H(X[t]|Y_B^{t-1},U,\overline{G}^{t-1},G_B[t]=1)$$
$$+ \epsilon_B H(0|Y_B^{t-1},U,\overline{G}^{t-1},G_B[t]=0)$$
$$= (1-\epsilon_B)H(X[t]|Y_B^{t-1},U,\overline{G}^{t-1},G_B[t]=1)$$
$$\stackrel{(b)}{=} (1-\epsilon_B)H(X[t]|Y_B^{t-1},U,\overline{G}^t)$$
$$\stackrel{(c)}{=} (1-\epsilon_B)H(X[t]|\tilde{Y}_A^{t-1},U,\overline{G}^t)$$
$$\stackrel{(d)}{\geq} (1-\epsilon_B)H(X[t]|Y_A^{t-1},U,\overline{G}^t)$$
$$= \frac{1-\epsilon_B}{1-\epsilon_A}H(Y_A[t]|Y_A^{t-1},U,\overline{G}^t)$$
$$= \frac{1-\epsilon_B}{1-\epsilon_A}H(Y_A[t]|Y_A^{t-1},U,\overline{G}^n), \tag{36}$$

where (a) holds because the channels are i.i.d., (b) holds because the transmitter has no CSIT and therefore, (c) holds because of statistical equivalence, and (d) holds because $H(\tilde{Y}_A^t|Y_A^t,U,\overline{G}^t) = 0$. Using (36), we arrive at the lemma statement:

$$H(Y_B^n|U,\overline{G}^n) = \sum_{t=1}^n H(Y_B[t]|Y_B^{t-1},U,G^n)$$
$$\geq \sum_{t=1}^n \frac{1-\epsilon_B}{1-\epsilon_A}H(Y_A[t]|Y_A^{t-1},U,G^n))$$
$$= \frac{1-\epsilon_B}{1-\epsilon_A}H(Y_A^n|U,\overline{G}^n). \tag{37}$$