Linear Network Codes: A Unified Framework for Source, Channel, and Network Coding *

Michelle Effros Muriel Médard Tracey Ho Siddharth Ray

David Karger Ralf Koetter Babak Hassibi

Submitted to *IEEE Transactions on Information Theory* on May 18, 2003.

Index Terms: COMPRESSION, ERROR CORRECTION, MULTIUSER INFORMATION THEORY, NETWORK CODING, ROUTING.

Abstract

We examine the issue of separation and code design for network data transmission environments. We demonstrate that source-channel (or source-network) separation holds for several canonical network examples when the whole network operates over a common finite field. Our approach uses linear codes. Our simple, unifying framework for these codes not only allows us to re-establish with economy the optimality of linear codes for single transmitter channels and for Slepian-Wolf source coding. It also enables us to establish the optimality of linear codes for multiple access and for erasure

*M. Effros (effros@caltech.edu) and B. Hassibi are with the Department of Electrical Engineering, 136-93, California Institute of Technology, Pasadena, CA 91125. M. Médard (medard@mit.edu), T. Ho (trace@mit.edu), and S. Ray (sray@mit.edu) are with the Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139. D. Karger (karger@lcs.mit.edu) is with the Laboratory for Computer Science (LCS), Massachusetts Institute of Technology, Cambridge, MA 02139. R. Koetter (koetter@uiuc.edu) is with the Coordinated Science Laboratory, University of Illinois at Urbana-Champain, Urbana, IL 61801. This work is supported in part by NSF grant CCR-0220039, University of Illinois subaward #02-194, Hewlett-Packard 008542-008, and Caltech's Lee Center for Advanced Networking.

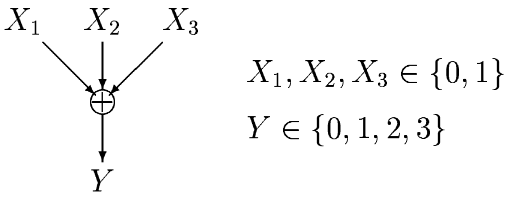


Figure 1: A linear network for which source-channel separation fails [1].

broadcast channels. This robustness of separation we show to be strongly predicated on the fact that noise and inputs are independent. Moreover, we show that source-channel separation holds for these networks. The linearity of both source and network coding blurs the delineation between source and network codes. Finally, we illustrate the fact that design for individual network modules may yield poor results when such modules are concatenated, entailing that end-to-end coding is necessary. Thus, we argue, it is the lack of decomposability into canonical network modules, rather than the lack of separation between source and channel coding, that presents major challenges for network coding.

I Introduction

The failure of source-channel separation in networks is often considered to be an impediment in applying information theoretic tools in network settings. The simple multiple access channel of Figure 1 gives one example of how separation can fail [1]. The receiver's channel output is the integer sum of the binary channel inputs of $m \geq 2$ users, yielding a channel output alphabet of size $m+1$. Since independent, uniformly distributed input signals fail to achieve the maximum mutual information between the transmitted and received signals, direct transmission of dependent source bits over the channel sometimes yields higher achievable transmission rates than Slepian-Wolf source coding followed by multiple access channel coding.

While this simple example may at first appear to irrefutably establish the failure of source-channel separation in networks, its simplicity is misleading. In particular, note that the alphabet size of the output is dependent on the number of transmitters. Thus, the network lacks a consistent digital framework. Replacing integer addition with binary addition to give a channel with input and output alphabets of the same cardinality yields a communication system for which separation holds.

In this paper, we argue that source-channel separation is more robust than counterexamples may suggest. We assert, however, that separate source and channel code design does not necessarily simplify the design of communication systems for digital networks. The operations of compression and channel coding are conceptual tools rather than necessary components. While modularity, such as that afforded by the separation theorem, is desirable in the design of components, the decomposition of a problem into modular tasks may increase complexity when the decomposition imposes unnecessary constraints.

In addition to examining traditional questions of source-channel separation, we also investigate a variety of other separation assumptions implicit in common network design techniques. By assuming independent data bits and lossless links, the network coding literature and other layered approaches to network design endorse a philosophy where source and channel coding are separated from network coding or routing. Through examples, we demonstrate the fragility of this assumed separation. Even in simple digital networks, neither separate source-network coding strategies nor separate channel-network coding techniques guarantee optimal communication performance.

Our network model requires the same finite alphabet at all nodes and additionally allows noise in the form of erasures.[1] Erasures are assumed to be channel-imposed, irreversible, and independent of the channel input so that the erasure symbol cannot be used as an additional symbol for coding. While our examples suggest the robustness of source-channel separation and fragility of source-network and channel-network separation in the resulting systems, we advocate an entirely unified approach, investigating independent, random, linear code design at all nodes of the network. For the examples given, it is not clear, even after the design is completed, what the appropriate decomposition of tasks should be.

We treat two important types of networks in detail: multiple access networks and degraded broadcast networks. For the networks we consider, optimal code construction is particularly simple. We show that random linear codes are sufficient and asymptotically

[1] While we focus primarily on erasure channels, we also briefly consider additive noise channels.

optimal for a wide array of problems. Our approach may be viewed, in the simplest way, as a generalization of information theoretic results known for single-receiver source codes and for single-transmitter, single-receiver channel codes. From the networking perspective, our results bear a different interpretation - compression, channel coding, and routing are not separable functions.

Finally, while the multiple access and broadcast networks considered here are important in their own right, we show that we cannot concatenate them arbitrarily and maintain end-to-end functionality. In effect, there is no separation of large networks into canonical elements. We argue that this lack of separation, rather than the oft-presumed lack of source-channel separation in networks, poses the real challenge in communication-network system design.

II Background

The use of random linear transformations in coding receives considerable attention in the literature. For channel coding, Elias [2] shows that random linear parity check codes, formed by Bernoulli(1/2) choices for the parity check entries in a systematic code's generator matrix, achieve capacity for the binary erasure channel and the binary symmetric channel. Elias also gives a construction for sliding parity check codes requiring fewer random binary digits. MacKay [3] proves that two families of error-correcting codes based on very sparse random parity check matrices – Gallager codes and MacKay-Neal codes (a special case of the former) – when optimally decoded, achieve information rates up to the Shannon limit for channels with symmetric stationery ergodic noise. MacKay also demonstrates empirically, for binary symmetric channels and Gaussian channels, that good decoding performance for these codes can be achieved with a practical sum-product decoding algorithm.

Linear channel coding for network systems has received far less attention. In this work, we consider both multiple access and degraded broadcast channels. In multiple access coding, the model of interest comprises a collection of transmitters sending information to a single receiver. The received signal is the sum of the transmitted signals with the possible inclusion of either erasures or additive noise. While this type of additive interference channel has received considerable attention in the literature (see, for example, [4, 5, 6, 7, 8, 9, 10, 11, 12, 13]) the majority of the work to date considers only the case where the incoming data streams interfere additively in the real field; one notable exception is the work of Poltyrev and Snyder [12], which treats a modulo-2 multiple access channel without noise in the case where a proper subset of the transmitters sends to the decoder at any given instant. We are unaware of prior work on linear coding for multiple access channels.

In broadcast networks, we consider physically and stochastically degraded channels with both additive noise and erasures. While the degraded broadcast channel is well understood, [14, 15], we are likewise unaware of any prior work on linear broadcast channel codes.

On the source coding side, Ancheta [16] presents universally optimal linear codes for lossless coding of binary sources; he also shows that the rate distortion function of a binary, stationary, memoryless source cannot be achieved by any linear transformation over a binary field into a sequence with rate lower than the entropy of the source. The syndrome-source-coding scheme described by Ancheta uses a linear error correcting code for data compression, treating the source sequence as an error pattern whose syndrome forms the compressed data.

In [17], Csiszar generalizes linear source coding techniques to allow linear multiple access source codes that achieve the optimal performance derived by Slepian and Wolf [18]. Csiszar demonstrates the universality of his proposed linear codes[2] and bounds the corresponding error exponents. These results are generalizable to single or multiple Markov sources.

Addressing the problem of practical encoding and decoding for multiple access source codes, [19, 20, 21, 22, 23] introduce the Distributed Source Coding Using Syndromes (DISCUS) framework, initially looking at sources with strongly structured statistical dependencies. Schonberg et al. [24] note that Csiszar's proof can be used to show that application of LDPC codes in the DISCUS framework approaches the Slepian-Wolf bound for general binary sources; they then demonstrate through simulation that belief propagation decoding works well in practice, with a small performance gap due to the finite block length and choice of parity check matrix. Uyematsu proposes a deterministic construction for linear multiple access source codes in [25]; the resulting codes achieve any point in the achievable rate region, with two-step encoding and decoding procedures (similar to concatenated codes

[2] In the given fixed-rate coding regime, a universal code is any code that achieves asymptotically negligible error probability on all sources for which the code's rate falls within the source's achievable rate region.

for channel coding) of complexity polynomial in the block length.

In other related work, multiple access source coding by randomly choosing among general block codes is considered as an exercise in [26]. Loeliger [27] considers averaging for sets of linear codes with basic symmetry properties and gives a general version of the Varshamov-Gilbert bound and a random coding bound that depend only on the size of the set of error patterns; these results extend corresponding prior results for more specific types of error patterns. Among the applications mentioned are burst error correction, and multiple access systems where each user considers the set of possible interference patterns arising from the activity of other users as well as channel noise.

Zhao and Effros introduce broadcast system source codes in [28, 29]. In a broadcast system source code, a single encoder describes multiple sources to be decoded at a collection of receivers. Sources may include both "common information" intended for more than one receiver and "specific information" intended for only one receiver. In the most general case, we allow a distinct source for every non-empty subset of the set of possible receivers. Design algorithms and performance bounds for lossless broadcast system source codes appear in [28, 29]. We know of no prior work on linear broadcast system source codes.

Network coding is a generalization of routing for transmitting independent bits through lossless networks. Work on linear network code design for multi-cast networks has recently become a topic of considerable interest (see, for example, [30, 31, 32]). Koetter and Médard give an algebraic framework in [32]. Reference [33] considers a randomized approach for independent or linearly correlated sources, while [34] and [35] give polynomial-time deterministic and randomized network code constructions for independent sources.

III Preliminaries and Generalizations

Since the focus of our paper is on the relationships between system components and concepts, we give all results in their simplest forms. In particular, we state our results and their corresponding derivations for independent, identically distributed (iid) random processes and focus on binary source and channel alphabets, modified only for the inclusion of the erasure noise model. For simplicity, all code constructions combine random linear encoding with typical set decoding. The definition of the typical set $A_\epsilon^{(n)}$ for a single random sequence $U_1, U_2, \ldots$ drawn iid according to probability mass function (pmf) p is

$$A_\epsilon^{(n)} = \left\{ u^n \in \mathcal{U}^n : -\frac{1}{n} \log p(u^n) < H(U) + \epsilon \right\}.$$

Given source alphabet $\mathcal{U}$, $H(U) = -\sum_{u \in \mathcal{U}} p(u) \log p(u)$ is the entropy of iid random process $U_1, U_2, \ldots$. By the Asymptotic Equipartition Property (AEP),

$$|A_\epsilon^{(n)}| \leq 2^{n(H(U)+\epsilon)}$$

and $\Pr(U^n \in A_\epsilon^{(n)}) \to 1$ as $n \to \infty$. In most cases, we use context to distinguish between typical sets. Thus $U^n \in A_\epsilon^{(n)}$ refers to the typical set for the pmf $p(u)$ of random variable U while $Z^n \in A_\epsilon^{(n)}$ refers to the typical set for the pmf $q(z)$ of random variable Z. Focusing on linear encoding and typical set decoding allows us to include the corresponding proofs and illuminates the relationships between them.

While state and prove our results in their simplest form for readability, we note that all of the results given here generalize widely from the forms that we state explicitly. Some of these generalizations are described below.

- While we focus on the binary alphabet, results generalize to arbitrary finite fields. The requirement that the finite field be the same for all sources, channel codewords, and additive noise processes cannot, however, be relaxed in general. The channel output alphabet is allowed to differ only in the inclusion of erasures. In our model, erasures propagate as erasures when the output of one channel is fed into the input of a subsequent channel.

- We state results for iid source and noise random processes; the results generalize to stationary, ergodic processes for which corresponding typical sets exist.

- We use non-systematic codes in channel coding; the results generalize to systematic codes.

- We use distribution-dependent typical set decoders; many of the results in this paper can be generalized to achieve universal coding performance and improved error exponents using the maximal entropy decoders of Csiszar [17].

- We ignore decoder complexity issues; good (sub-optimal) decoders with lower complexity can be derived for many of the systems described here using sparse matrix techniques like the the low-density parity-check (LDPC) coding techniques developed by Gallager [36], McKay [3], and others.
- We give results for the smallest generalizable instances of each network type (e.g., two-receiver broadcast channels and three-receiver broadcast system source codes); our results generalize to larger systems.

IV Single-Transmitter, Single-Receiver Networks

We begin by examining simple forms of some of the prior results described in Section II. In particular, we give simple new proofs for the linear source and channel coding theorems for single-transmitter, single-receiver networks [2, 16, 17]. These new derivations demonstrate the relationships between these algorithms and random linear network coding techniques. We further provide a linear source coding converse. Finally, we extend the given random design arguments to design linear joint source-channel codes for the single-transmitter, single-receiver network.

Given a single-transmitter, single-receiver network, source coding is equivalent to network coding of compressible source sequences. We say that a network code accomplishes optimal source coding on a noise-free network if that code can be used to transmit any source with entropy lower than the network capacity with asymptotically negligible error probability.

Shannon's achievability result for lossless source coding demonstrates that for $U_1, U_2, \ldots$ drawn iid from a Bernoulli(p) distribution and any $\epsilon > 0$, there exists a fixed-rate-($H(U)+\epsilon$) code for which the probability of decoding error can be made arbitrarily small as the coding dimension n grows without bound. The converse to Shannon's source coding theorem proves that asymptotically negligible error probabilities cannot be achieved with rates lower than $H(U)$. We begin by proving that the expected error probability of a randomly chosen, rate-R, linear source code approaches zero as n grows without bound for any source U with $H(U) < R$. The fixed-rate, linear encoder is independent of the source distribution; we use distribution-dependent typical set decoders for simplicity. We first describe the source encoder and decoder for a fixed linear code and then give the random coding result.

Let a_n be an $\lceil nR \rceil \times n$ matrix with coefficients in the binary field $\mathbb{F}_2$. The encoder for the linear source code based on a_n is

$$\alpha_n(u^n) = a_n \mathbf{u},$$

where $u^n = \mathbf{u}^t \in (\mathbb{F}_2)^n$ is an arbitrary source sequence with blocklength n. The corresponding decoder is

$$\beta_n(v^{\lceil nR \rceil}) = \begin{cases} u^n & \text{if } u^n \in A_\epsilon^{(n)} \text{ and } a_n\mathbf{u} = \mathbf{v} \text{ and } \;\nexists \hat{\mathbf{u}}^n \in A_\epsilon^{(n)} \cap \{\mathbf{u}\}^c \text{ s.t. } a_n\hat{\mathbf{u}} = \mathbf{v} \\ \hat{U}^n & \text{otherwise,} \end{cases}$$

where $v^{\lceil nR \rceil} = \mathbf{v}^t \in (\mathbb{F}_2)^{\lceil nR \rceil}$ and decoding to $\hat{U}^n$ denotes a random decoder output (which yields a decoding error by assumption). The error probability for source code a_n is

$$P_e(a_n) = \Pr(\beta_n(\alpha_n(U^n)) \neq U^n).$$

Theorem 1 *Let $U_1, U_2, \ldots, U_n$ be drawn iid according to distribution $p(u)$. Let $\{A_n\}_{n=1}^\infty$ be a sequence of rate-R linear source codes. Each A_n is an $\lceil nR \rceil \times n$ matrix with coefficients drawn iid Bernoulli$(1/2)$. For any $R > H(U)$, $EP_e(A_n) \to 0$ as $n \to \infty$.*

Proof: We design a sequence $\{A_n\}_{n=1}^\infty$ of codes at random and show that if the rate is chosen appropriately, then the expected error probability $EP_e(A_n)$ of the randomly chosen code decays to zero as n grows without bound. Using the above encoder and decoder definitions and letting $\mathbf{w}^t \in \mathbb{F}_2^n$ be an arbitrary nonzero vector,

$$\begin{aligned} EP_e^{(n)} &= E\Pr(\beta_n(\alpha_n(U^n)) \neq U^n) \\ &= \sum_{u^n \notin A_\epsilon^{(n)}} p(u^n)\Pr(\beta_n(\alpha_n(u^n)) \neq u^n) + \sum_{u^n \in A_\epsilon^{(n)}} p(u^n)\Pr(\beta_n(\alpha_n(u^n)) \neq u^n) \\ &\leq \epsilon_n + \sum_{u^n, \hat{u}^n \in A_\epsilon^{(n)}} p(u^n) 1(\hat{\mathbf{u}} \neq \mathbf{u}) \Pr(A_n\hat{\mathbf{u}} = A_n\mathbf{u}) \qquad (1) \\ &\leq \epsilon_n + \sum_{u^n \in A_\epsilon^{(n)}} p(u^n) 2^{n(H(U)+\epsilon)} \Pr(A_n\mathbf{w} = \mathbf{0}) \qquad (2) \\ &\leq \epsilon_n + 2^{n(H(U)+\epsilon)} 2^{-\lceil nR \rceil} \qquad (3) \end{aligned}$$

for some $\epsilon_n \to 0$. Equation (1) and the bound on the size of the typical set follow from the AEP. The symmetry represented by the introduction of $\mathbf{w}$ in (2) and the bound on the corresponding probability in (3) result from the following argument. Let k be the number of ones in an arbitrary $\mathbf{w} \neq \mathbf{0}$. Then each coefficient of vector $A_n\mathbf{w}$ is the sum of k independent Bernoulli(1/2) random variables. Since summing iid Bernoulli(1/2) random variables yields a Bernoulli(1/2) random variable and the rows of A_n are chosen independently, $A_n\mathbf{w}$ is uniformly distributed over its $2^{\lceil nR \rceil}$ possible outcomes.

By (3), $EP_e^{(n)} \to 0$ as $n \to \infty$ provided that $\lceil nR \rceil > n(H(U) + \epsilon)$. □

Lemma 1 provides a form of converse to Theorem 1. While Theorem 1 shows that linear source codes are asymptotically optimal, Lemma 1 shows that any fixed linear code yields statistically dependent output symbols. An immediate consequence of this observation is that linear source codes cannot achieve the entropy bound for non-uniform sources (since achieving the entropy bound would necessarily yield an incompressible data sequence). This result highlights one difference between the fixed-rate, asymptotically lossless linear codes investigated here and the more typically applied variable-rate, truly lossless source coding schemes like Huffman and arithmetic codes. Variable-rate schemes can achieve lossless performance for any blocklength and precisely achieve the entropy for dyadic distributions. We address the advantages of fixed-rate codes later in this section by showing how fixed-rate, linear source and channel codes combine naturally to give linear joint source-channel codes. The proof of Lemma 1 relies on a recent extension of the Darmois-Skitovich theorem to finite Abelian groups.

Lemma 1 *Given any $n > 1$, let $p_1, \ldots, p_n$ be non-uniform probability mass functions on the mutually independent random variables $U_1, \ldots, U_n$. Defining $\mathbf{V} = (V_1, \ldots, V_k)^t$ and $\mathbf{U} = (U_1, \ldots, U_n)$, let*

$$\mathbf{V} = a\mathbf{U}$$

for an arbitrary $k \times n$ matrix a. If $V_1, V_2, \ldots, V_k$ are mutually independent, then matrix a has at most one non-zero element in each column.

Proof: The proof uses the analogue of the Darmois-Skitovich theorem for discrete periodic Abelian groups by Fel'dman [37]. Let us proceed by contradiction. Suppose that the jth column of a has non-zero elements in positions i and $\hat{i}$ ($\hat{i} \neq i$). Then $V_{\hat{i}}$ and V_i both experience a non-zero contribution from U_j. In this case, the independence of $V_{\hat{i}}$ and V_i requires that p_j be a uniform probability mass function, which gives a contradiction. □

Just as source coding can be viewed as an extension of network coding to applications with statistically dependent input symbols, channel coding can be viewed as an extension of network coding to unreliable channels. Prior network coding results address the issue of robust communication over unreliable channels by considering strategies for working with non-ergodic link failures [32, 33]. We here investigate ergodic failures. A network code designed for a single-transmitter, single-receiver network with ergodic failures is a channel code for the erasure channel. We say that a network code accomplishes optimal channel coding on the given channel if the network code can be used to transmit any source with rate lower than the noisy channel capacity with asymptotically negligible error probability.

To accomplish linear channel coding for the erasure channel, we use an $n \times \lfloor nR \rfloor$ linear generator matrix b_n and a conceptually simple non-linear decoder. The linear channel encoder is defined by

$$\gamma(v^{\lfloor nR \rfloor}) = b_n \mathbf{v}.$$

Let X^n denote the channel input and Y^n denote the corrupted channel output. For any $y^n = \mathbf{y}^t \in \{0, 1, E\}^n$ define the decoder as

$$\delta_n(y^n) = \begin{cases} v^n & \text{if } (b_n\mathbf{v})_i = y_i \text{ for all } i \text{ s.t. } y_i \in \mathbb{F}_2 \\ & \text{ and } \nexists \hat{\mathbf{v}} \neq \mathbf{v} \text{ s.t. } (b_n\hat{\mathbf{v}})_i = y_i \text{ for all } i \text{ s.t. } y_i \in \mathbb{F}_2 \\ \hat{V}^{\lfloor nR \rfloor} & \text{otherwise,} \end{cases}$$

where for any $\mathbf{v} \in \mathbb{F}_2^{\lfloor nR \rfloor}$, $(b_n\mathbf{v})_i$ is the ith component of the vector $b_n\mathbf{v}$. Again, decoding to $\hat{V}^{\lfloor nR \rfloor}$ denotes a random decoder output.

Theorem 2 *Consider an erasure channel with input and output alphabets $\mathbb{F}_2$ and $\{0, 1, E\}$, respectively. The erasure sequence $Z_1, Z_2, \ldots$ is drawn iid according to distribution $q(z)$, where $Z_i = 1$ denotes the erasure event, and $Z_i = 0$ designates a successful transmission. The channel noise is independent of the channel input by assumption. Let $\{B_n\}_{n=1}^{\infty}$ describe a sequence of channel codes. Each B_n is an $n \times \lfloor nR \rfloor$ matrix with elements chosen iid Bernoulli$(1/2)$. If $R < 1 - q(1)$, then the expected error probability $EP_e(B_n) \to 0$ as $n \to \infty$.*

Proof: For the erasure channel, we can immediately decode Z^n from the received string Y^n. For any $z^n \in \mathbb{F}_2^n$, define $\mathcal{E}(z^n) = \{\mathbf{e} \in \mathbb{F}_2^n : e_i = z_i \quad \forall i \text{ s.t. } z_i = 0\}$. A decoding error occurs if there exists a $\hat{\mathbf{v}} \neq \mathbf{V}$ for which $B_n\mathbf{V} - B_n\hat{\mathbf{v}} = B_n(\mathbf{V} - \hat{\mathbf{v}}) \in \mathcal{E}(Z^n)$, since any such $\hat{\mathbf{v}}$ would be mapped to the same channel output by Z^n. For any z^n with $\sum_{i=1}^n z_i = k$, $|\mathcal{E}(z^n)| = 2^k$. Using the definition of the typical set, $z^n \in A_\epsilon^{(n)}$ implies that $\sum_{i=1}^n z_i \leq n(q(1) + \epsilon')$, where $\epsilon' = \epsilon / \log(q(1)/q(0))$. Thus for any fixed $z^n \in A_\epsilon^{(n)}$ and $\mathbf{w}^t \in \mathbb{F}_2^{\lfloor nR \rfloor}$, $\Pr(B_n\mathbf{w} \in \mathcal{E}(z^n)) \leq 2^{-n}2^{n(q(1)+\epsilon')}$ (since $B_n\mathbf{w}$ is uniformly distributed by the argument in the proof of Theorem 1), giving

$$\begin{aligned}
&EP_e^{(n)}(B_n) \\
&= E\Pr(\text{Error} \wedge Z^n \notin A_\epsilon^{(n)}(q)) + E\Pr(\text{Error} \wedge Z^n \in A_\epsilon^{(n)}(q)) \\
&\leq \epsilon_n + \sum_{v^{\lfloor nR \rfloor}, \hat{v}^{\lfloor nR \rfloor} \in \mathbb{F}_2^{\lfloor nR \rfloor}} \sum_{z^n \in A_\epsilon^{(n)}(q)} p(v^{\lfloor nR \rfloor}) q(z^n) 1(\hat{\mathbf{v}} \neq \mathbf{v}) \Pr(B_n(\mathbf{v} - \hat{\mathbf{v}}) \in \mathcal{E}(z^n)) \\
&\leq \epsilon_n + \sum_{v^{\lfloor nR \rfloor} \in \mathbb{F}_2^{\lfloor nR \rfloor}} \sum_{z^n \in A_\epsilon^{(n)}(q)} p(v^{\lfloor nR \rfloor}) q(z^n) 2^{\lfloor nR \rfloor} 2^{-n} 2^{n(q(1)+\epsilon')} \\
&\leq \epsilon_n + 2^{-n(1-q(1)-\epsilon')+\lfloor nR \rfloor}
\end{aligned}$$

for some $\epsilon_n \to 0$. Here $A_\epsilon^{(n)}(p)$ is the typical set for the source distribution and $A_\epsilon^{(n)}(q)$ is the typical set for the noise. The expected error probability decays to zero as n grows without bound provided that $R < 1 - q(1) - \epsilon'$. □

By Shannon's separation theorem, we can achieve optimal communication over the given erasure channel by concatenating optimal source and channel codes. Concatenating the optimal linear source and channel codes of Theorems 1 and 2 yields an optimal linear source-channel code. Given source code a_n and channel code b_n, the joint source-channel encoder multiplies the source input by a single $n \times n$ matrix $c_n = b_n a_n$ and transmits the output across the channel. The corresponding decoder is $\beta_n(\delta_n(\cdot))$.

As an alternative to the above approach, where we design separate random linear source and channel codes and concatenate them together, we can design a joint source-channel code at random and decode in a single typical set decoding argument. While we stick with the traditional name of joint source-channel coding, we note that the code does not perform the separate functions of source and channel coding jointly. Instead, the code maps source sequences to channel inputs in a manner that allows robust communication without any explicit or implicit compression or addition of channel coding redundancy.

The joint source-channel code's encoder is defined by

$$\zeta(u^n) = c_n\mathbf{u}.$$

Denote the random channel input and output by X^n and Y^n, respectively. For any $y^n = \mathbf{y}^t \in \{0, 1, E\}^n$ the decoder is defined by

$$\eta_n(y^n) = \begin{cases} u^n & \text{if } (c_n\mathbf{u})_i = y_i \text{ for all } i \text{ s.t. } \; y_i \in \mathbb{F}_2 \\ & \quad \text{and } \; \nexists \hat{\mathbf{u}} \neq \mathbf{u} \text{ s.t. } (c_n\hat{\mathbf{u}})_i = y_i \text{ for all } i \text{ s.t. } \; y_i \in \mathbb{F}_2 \\ \hat{U}^n & \text{otherwise.} \end{cases}$$

Here, $(c_n\mathbf{u})_i$ denotes the ith component of vector $c_n\mathbf{u}$. The error probability for code c_n is

$$P_e(c_n) = \Pr(\eta_n(\zeta_n(Y^n)) \neq U^n),$$

where U^n and Y^n are the random source vector and channel output, respectively. Theorem 5 demonstrates that the expected error probability for a randomly chosen linear code C_n decays to zero as n grows without bound.

Theorem 3 *Consider the random source $U_1, U_2, \ldots$ drawn iid according to distribution $p(u)$, and let $Z_1, Z_2, \ldots$ be the channel's random erasures, where $Z_1, Z_2, \ldots$ are drawn iid according to distribution $q(z)$ and are independent of the source. (Again $Z_i = 1$ denotes an erasure event.) Assume that the source and channel input alphabets are equal to the binary field $\mathbb{F}_2$. Let $\{C_n\}_{n=1}^{\infty}$ describe a sequence of joint source-channel codes. Each C_n is an $n \times n$ matrix with elements chosen iid Bernoulli$(1/2)$. If $H(U) < 1 - q(1)$, then the expected error probability $EP_e(C_n) \to 0$ as $n \to \infty$.*

Proof: Again, we can immediately decode Z^n from the received string Y^n, and a decoding error occurs if there exists a $\hat{\mathbf{u}} \neq \mathbf{U}$ for which $C_n(\mathbf{U} - \hat{\mathbf{u}}) \in \mathcal{E}(Z^n)$. Thus

$$\begin{aligned} EP_e^{(n)}(C_n) &= E\Pr\left(\text{Error} \wedge \left(U^n \notin A_\epsilon^{(n)}(p) \vee Z^n \notin A_\epsilon^{(n)}(q)\right)\right) \\ &\quad + E\Pr\left(\text{Error} \wedge U^n \in A_\epsilon^{(n)}(p) \wedge Z^n \in A_\epsilon^{(n)}(q)\right) \\ &\leq 2\epsilon_n + \sum_{u^n, \hat{u}^n \in A_\epsilon^{(n)}(p)} \sum_{z^n \in A_\epsilon^{(n)}(q)} p(u^n)q(z^n)1(\hat{\mathbf{u}} \neq \mathbf{u}) \Pr(C_n(\mathbf{u} - \hat{\mathbf{u}}) \in \mathcal{E}(z^n)) \\ &\leq 2\epsilon_n + \sum_{u^n \in A_\epsilon^{(n)}(p)} \sum_{z^n \in A_\epsilon^{(n)}(q)} p(u^n)q(z^n)2^{n(H(U)+\epsilon)}2^{-n}2^{n(q(1)+\epsilon')} \\ &\leq 2\epsilon_n + 2^{-n(1-q(1)-\epsilon'-H(U)-\epsilon)} \end{aligned}$$

for some $\epsilon_n \to 0$. Thus the expected error probability decays to zero as n grows without bound provided that $H(U) < 1 - q(1) - \epsilon - \epsilon'$. □

While we focus primarily on the erasure channel model, we note that both the channel coding theorem and the joint source-channel coding extend easily to additive noise models. This model may be viewed either as true noise or as the signal of another user that has been combined with the desired signal at some node of a network code. The second interpretation is only useful when the interfering signal is not iid uniform; we treat interference channels in detail in Section V.

We begin with the additive noise channel's channel coding theorem. Let a_n be an $\lceil n(1-R)\rceil \times n$ matrix with coefficients in $\mathbb{F}_2$. For channel coding, a_n plays the traditional role of the parity check matrix. Following Csiszar [17], however, we interpret a_n as a source code on the noise. For any matrix a_n, we can design an $n \times \lfloor nR \rfloor$ matrix b_n such that b_n has full rank and $a_n b_n = \mathbf{0}$. Matrix b_n plays the role of the generator matrix for the desired channel code. We design b_n to have full rank so that each length-$\lfloor nR \rfloor$ input message maps to a distinct channel codeword. We force $a_n b_n = \mathbf{0}$ so that each codeword is in the null space of a_n, making possible separation of the encoded message from the additive noise.

More precisely, the channel encoder is defined by

$$\gamma(v^{n-k}) = b_n \mathbf{v}.$$

The channel output for a random channel input $b_n \mathbf{V}$ is

$$\mathbf{Y} = b_n \mathbf{V} + \mathbf{Z}.$$

In decoding the channel output, the receiver first multiplies $\mathbf{Y}$ by a_n to give

$$a_n \mathbf{Y} = a_n(b_n \mathbf{V} + \mathbf{Z}) = a_n \mathbf{Z}.$$

The result of this multiplication is a source coded description of the error signal $\mathbf{Z}$. Thus the decoding procedure involves applying source decoder β_n to $a_n \mathbf{Y}$. The error is decoded correctly with high probability. The receiver then subtracts the error estimate from the received $\mathbf{Y}$ to yield, with high-probability, $b_n \mathbf{V}$. Since b_n has full rank, the receiver can recover $\mathbf{V}$ perfectly from $b_n \mathbf{V}$. Thus the channel code's error probability equals the error probability for the corresponding source code on the error signal Z^n. Given this insight, the channel coding theorem is an immediate extension of the source coding theorem.

Theorem 4 *Consider an additive noise channel with input, output, and noise alphabets all equal to the binary field $\mathbb{F}_2$. Let noise $Z_1, Z_2, \ldots$ be drawn iid according to distribution $q(z)$. The channel noise is independent of the channel input. Let $\{(B_n, A_n)\}_{n=1}^{\infty}$ describe a sequence of channel codes. Each A_n is $\lceil n(1-R)\rceil$ matrix with elements chosen iid Bernoulli$(1/2)$. Each B_n is designed to match the corresponding A_n as described above. If $R < 1 - H(Z)$, then the expected error probability $EP_e(B_n, A_n) \to 0$ as $n \to \infty$.*

Proof: As in the proof of Theorem 1, we choose a sequence $\{A_n\}_{n=1}^{\infty}$ of matrices at random. This is our source code for the noise. For each A_n, we design an $n \times (n-k)$ matrix B_n such that B_n has full rank and $A_n B_n = \mathbf{0}_{k\times(n-k)}$. By the argument given above, the error probability for the given channel code equals the error probability for the corresponding source code on the error signal Z^n. By Theorem 1, the expected value of this error probability goes to zero as n grows without bound for all $\lceil n(1-R)\rceil > nH(Z)$, giving an asymptotically negligible error probability for any $R < 1 - H(Z)$. □

Finally, we consider a linear joint source-channel code for the additive noise channel. Again, given $n \times n$ matrix c_n, we define the encoder as

$$\zeta(u^n) = c_n \mathbf{u}.$$

Given random channel input $c_n\mathbf{U}$, the channel output is

$$\mathbf{Y} = c_n \mathbf{U} + \mathbf{Z}.$$

The decoder is

$$\eta_n(y^n) = \begin{cases} u^n & \text{if } u^n \in A_\epsilon^{(n)}(p) \text{ and } \exists z^n \in A_\epsilon^{(n)}(q) \text{ s.t. } c_n\mathbf{u} + \mathbf{z} = \mathbf{y} \\ & \quad \text{and } \nexists(\hat{\mathbf{u}}^n, \hat{\mathbf{z}}^n) \in (A_\epsilon^{(n)}(p) \cap \{\mathbf{u}\}^c) \times A_\epsilon^{(n)}(q) \text{ s.t. } c_n\hat{\mathbf{u}} + \hat{\mathbf{z}} = \mathbf{y} \\ \hat{U}^n & \text{otherwise.} \end{cases}$$

The error probability for code c_n is

$$P_e(c_n) = \Pr(\eta_n(\zeta_n(U^n) + Z^n) \neq U^n).$$

Theorem 5 demonstrates that the expected error probability for a randomly chosen linear code C_n decays to zero as n grows without bound.

Theorem 5 *Consider the random source $U_1, U_2, \ldots$ drawn iid according to distribution $p(u)$, and let $Z_1, Z_2, \ldots$ be the channel's random additive noise, where $Z_1, Z_2, \ldots$ are drawn iid according to distribution $q(z)$ and are independent of the source. Assume that the source, channel input, channel output, and noise alphabets are all equal to the binary field $\mathbb{F}_2$. Let $\{C_n\}_{n=1}^{\infty}$ describe a sequence of joint source-channel codes. Each C_n is an $n \times n$ matrix with elements chosen iid Bernoulli$(1/2)$. If $H(U) < 1 - H(Z)$, then the expected error probability $EP_e(C_n) \to 0$ as $n \to \infty$.*

Proof: An error occurs if two source sequences are mapped to the same channel input vector or if there exist distinct noise vectors that map distinct channel input vectors to the same channel output. In the first case, $C_n\mathbf{U} = C_n\hat{\mathbf{u}}$ for some $\hat{\mathbf{u}} \neq \mathbf{U}$, and in the second case, $C_n\mathbf{U} + \mathbf{Z} = C_n\hat{\mathbf{u}} + \hat{\mathbf{z}}$ for some $\hat{\mathbf{u}} \neq \mathbf{U}$ and $\hat{\mathbf{z}} \neq \mathbf{Z}$. Restricting our attention to typical source and noise vectors, an error occurs if there exists a $\hat{\mathbf{u}} \in A_\epsilon^{(n)}(p)$ such that $\hat{\mathbf{u}} \neq \mathbf{U}$ and $C_n(\hat{\mathbf{u}} - \mathbf{U}) \in \{\mathbf{0}\} \cup \{\hat{\mathbf{z}} - \mathbf{Z} : \hat{\mathbf{z}} \in A_\epsilon^{(n)}(q)\}$. For any fixed $\mathbf{u} - \hat{\mathbf{u}} \neq \mathbf{0}$ and randomly chosen C_n, the coefficients of vector $C_n(\mathbf{u} - \hat{\mathbf{u}})$ are sums of fixed numbers of iid Bernoulli$(1/2)$ values. Thus $\Pr(C_n(\hat{\mathbf{u}} - \mathbf{u}) = \mathbf{w}) = 2^{-n}$ for all $\mathbf{w} \in \mathbb{F}_2^n$, and

$$\begin{aligned}
& EP_e^{(n)}(C_n) \\
= \;& E\Pr\left(\text{Error } \wedge \left(U^n \notin A_\epsilon^{(n)}(p) \vee Z^n \notin A_\epsilon^{(n)}(q)\right)\right) \\
& + E\Pr\left(\text{Error } \wedge U^n \in A_\epsilon^{(n)}(p) \wedge Z^n \in A_\epsilon^{(n)}(q)\right) \\
\leq \;& 2\epsilon_n + \sum_{(u^n, z^n), (\hat{u}^n, \hat{z}^n) \in A_\epsilon^{(n)}(p) \times A_\epsilon^{(n)}(q)} p(u^n) q(z^n) 1(\hat{\mathbf{u}} \neq \mathbf{u}) \Pr\left(C_n(\mathbf{u} - \hat{\mathbf{u}}) = \hat{\mathbf{z}} - \mathbf{z}\right) \\
\leq \;& 2\epsilon_n + \sum_{(u^n, z^n) \in A_\epsilon^{(n)}(p) \times A_\epsilon^{(n)}(q)} p(u^n) q(z^n) 2^{n(H(U)+\epsilon)} 2^{n(H(q)+\epsilon)} 2^{-n} \\
\leq \;& 2\epsilon_n + 2^{-n(1-H(q)-H(U)-2\epsilon)}
\end{aligned}$$

for some $\epsilon_n \to 0$. The error probability goes to zero provided that $H(U) < 1 - H(q) - 2\epsilon$. □

V Multiple Access Systems

The techniques applied in the previous section for single-transmitter, single-receiver systems can also be applied to the design of linear source and channel codes for networks. We begin with a simple re-derivation of the linear multiple access source codes first studied by Csiszar [17]. We then consider linear coding for multiple access channel codes.

Given $\lceil nR_1 \rceil \times n$ matrix $a_{1,n}$ and $\lceil nR_2 \rceil \times n$ matrix $a_{2,n}$, we associate with $(a_{1,n}, a_{2,n})$ a blocklength-n, two-transmitter, linear multiple access source code as follows. For any $u_1^n = \mathbf{u}_1^t \in (\mathbb{F}_2)^n$ and $u_2^n = \mathbf{u}_2^t \in (\mathbb{F}_2)^n$, encoders 1 and 2 are defined by

$$\begin{aligned} \alpha_{1,n}(u_1^n) &= a_{1,n}\mathbf{u}_1 \\ \alpha_{2,n}(u_2^n) &= a_{2,n}\mathbf{u}_2 \end{aligned}$$

For any $v_1^{\lceil nR_1 \rceil} = \mathbf{v}_1^t \in (\mathbb{F}_2)^{\lceil nR_1 \rceil}$ and $v_2^{\lceil nR_2 \rceil} = \mathbf{v}_2^t \in (\mathbb{F}_2)^{\lceil nR_2 \rceil}$, the decoder is defined by

$$\beta_n(v_1^{\lceil nR_1 \rceil}, v_2^{\lceil nR_2 \rceil}) = \begin{cases} (u_1^n, u_2^n) & \text{if } (u_1^n, u_2^n) \in A_\epsilon^{(n)} \text{ and } (a_{1,n}\mathbf{u}_1, a_{2,n}\mathbf{u}_2) = (\mathbf{v}_1, \mathbf{v}_2) \text{ and} \\ & \nexists(\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2) \in A_\epsilon^{(n)} \cap \{(\mathbf{u}_1, \mathbf{u}_2)\}^c \text{ s.t.} \\ & (a_{1,n}\hat{\mathbf{u}}_1, a_{2,n}\hat{\mathbf{u}}_2) = (\mathbf{v}_1, \mathbf{v}_2) \\ (\hat{U}_1^n, \hat{U}_2^n) & \text{otherwise.} \end{cases}$$

Again, decoding to $(\hat{U}_1^n, \hat{U}_2^n)$ denotes an error event.

Theorem 6 *Consider source sequence $(U_{1,1}, U_{2,1}), (U_{1,2}, U_{2,2}), \ldots$ drawn iid according to distribution $p(u_1, u_2)$ on $(\mathbb{F}_2)^2$. Let $\{(A_{1,n}, A_{2,n})\}_{n=1}^\infty$ be a sequence of rate-(R_1, R_2) linear multiple-access source codes with coefficients chosen iid Bernoulli$(1/2)$. Then for any rates*

$$\begin{aligned} R_1 &> H(U_1|U_2) \\ R_2 &> H(U_2|U_1) \\ R_1 + R_2 &> H(U_1, U_2), \end{aligned}$$

$\{(A_{1,n}, A_{2,n})\}_{n=1}^\infty$ achieves expected error probability $EP_e(A_{1,n}, A_{2,n}) \to 0$ as $n \to \infty$.

Proof: An error occurs if either or both of the source sequences is decoded in error. Thus, following an argument very similar to those seen previously,

$$EP_e(A_{1,n}, A_{2,n})$$

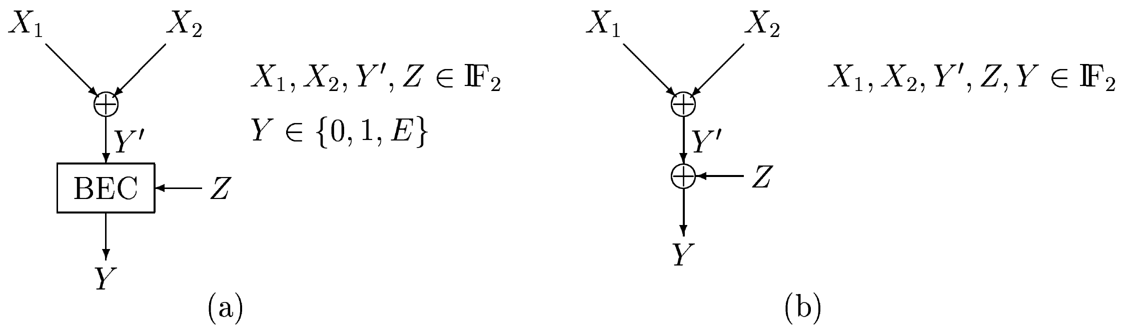

X_1
X_2
Y'
BEC
Z
Y
$X_1, X_2, Y', Z \in \mathbb{F}_2$
$Y \in \{0, 1, E\}$
(a)
X_1
X_2
Y'
Z
Y
$X_1, X_2, Y', Z, Y \in \mathbb{F}_2$
(b)

Figure 2: Binary additive multiple access channels with (a) erasures and (b) additive noise. In both cases, $Z_1, Z_2, \ldots$ are iid and independent of the channel inputs.

$$\begin{aligned}
&= E\Pr(\beta_n(\alpha_{1,n}(U_1^n), \alpha_{2,n}(U_2^n))) \neq (U_1^n, U_2^n) \wedge (U_1^n, U_2^n) \notin A_\epsilon^{(n)}) \\
&\quad + E\Pr(\beta_n(\alpha_{1,n}(U_1^n), \alpha_{2,n}(U_2^n))) \neq (U_1^n, U_2^n) \wedge (U_1^n, U_2^n) \in A_\epsilon^{(n)}) \\
&\leq \epsilon_n + \sum_{(u_1^n, u_2^n) \in A_\epsilon^{(n)}} p(u_1^n, u_2^n) \sum_{\hat{u}_2^n : (u_1^n, u_2^n) \in A_\epsilon^{(n)}} 1(\hat{u}_2^n \neq u_2^n) \Pr(A_{2,n}(\mathbf{u}_2 - \hat{\mathbf{u}}_2) = \mathbf{0}) \\
&\quad + \sum_{(u_1^n, u_2^n) \in A_\epsilon^{(n)}} p(u_1^n, u_2^n) \sum_{\hat{u}_1^n : (\hat{u}_1^n, u_2^n) \in A_\epsilon^{(n)}} 1(\hat{u}_1^n \neq u_1^n) \Pr(A_{1,n}(\mathbf{u}_1 - \hat{\mathbf{u}}_1) = \mathbf{0}) \\
&\quad + \sum_{(u_1^n, u_2^n), (\hat{u}_1^n, \hat{u}_2^n) \in A_\epsilon^{(n)}} p(u_1^n, u_2^n) 1(\hat{u}_1^n \neq u_1^n) 1(\hat{u}_2^n \neq u_2^n) \\
&\qquad \cdot \Pr((A_{1,n}(\mathbf{u}_1 - \hat{\mathbf{u}}_1), A_{2,n}(\mathbf{u}_2 - \hat{\mathbf{u}}_2)) = (\mathbf{0}, \mathbf{0})) \\
&\leq \epsilon_n + 2^{n(H(U_1|U_2)+2\epsilon)} \Pr(A_{1,n}\mathbf{w} = \mathbf{0}) + 2^{n(H(U_2|U_1)+2\epsilon)} \Pr(A_{2,n}\mathbf{w} = \mathbf{0}) \\
&\quad + 2^{n(H(U_1,U_2)+\epsilon)} \Pr(A_{1,n}\mathbf{w}_1 = \mathbf{0} \ \wedge \ A_{2,n}\mathbf{w}_2 = \mathbf{0}) \\
&= \epsilon_n + 2^{-(\lceil nR_1 \rceil - n(H(U_1|U_2)+2\epsilon))} + 2^{-(\lceil nR_2 \rceil - n(H(U_2|U_1)+2\epsilon))} \\
&\quad + 2^{-(\lceil nR_1 \rceil + \lceil nR_2 \rceil - n(H(U_1,U_2)+\epsilon))}
\end{aligned}$$

for arbitrary, non-zero $\mathbf{w}^t, \mathbf{w}_1^t, \mathbf{w}_2^t \in \mathbb{F}_2^n$ and some $\epsilon_n \to 0$. Thus for all $(\lceil nR_1 \rceil, \lceil nR_2 \rceil)$ satisfying $\lceil nR_1 \rceil > n(H(U_1|U_2) + 2\epsilon)$, $\lceil nR_2 \rceil > n(H(U_2|U_1) + 2\epsilon)$, and $\lceil nR_1 \rceil + \lceil nR_2 \rceil > n(H(U_1, U_2) + \epsilon)$, $EP_e(A_{1,n}, A_{2,n}) \to 0$ as n grows without bound. □

Application of the above linear channel coding techniques to achieve linear multiple access channel codes is more straightforward than the corresponding source coding result. In particular, we consider the two additive multiple access channels shown in Figure 2. The first is the additive multiple access channel with erasures, and the second is the additive multiple access channel with additive noise. The additive channel with interference only (no channel noise) can be viewed as a special case of either of the noisy models where errors or erasures occur with probability zero. Let X_1^n and X_2^n denote the random channel inputs, and use Y^n to denote the corresponding random channel output. Then Y^n equals $X_1^n + X_2^n$ corrupted by erasures in the erasure channel model, and $Y^n = X_1^n + X_2^n + Z^n$ for iid additive binary noise Z^n in the additive noise channel model. Both examples use addition over the binary field. All noise is independent of the channel input.

We begin by deriving the multiple access capacities of both the additive multiple access channel with erasures and the additive multiple access channel with additive noise.

Lemma 2 *The multiple access capacities of both the additive multiple access channel with erasures and the additive multiple access channel with additive noise equal the rate region achieved by time-sharing between the points $(C, 0)$ and $(0, C)$, respectively, where $C = 1 - q(1)$ for the erasure model and $C = 1 - H(Z)$ for the additive noise model.*

Proof: The cooperative capacity for each multiple access channel is equal to the capacity of the corresponding single-transmitter, single-receiver channel. Since the multiple access capacity without cooperation cannot exceed the cooperative capacity and the above time-sharing solution achieves the cooperative capacity, we have the desired result. □

Since time-sharing between two linear codes can itself be described as a linear code, the time-sharing solution demonstrates not only that the end points are achievable by linear codes but also that all points in the set of achievable rates are achievable by linear multiple access channel codes. The following argument demonstrates the construction of linear multiple access channel codes from linear channel codes for single-transmitter, single-receiver networks.

Matrix pair $(b_{n,1}, b_{n,2})$ denotes a linear multiple access channel code with encoders

$$\begin{aligned}\gamma_1(v_1^{\lfloor nR_1 \rfloor}) &= b_{1,n}\mathbf{v}_1 \\ \gamma_2(v_2^{\lfloor nR_2 \rfloor}) &= b_{2,n}\mathbf{v}_2.\end{aligned}$$

We build matrices $(b_{n,1}, b_{n,2})$ from the linear code for the corresponding single-transmitter, single-receiver channel. Let $\{b_n\}_{n=1}^{\infty}$ be a sequence of rate-R single-transmitter, single-receiver channel codes for the given channel model, then matrix pair $(b_{n,1}^0, b_{n,2}^0) = (b_n, \mathbf{0}_{nR\times n})$

describes a multiple access channel code that achieves rate pair $(R, 0)$. Similarly, matrix pair $(b^1_{n,1}, b^1_{n,2}) = (\mathbf{0}_{nR\times n}, b_n)$ describes a multiple access channel code achieving rate pair $(0, R)$. The multiple access channel code achieving the $(\lambda, 1-\lambda)$ time-sharing solution between $(R, 0)$ and $(0, R)$ is a linear code with

$$[b^\lambda_{1,n}, b^\lambda_{2,n}] = \left(\begin{bmatrix} b_{\lambda n} & \mathbf{0}_{\lambda n\times(1-\lambda)nR} \\ \mathbf{0}_{(1-\lambda)n\times\lambda nR} & \mathbf{0}_{(1-\lambda)n\times(1-\lambda)nR} \end{bmatrix}, \begin{bmatrix} \mathbf{0}_{\lambda n\times\lambda nR} & \mathbf{0}_{\lambda n\times(1-\lambda)nR} \\ \mathbf{0}_{(1-\lambda)n\times\lambda nR} & b_{(1-\lambda)n} \end{bmatrix} \right).$$

We decoding the first λn channel outputs with the decoder for $b_{\lambda n}$ and the remaining outputs with the decoder for $\beta_{(1-\lambda)n}$. The resulting codes lead immediately to Theorems 7 and 8.

Theorem 7 *Consider a multiple access channel with input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{F}_2$ and output alphabet $\mathcal{Y} = \{0, 1, E\}$. If the channel inputs at time i are $X_{1,i}$ and $X_{2,i}$, then the channel output at time i is the binary sum $X_{1,i} + X_{2,i}$ with probability $q(0)$ and E with probability $q(1)$. Erasures are iid and independent of the channel inputs. Let $\{(B_{1,n}, B_{2,n})\}_{n=1}^\infty$ describe a sequence of rate-(R_1, R_2) multiple access channel codes. Matrices $B_{1,n}$ and $B_{2,n}$ take the forms*

$$B_{1,n} = \begin{bmatrix} B_{\lambda n} & \mathbf{0}_{\lambda n\times(1-\lambda)nR} \\ \mathbf{0}_{(1-\lambda)n\times\lambda nR} & \mathbf{0}_{(1-\lambda)n\times(1-\lambda)nR} \end{bmatrix} \text{ and } B_{2,n} = \begin{bmatrix} \mathbf{0}_{\lambda n\times\lambda nR} & \mathbf{0}_{\lambda n\times(1-\lambda)nR} \\ \mathbf{0}_{(1-\lambda)n\times\lambda nR} & B_{(1-\lambda)n} \end{bmatrix},$$

where $B_{\lambda n}$ and $B_{(1-\lambda)n}$ are $\lfloor \lambda nR\rfloor \times \lfloor \lambda n\rfloor$ and $(\lfloor nR\rfloor - \lfloor \lambda nR\rfloor) \times (n - \lfloor \lambda n\rfloor)$ matrices, respectively, with coefficients chosen iid Bernoulli$(1/2)$. For any $\lambda \in [0, 1]$ and $R < 1 - q(1)$, the given sequence of linear multiple access channel codes gives expected error probability $EP_e(B_{1,n}, B_{2,n}) \to 0$ as $n \to \infty$. Thus all rates (R_1, R_2) with $R_1 + R_2 < 1 - q(1)$ are achievable.

Theorem 8 *Consider a multiple access channel with input-independent, additive noise. Suppose that the input alphabets, output alphabet, and noise alphabet are all equal to the binary field $\mathbb{F}_2$. Let noise $Z_1, Z_2, \ldots$ be drawn iid according to distribution $q(z)$. If the channel inputs at time i are $X_{1,i}$ and $X_{2,i}$, then the channel output at time i is $Y_i = X_{1,i}+X_{2,i}+Z_i$. Let $\{(B_{1,n}, B_{2,n}, A_n)\}_{n=1}^\infty$ describe a sequence of rate-(R_1, R_2) multiple access channel codes. Matrix A_n takes form*

$$A_n = \begin{bmatrix} A_{\lambda n} & \mathbf{0} \\ \mathbf{0} & A_{(1-\lambda)n} \end{bmatrix},$$

where $A_{\lambda n}$ and $A_{(1-\lambda)n}$ are $\lceil (1-R)\lambda n \rceil \times \lambda n$ and $\lceil (1-R)(1-\lambda)n \rceil \times (1-\lambda)n$ matrices, respectively, with entries chosen iid Bernoulli$(1/2)$. Matrices $B_{1,n}$ and $B_{2,n}$ take the forms

$$B_{1,n} = \begin{bmatrix} B_{\lambda n} & \mathbf{0}_{\lambda n \times (1-\lambda)nR} \\ \mathbf{0}_{(1-\lambda)n \times \lambda nR} & \mathbf{0}_{(1-\lambda)n \times (1-\lambda)nR} \end{bmatrix} \text{ and } B_{2,n} = \begin{bmatrix} \mathbf{0}_{\lambda n \times \lambda nR} & \mathbf{0}_{\lambda n \times (1-\lambda)nR} \\ \mathbf{0}_{(1-\lambda)n \times \lambda nR} & B_{(1-\lambda)n} \end{bmatrix},$$

where $B_{\lambda n}$ and $B_{(1-\lambda)n}$ are the generator matrices corresponding to random parity check matrices $A_{\lambda n}$ and $A_{(1-\lambda)n}$, respectively. For any $\lambda \in [0,1]$ and $R < 1 - H(Z)$, the given sequence of linear multiple access channel codes gives expected error probability $EP_e(B_{1,n}, B_{2,n}, A_n) \to 0$ as $n \to \infty$. Thus all rates (R_1, R_2) with $R_1 + R_2 < 1 - H(Z)$ are achievable.

While the proofs of Theorems 7 and 8 take slightly different approaches, this difference is not essential. The proof methodology from Theorem 7, which uses direct typical set decoding rather than building a parity-check matrix, can be adapted to the additive noise multiple access channel.

Given the above source and channel coding theorems, we next tackle the issue of source-channel separation for our multiple access channels.

Theorem 9 *Consider a multiple access channel with input alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{F}_2$ and output alphabet $\mathcal{Y} = \{0, 1, E\}$. If the channel inputs at time i are $X_{1,i}$ and $X_{2,i}$, then the channel output at time i is the binary sum $X_{1,i} + X_{2,i}$ with probability $q(0)$ and E with probability $q(1)$; the erasure events are iid. If source pair $(U_{1,1}, U_{2,1}), (U_{1,2}, U_{2,2}), \ldots$ is drawn iid according to distribution $p(u_1, u_2)$ with $H(U_1, U_2) < 1 - q(1)$, then there exists a sequence of joint source-channel codes with probability of error $P_e^{(n)} \to 0$. Conversely, if $H(U_1, U_2) > 1 - q(1)$, then the probability of error for any communication system is bounded away from zero. Thus source-channel separation holds for the multiple access erasure channel.*

Proof: By Theorem 6, the Slepian-Wolf region for the given source is $R_1 > H(U_1|U_2)$, $R_2 > H(U_2|U_1)$, and $R_1 + R_2 > H(U_1, U_2)$. By Theorem 7, the capacity region for the given channel is $R_1 + R_2 > 1 - q(1)$. If $H(U_1, U_2) < 1 - q(1)$, then the regions overlap, and the given source can reliably communicated across the given channel with separate source and channel coding schemes. (Here $H(U_1, U_2) < 1 - q(1)$ implies that $H(U_1|U_2) < 1 - q(1)$

and $H(U_2|U_1) < 1 - q(1)$ and that there exists some (R_1, R_2) with $R_1 > H(U_1|U_2)$, $R_2 > H(U_2|U_1)$, and $R_1 + R_2 > H(U_1, U_2)$.)

To prove the converse, note that separation holds for the channel with vector input (X_1, X_2) and scalar output Y. Thus even if the two transmitters could cooperate, no source pair (U_1, U_2) with $H(U_1, U_2) > 1 - q(1)$ could be reliably transmitted across the given communication system, giving the desired result. □

Theorem 10 *Consider a multiple access channel with input-independent, additive noise. Suppose that the input alphabets, output alphabet, and noise alphabet are all equal to the binary field $\mathbb{F}_2$. Let noise $Z_1, Z_2, \ldots$ be drawn iid according to distribution $q(z)$. If source pair $(U_{1,1}, U_{2,1}), (U_{1,2}, U_{2,2}), \ldots$ is drawn iid according to distribution $p(u_1, u_2)$ with $H(U_1, U_2) < 1 - H(Z)$, then there exists a sequence of joint source-channel codes with probability of error $P_e^{(n)} \to 0$. Conversely, if $H(U_1, U_2) > 1 - H(Z)$, then the probability of error is bounded away from zero.*

Proof: As in the previous proof, given input-independent noise, separation holds for the channel with vector input (X_1, X_2) and scalar output $Y = X_1 + X_2 + Z$, making reliable communication of any source with $H(U_1, U_2) > 1 - H(Z)$ impossible for the given channel. Reliable communication for any source with $H(U_1, U_2) < 1 - H(Z)$ is achieved by separate source and channel coding since the Slepian-Wolf region and capacity region again overlap. □

We next demonstrate the performance of random linear codes in joint source-channel coding across the given linear channel.

Theorem 11 *Consider the random source $(U_{1,1}, U_{2,1}), (U_{1,2}, U_{2,2}), \ldots$ drawn iid according to distribution $p(u_1, u_2)$, and let $Z_1, Z_2, \ldots$ be the channel's random erasures, where $Z_1, Z_2, \ldots$ are drawn iid according to distribution $q(z)$, all Z_i are independent of the source, and $Z_i = 1$ denotes an erasure in channel use i. Assume that the source and channel input alphabets are equal to the binary field $\mathbb{F}_2$. Let $\{(C_{n,1}, C_{n,2})\}_{n=1}^{\infty}$ describe a sequence of linear joint source-channel codes. Each $C_{i,n}$ $(i \in \{1, 2\})$ is an $n \times n$ matrix with elements chosen iid Bernoulli$(1/2)$. If $H(U_1, U_2) < 1 - q(1)$, then the expected error probability $EP_e(C_n) \to 0$ as $n \to \infty$.*

Proof: Again, we begin by noting the erasure positions in Y^n and using them to reconstruct Z^n. A decoding error occurs if there exists a $\hat{\mathbf{u}}_1 \neq \mathbf{U}_1$ for which $C_{1,n}(\mathbf{U}_1 - \hat{\mathbf{u}}_1) \in \mathcal{E}(Z^n)$, a $\hat{\mathbf{u}}_2 \neq \mathbf{U}_2$ for which $C_{2,n}(\mathbf{U}_2 - \hat{\mathbf{u}}_2) \in \mathcal{E}(Z^n)$, or a $\hat{\mathbf{u}}_1 \neq \mathbf{U}_1$ and $\hat{\mathbf{u}}_2 \neq \mathbf{U}_2$ for which $C_{1,n}(\mathbf{U}_1 - \hat{\mathbf{u}}_1) + C_{2,n}(\mathbf{U}_2 - \hat{\mathbf{u}}_2) \in \mathcal{E}(Z^n)$. Thus

$$
\begin{aligned}
&EP_e^{(n)}(C_{1,n}, C_{2,n}) \\
&= \; E\Pr\left(\text{Error} \wedge \left((U_1^n, U_2^n) \notin A_\epsilon^{(n)}(p) \vee Z^n \notin A_\epsilon^{(n)}(q)\right)\right) \\
&\quad +E\Pr\left(\text{Error} \wedge (U_1^n, U_2^n) \in A_\epsilon^{(n)}(p) \wedge Z^n \in A_\epsilon^{(n)}(q)\right) \\
&\leq \; 2\epsilon_n + \sum_{(u_1^n,u_2^n)\in A_\epsilon^{(n)}(p)} \; \sum_{z^n\in A_\epsilon^{(n)}(q)} p(u_1^n, u_2^n) q(z^n) \\
&\quad \cdot \left[\sum_{\hat{u}_1^n \neq u_1^n : (\hat{u}_1^n, u_2^n) \in A_\epsilon^{(n)}(p)} \Pr(C_{1,n}(\mathbf{u}_1 - \hat{\mathbf{u}}_1) \in \mathcal{E}(z^n)) \right. \\
&\quad + \sum_{\hat{u}_2^n \neq u_2^n : (u_1^n, \hat{u}_2^n) \in A_\epsilon^{(n)}(p)} \Pr(C_{2,n}(\mathbf{u}_2 - \hat{\mathbf{u}}_2) \in \mathcal{E}(z^n)) \\
&\quad \left. + \sum_{\hat{u}_1^n \neq u_1^n, \hat{u}_2^n \neq u_2^n : (\hat{u}_1^n, \hat{u}_2^n) \in A_\epsilon^{(n)}(p)} \Pr(C_{1,n}(\mathbf{u}_1 - \hat{\mathbf{u}}_1) + C_{2,n}(\mathbf{u}_2 - \hat{\mathbf{u}}_2) \in \mathcal{E}(z^n)) \right] \\
&\leq \; 2\epsilon_n + \sum_{(u_1^n,u_2^n)\in A_\epsilon^{(n)}(p)} \; \sum_{z^n\in A_\epsilon^{(n)}(q)} p(u_1^n, u_2^n) q(z^n) \\
&\quad \cdot \left[2^{n(H(U_1|U_2)+\epsilon)} 2^{-n} 2^{n(q(1)+\epsilon')} + 2^{n(H(U_2|U_1)+\epsilon)} 2^{-n} 2^{n(q(1)+\epsilon')} + 2^{n(H(U_1,U_2)+\epsilon)} 2^{-n} 2^{n(q(1)+\epsilon')} \right] \\
&\leq \; 2\epsilon_n + 2^{-n(1-q(1)-\epsilon'-H(U_1|U_2)-\epsilon)} + 2^{-n(1-q(1)-\epsilon'-H(U_2|U_1)-\epsilon)} + 2^{-n(1-q(1)-\epsilon'-H(U_1,U_2)-\epsilon)}.
\end{aligned}
$$

for some $\epsilon_n \to 0$. Thus the expected error probability decays to zero as n grows without bound provided that $\max\{H(U_1|U_2), H(U_2|U_1), H(U_1,U_2)\} = H(U_1,U_2) < 1 - q(1) - \epsilon - \epsilon'$. □

Theorem 12 *Consider the random source $(U_{1,1}, U_{2,1}), (U_{1,2}, U_{2,2}), \ldots$ drawn iid according to distribution $p(u_1, u_2)$, and let $Z_1, Z_2, \ldots$ be the channel's random additive noise, where $Z_1, Z_2, \ldots$ are drawn iid according to distribution $q(z)$, and Z_i are independent of the source. Assume that the source, channel input, channel output, and noise alphabets are all equal to the binary field $\mathbb{F}_2$. Let $\{(C_{1,n}, C_{2,n})\}_{n=1}^{\infty}$ describe a sequence of linear joint source-channel codes. Each $C_{i,n}$ $(i \in \{1,2\})$ is an $n \times n$ matrix with elements chosen iid Bernoulli$(1/2)$. If*

$H(U_1, U_2) < 1 - H(Z)$, then the expected error probability $EP_e(C_{1,n}, C_{2,n}) \to 0$ as $n \to \infty$.

Proof: An error occurs if two values of u_1^n are mapped to the same value of x_1^n, two values of u_2^n are mapped to the same value of x_2^n, or if there exist distinct noise vectors that map distinct source vectors to the same channel output. In the first case, $C_{1,n}\mathbf{U}_1 = C_{1,n}\hat{\mathbf{u}}_1$ for some $\hat{\mathbf{u}}_1 \neq \mathbf{U}_1$; in the second case, $C_{2,n}\mathbf{U}_2 = C_{2,n}\hat{\mathbf{u}}_2$ for some $\hat{\mathbf{u}}_2 \neq \mathbf{U}_2$; and in the third case, $C_{1,n}\mathbf{U}_1 + \mathbf{Z} = C_{1,n}\hat{\mathbf{u}}_1 + \hat{\mathbf{z}}$ for some $\hat{\mathbf{u}}_1 \neq \mathbf{U}_1$ and $\hat{\mathbf{z}} \neq \mathbf{Z}$, $C_{2,n}\mathbf{U}_2 + \mathbf{Z} = C_{2,n}\hat{\mathbf{u}}_2 + \hat{\mathbf{z}}$ for some $\hat{\mathbf{u}}_2 \neq \mathbf{U}_2$ and $\hat{\mathbf{z}} \neq \mathbf{Z}$, or $C_{1,n}\mathbf{U}_1 + C_{2,n}\mathbf{U}_2 + \mathbf{Z} = C_{1,n}\hat{\mathbf{u}}_1 + C_{2,n}\hat{\mathbf{u}}_2 + \hat{\mathbf{z}}$ for some $\hat{\mathbf{u}}_2 \neq \mathbf{U}_2$, $\hat{\mathbf{u}}_2 \neq \mathbf{U}_2$, and $\hat{\mathbf{z}} \neq \mathbf{Z}$. Thus, setting $\mathcal{F}(z^n) = \{\hat{\mathbf{z}} - \mathbf{z} : \hat{\mathbf{z}} \neq \mathbf{z}, \hat{\mathbf{z}}^t \in A_\epsilon^{(n)}(q)\}$ and restricting our attention to typical error sequences, we sum up the error events as: $C_{1,n}(\mathbf{U}_1 - \hat{\mathbf{u}}_1) \in \{\mathbf{0}\} \cup \mathcal{F}(Z^n)$, $C_{2,n}(\mathbf{U}_2 - \hat{\mathbf{u}}_2) \in \{\mathbf{0}\} \cup \mathcal{F}(Z^n)$, and $C_{1,n}(\mathbf{U}_1 - \hat{\mathbf{u}}_1) + C_{2,n}(\mathbf{U}_2 - \hat{\mathbf{u}}_2) \in \mathcal{F}(Z^n)$. We then bound the expected error probability as

$$\begin{aligned}
&EP_e^{(n)}(C_{1,n}, C_{2,n}) \\
&= E\Pr\left(\text{Error} \wedge \left((U_1^n, U_2^n) \notin A_\epsilon^{(n)}(p) \vee Z^n \notin A_\epsilon^{(n)}(q)\right)\right) \\
&\quad + E\Pr\left(\text{Error} \wedge (U_1^n, U_2^n) \in A_\epsilon^{(n)}(p) \wedge Z^n \in A_\epsilon^{(n)}(q)\right) \\
&\leq 2\epsilon_n + \sum_{(u_1^n, u_2^n) \in A_\epsilon^{(n)}(p)} \sum_{z^n \in A_\epsilon^{(n)}(q)} p(u_1^n, u_2^n) q(z^n) \\
&\quad \cdot \Bigg[\sum_{\hat{u}_1^n \neq u_1^n : (\hat{u}_1^n, u_2^n) \in A_\epsilon^{(n)}(p)} \Pr(C_{1,n}(\mathbf{u}_1 - \hat{\mathbf{u}}_1) \in \{\mathbf{0}\} \cup \mathcal{F}(z^n)) \\
&\quad + \sum_{\hat{u}_2^n \neq u_2^n : (u_1^n, \hat{u}_2^n) \in A_\epsilon^{(n)}(p)} \Pr(C_{2,n}(\mathbf{u}_2 - \hat{\mathbf{u}}_2) \in \{\mathbf{0}\} \cup \mathcal{F}(z^n)) \\
&\quad + \sum_{\hat{u}_1^n \neq u_1^n, \hat{u}_2^n \neq u_2^n : (\hat{u}_1^n, \hat{u}_2^n) \in A_\epsilon^{(n)}(p)} \Pr(C_{1,n}(\mathbf{u}_1 - \hat{\mathbf{u}}_1) + C_{2,n}(\mathbf{u}_2 - \hat{\mathbf{u}}_2) \in \mathcal{F}(z^n)) \Bigg] \\
&\leq 2\epsilon_n + \sum_{(u_1^n, u_2^n) \in A_\epsilon^{(n)}(p)} \sum_{z^n \in A_\epsilon^{(n)}(q)} p(u_1^n, u_2^n) q(z^n) \left[2^{n(H(U_1|U_2)+\epsilon)} 2^{-n} 2^{n(H(Z)+\epsilon)} \right. \\
&\quad \left. + 2^{n(H(U_2|U_1)+\epsilon)} 2^{-n} 2^{n(H(Z)+\epsilon)} + 2^{n(H(U_1,U_2)+\epsilon)} 2^{-n} 2^{n(H(Z)+\epsilon)} \right] \\
&\leq 2\epsilon_n + 2^{-n(1-H(Z)-H(U_1|U_2)-2\epsilon)} + 2^{-n(1-H(Z)-H(U_2|U_1)-2\epsilon)} + 2^{-n(1-H(Z)-H(U_1,U_2)-2\epsilon)}
\end{aligned}$$

for some $\epsilon_n \to 0$. Thus the expected error probability decays to zero as n grows without bound provided that $\max\{H(U_1|U_2), H(U_2|U_1), H(U_1, U_2)\} = H(U_1, U_2) < 1 - H(Z) - 2\epsilon$.

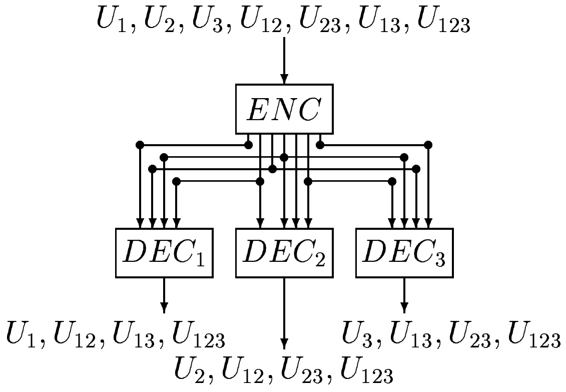

$U_1, U_2, U_3, U_{12}, U_{23}, U_{13}, U_{123}$
ENC
DEC_1
DEC_2
DEC_3
$U_1, U_{12}, U_{13}, U_{123}$
$U_3, U_{13}, U_{23}, U_{123}$
$U_2, U_{12}, U_{23}, U_{123}$

Figure 3: A broadcast system source code with three receivers.

□

VI Broadcast Systems

The next simple model under consideration is the broadcast system, where one transmitter sends information to a collection of receivers.

A broadcast system source code comprises a single encoder and a collection of decoders. Since the case with two receivers has special structure absent from general broadcast system source codes [28, 29], we focus on the three-receiver system of Figure 3. The results given simplify easily to the two-receiver case and generalize to more receivers. Note that, since we consider discrete channels, the degraded broadcast channel converses of [38] or of [39], which allows no or partial common information, are applicable. In the given broadcast system source coding model, samples of source vector $(U_1, U_2, U_3, U_{12}, U_{23}, U_{13}, U_{123})$ are drawn iid from some distribution $p(u_1, u_2, u_3, u_{12}, u_{23}, u_{13}, u_{123})$. The source description contains components of rates R_1, R_2, R_3, R_{12}, R_{23}, R_{13}, and R_{123}. Decoder 1 receives the rate R_1, R_{12}, R_{13}, and R_{123} descriptions and uses them to decode $(U_1, U_{12}, U_{13}, U_{123})$. Decoder 2 receives the rate R_2, R_{12}, R_{23}, and R_{123} descriptions and uses them to decode $(U_2, U_{12}, U_{23}, U_{123})$. Decoder 3 receives the rate R_3, R_{13}, R_{23}, and R_{123} descriptions and uses them to decode $(U_3, U_{13}, U_{23}, U_{123})$. While several receivers decode the common information, each has a different subset of the descriptions with which to decode.

Theorem 13 proves the optimality of linear broadcast system source codes. In this case, the linear encoder is a matrix of dimension

$$(\lceil nR_1 \rceil + \lceil nR_2 \rceil + \lceil nR_3 \rceil + \lceil nR_{12} \rceil + \lceil nR_{23} \rceil + \lceil nR_{13} \rceil + \lceil nR_{123} \rceil) \times n.$$

The first $\lceil nR_1 \rceil$ bits of the output go to receiver 1 only. The subsequent $\lceil nR_2 \rceil$ and $\lceil nR_3 \rceil$ bits similarly go to receivers 2 and 3, respectively. Next come, in order, the rate-R_{12}, R_{23}, R_{13}, and R_{123} descriptions. We again use typical set decoding.

Theorem 13 *Consider samples of source vector* $(U_1, U_2, U_3, U_{12}, U_{23}, U_{13}, U_{123})$ *drawn iid according to distribution* $p(u_1, u_2, u_3, u_{12}, u_{23}, u_{13}, u_{123})$ *on* $(\mathbb{F}_2)^7$*. Let* $\{A_n\}_{n=1}^{\infty}$ *be a sequence of rate-*$(R_1, R_2, R_3, R_{12}, R_{23}, R_{13}, R_{123})$ *linear broadcast system source codes with coefficients chosen iid Bernoulli*(1/2)*. For any* $s \subseteq \{1, 2, 3, 12, 23, 13, 123\}$*, let* $u_s = (u_a)_{a \in s}$*, and let* $(nR)_s = \sum_{a \in s} \lceil nR_a \rceil$*. Then for any rates satisfying*

$$(nR)_s \geq H(U_s | U_{S_1 - s}) \quad \forall \quad s \subseteq S_1 = \{1, 12, 13, 123\}, s \neq \phi$$

$$(nR)_s \geq H(U_s | U_{S_2 - s}) \quad \forall \quad s \subseteq S_2 = \{2, 12, 23, 123\}, s \neq \phi$$

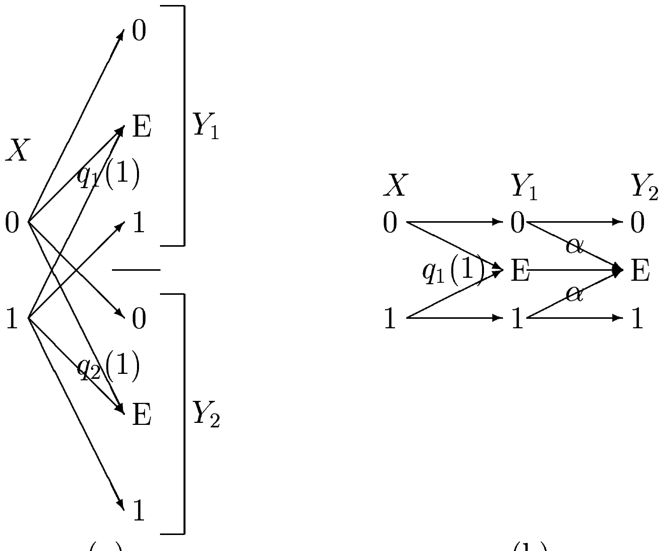
$$(nR)_s \geq H(U_s | U_{S_3 - s}) \quad \forall \quad s \subseteq S_3 = \{3, 13, 23, 123\}, s \neq \phi$$

$\{A_n\}_{n=1}^{\infty}$ *achieves expected error probability* $EP_e(A_n) \to 0$ *as* $n \to \infty$*.*

Proof: Given the linear structure of the code, we can break encoder matrix A_n into a collection of $\lceil nR_a \rceil \times n$ sub-matrices, $a \in \{1, 2, 3, 12, 23, 13, 123\}$, such that

$$A_n = \begin{bmatrix} A_{1,n} \\ A_{2,n} \\ A_{3,n} \\ A_{12,n} \\ A_{23,n} \\ A_{13,n} \\ A_{123,n} \end{bmatrix}.$$

We begin by bounding the expected probability of decoding in error at receiver 1, here denoted as $EP_e(A_{1,n}, A_{12,n}, A_{13,n}, A_{123,n})$ The arguments for receivers 2 and 3 are similar. By the union bound, the code error probability is bounded by the sum of the individual decoder error probabilities.



(a) (b)

Figure 4: (a) The erasure broadcast channel and (b) a physically degraded channel with the same capacity ($\alpha = (q_2(1) - q_1(1))/(1 - q_1(1))$ and all erasures propagate as erasures).

An error occurs at receiver 1 if any subset of the desired sources is decoded in error. Thus, following our standard approach,

$$\begin{aligned} EP_e(A_{1,n}, A_{12,n}, A_{13,n}, A_{123,n}) &\leq \epsilon_n + \sum_{(u_1^n, u_{12}^n, u_{13}^n, u_{123}^n) \in A_\epsilon^{(n)}} p(u_1^n, u_{12}^n, u_{13}^n, u_{123}^n) \\ &\quad \cdot \sum_{s \subseteq S_1 : s \neq \phi} \sum_{\hat{u}_s^n \neq u_s^n : (\hat{u}_s^n, u_{S_1 - s}^n) \in A_\epsilon^{(n)}} \Pr(A_{s,n}(\mathbf{u}_s - \hat{\mathbf{u}}_s) = \mathbf{0}) \\ &\leq \epsilon_n + \sum_{s \subseteq S_1 : s \neq \phi} 2^{n(H(U_s | U_{S_1 - s}) + 2\epsilon)} 2^{-(nR)_s} \end{aligned}$$

for some $\epsilon_n \to 0$. □

We next consider the erasure broadcast channel models shown in Figure 4 (a) and (b). A single channel input is sent to receivers 1 and 2. In the first model, the output at receiver 1 is an erasure with probability $q_1(1)$ and the transmitted value with probability $q_1(0)$; likewise, the output at receiver 2 is an erasure with probability $q_2(1)$ and is otherwise received correctly. Without loss of generality, assume that $q_1(1) \leq q_2(1)$. In this model, erasures are assumed to be independent events. In the model of Figure 4(b), the erasure probabilities for the two receivers are the same, but the erasures are dependent random variables, with all erasures at the first receiver propagating to the second receiver. By [1, Theorem 14.6.1], the capacity of the broadcast channel depends only on the conditional marginal distributions $p(y_1|x)$ and $p(y_2|x)$, thus the capacity of the two channels shown and all channels with the same $p(y_1|x)$ and $p(y_2|x)$ (regardless of the statistical dependencies between erasure events Z_1 and Z_2) are identical.[3] Note that the elegant and simple converse for degraded BSC broadcast channels of [40], which relies on properties of binary sequences, might be readily extended to our model, albeit without the generality of [38, 39].

Lemma 3 proves time-sharing to be optimal for broadcast coding over the given family of channels. The result of Theorem 14, proving the rates achievable by linear broadcast channel codes on the erasure broadcast channel is then immediate by the previous linearity of time-sharing argument. The given bound is optimal for the case of no common information. No converse exists for the case of common information, but the given linear coding achievability results agree with the best known achievability results on the binary erasure channel.

Lemma 3 *Consider a binary erasure channel with output alphabets* $\{0, 1, E\}$ *at each of two receivers. The erasure sequences* $Z_{1,1}, Z_{1,2}, \ldots$ *and* $Z_{2,1}, Z_{2,2}, \ldots$ *are drawn iid according to distributions* $q_1(z_1)$ *and* $q_2(z_2)$*, respectively, where* $Z_{i,j} = 1$ *denotes an erasure event at receiver* i *at time* j*. The joint distribution* $q(z_1, z_2)$ *may be any distribution with the given marginals, but the channel noise is independent of the channel input by assumption. The capacity region for sending independent information to the two receivers is described by*

$$\frac{R_1}{1 - q_1(1)} + \frac{R_2}{1 - q_2(1)} \leq 1.$$

For any achievable independent information rate pair (R_1, R_2)*, the rate triple* $(R_1', R_2', R_{12}') = (R_1, R_2 - R_0, R_0)$ *with common information rate* R_{12}' *and independent information rates* R_1' *and* R_2' *is also achievable for any* $R_0 < R_2$*.*

Proof: By [1, Theorem 14.6.1,Theorem 14.6.2], the capacity of the given channel is the convex hull of the closure of all (R_1, R_2) satisfying $R_2 \leq I(W; Y_2)$ and $R_1 \leq I(X; Y_1|W)$ for some joint distribution $p(w)p(x|w)p(y_1|x)p(y_2|y_1)$. Here W is an auxiliary random variable with alphabet size 2 and $p(y_2|y_1)$ is derived from the physically degraded channel model. By a

[3] All channel models considered here assume Z_1 and Z_2 are independent of the channel input.

symmetry argument, the optimal W is a uniform binary random variable with $p(x|w) = 1-\beta$ if $x = w$ and $p(x|w) = \beta$ otherwise. Thus

$$\begin{aligned}
R_1 &\leq I(X;Y_1|W) \\
&= I(X;Y_1) - I(W;Y_1) \\
&= (1-q_1(1)) - [H((1-q_1(1))/2, q_1(1), (1-q_1(1))/2) \\
&\quad -H((1-\beta)(1-q_1(1)), q_1(1), \beta(1-q_1(1)))] \\
&= (1-q_1(1))H(\beta) \\
R_2 &\leq I(W;Y_2) \\
&= H((1-q_1(1))(1-\alpha)/2, q_1(1)+(1-q_1(1))\alpha, (1-q_1(1))(1-\alpha)/2) \\
&\quad -H((1-\beta)(1-q_1(1))(1-\alpha), q_1(1)+(1-q_1(1))\alpha, \beta(1-q_1(1))(1-\alpha)) \\
&= (1-q_1(1))(1-\alpha)(1-H(\beta)) \\
&= (1-q_2(1))(1-H(\beta)).
\end{aligned}$$

Varying $H(\beta)$ from 0 to 1 gives the independent coding result. The common information result comes from [1, Theorem14.6.4]. □

Theorem 14 *Consider an erasure channel with input alphabet $\mathbb{F}_2$ and output alphabets $\{0,1,E\}$ at each of two receivers. The erasure sequences $Z_{1,1}, Z_{1,2}, \ldots$ and $Z_{2,1}, Z_{2,2}, \ldots$ are drawn iid according to distributions $q_1(z_1)$ and $q_2(z_2)$, respectively, where $Z_{i,j} = 1$ denotes an erasure event at receiver i at time j. The joint distribution $q(z_1, z_2)$ may be any distribution with the given marginals, but the channel noise is independent of the channel input by assumption. Let $\{B_n\}_{n=1}^{\infty}$ describe a sequence of channel codes. Each B_n is an $n \times (\lfloor nR_1 \rfloor + \lfloor nR_2 \rfloor)$ matrix with elements chosen iid Bernoulli$(1/2)$. If $R_1/(1-q_1(1)) + R_2/(1-q_2(1)) < 1$, then the expected error probability $EP_e(B_n) \to 0$ as $n \to \infty$.*

To date, there exist no results to prove the optimality of linear broadcast codes for the additive noise broadcast channel model. In this case, time-sharing is not the optimal solution [1], and direct application of the techniques used in this paper fail to achieve the optimal performance. The stumbling block is that we cannot apply the construction used to build channel input X from the auxiliary random variable W to be decoded by the second receiver. (See the proof of Lemma 3.) In particular, we cannot achieve the appropriate (non-uniform) cross-over probability from the auxiliary random variable to X using an additive signal created by a linear code. In this case, as in Theorem 14, the time-sharing solution is achievable with linear coding. While the time-sharing solution gives a bound on the performance achievable by linear coding, the time-sharing solution is sub-optimal for this problem. Linear coding performance beyond the time-sharing bound may or may not be possible. The following argument describes one possible strategy for trying to move linear codes beyond the time-sharing bound. Consider a systematic code with a low density parity-check matrix. Let the encoding matrix be

$$B_n = \begin{bmatrix} \multicolumn{2}{c}{I} \\ P_{11} & P_{21} \\ \mathbf{0} & P_{22} \end{bmatrix},$$

where I is the $(\lfloor nR_1 \rfloor + \lfloor nR_2 \rfloor) \times (\lfloor nR_1 \rfloor + \lfloor nR_2 \rfloor)$ identity matrix and P_{11}, P_{21}, and P_{22} have dimensions

$$nR_1 \frac{H(Z_1)}{1-H(Z_1)} \times nR_1, \quad nR_1 \frac{H(Z_1)}{1-H(Z_1)} \times nR_2, \quad \left(n - nR_2 - nR_1 \frac{1}{1-H(Z_1)}\right) \times nR_2,$$

respectively. (We here drop the rounding notation for readability but note that all of the above quantities must be integers.) For each $i \in \{1,2\}$, let $\mathbf{Z}_i^t = [\mathbf{Z}_{i1}^t \mathbf{Z}_{i2}^t \mathbf{Z}_{i3}^t \mathbf{Z}_{i4}^t]$, where the sub-vectors have lengths nR_1, nR_2, $nR_1 H(Z_1)/(1-H(Z_1))$, and $n - nR_2 - nR_1/(1-H(Z_1))$, respectively. Applying the above code, the channel output at receiver 2 is

$$\mathbf{Y}_2 = \begin{bmatrix} \mathbf{V}_1 + \mathbf{Z}_{21} \\ \mathbf{V}_2 + \mathbf{Z}_{22} \\ P_{11}\mathbf{V}_1 + P_{21}\mathbf{V}_2 + \mathbf{Z}_{23} \\ P_{22}\mathbf{V}_2 + \mathbf{Z}_{24} \end{bmatrix}.$$

If the decoder at that receiver applies parity check matrix P_{11} to the the received $\mathbf{V}_1 + \mathbf{Z}_{21}$ and subtracts off the outcome from the third component of $\mathbf{Y}$ then the modified signal is

$$\mathbf{Y}_2 - \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ P_{11}(\mathbf{V}_1 + \mathbf{Z}_{21}) \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1 + \mathbf{Z}_{21} \\ \mathbf{V}_2 + \mathbf{Z}_{22} \\ P_{21}\mathbf{V}_2 + \mathbf{Z}_{23} + P_{11}\mathbf{Z}_{21} \\ P_{22}\mathbf{V}_2 + \mathbf{Z}_{24} \end{bmatrix}.$$

Decoder 2 thereby recovers more of its parity check symbols at the expense of increasing the corresponding error probability in those symbols. When the density of parity check matrix P_{11} is low, the increase in error probability for symbols $P_{21}\mathbf{V}_2$ may also be low enough to make those parity check bits useful in decoding the description of $\mathbf{V}_2$. Receiver 1 uses the same technique to decode $\mathbf{V}_2$, then subtracts off its impact on the parity check bits for $\mathbf{V}_1$, and finally decodes $\mathbf{V}_1$.

VII Input-Dependent Noise

By assuming that the channel noise is independent of the channel input, the theorems of the previous section rule out asymmetrical channels like the Z-channel. Unfortunately, the above techniques do not extend to the case where the noise random variable is dependent on the channel input. For example, in the single-transmitter, single-receiver network, more careful choice of B_n can yield an arbitrary desired distribution on the channel input vectors. Unfortunately, it cannot do so in a manner that allows the set of codewords to cover the typical set. In particular, if row i of the generator matrix b_n is nonzero, then the ith coefficient of channel codeword $X^n = b_n\mathbf{V}$ is a mixture of iid, Bernoulli(1/2) random values (assuming that the message $\mathbf{V}$ is chosen uniformly over the space $\mathbb{F}_2^{\lfloor nR \rfloor}$ of possible messages), and thus is itself Bernoulli(1/2). While we can make the distribution on X_i Bernoulli(q) for some $q < 1/2$ by mixing proportion $(1-2q)$ all-zero rows with proportion $2q$ non-zero rows, the number of distinct channel codewords that result is at most $2^{n(2q)} < 2^{nH(q)}$ for all $q \in (0, 1/2)$ by the strict concavity of the the entropy function.

The above observation of the failure of linear codes for input-dependent noise demonstrates that while separation holds for channels like the Z-channel, separation does not hold in general for *linear* codes on single-transmitter, single-receiver channels with input-dependent noise. For example, consider an additive noise channel with input-dependent noise $Z_1, Z_2, \ldots$, and suppose that linear channel codes cannot achieve the capacity of the given channel. Now consider a source $U_1, U_2, \ldots$ such that the statistic of source U are precisely the optimal input statistics for the given channel. For this example, optimal linear source coding followed by optimal linear channel coding would fail to achieve the optimal performance but joint coding – in this case, no coding – would succeed. Unfortunately, while the above example demonstrates a case where linear joint source-channel coding achieves the optimal performance while separated linear codes fail, the linear joint source-channel coding techniques used in this paper fail to achieve the optimal performance in this example.

The above observation about the failure of source-channel separation for linear coding on channels with input-dependent noise is one of many examples where separation fails owing to input-dependent noise. In that example, separation fails for linear codes but does not fail in general. Theorem 15, which treats the additive multiple access channel with additive noise, provides an example where source-channel separation fails more generally owing to input-dependent noise. The same phenomena may be observed in erasure channels.

Theorem 15 *Consider a multiple access channel where the input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$, output alphabet $\mathcal{Y}$, and noise alphabet $\mathcal{Z}$ are all equal to the binary field $\mathbb{F}_2$. Let $Z_1, Z_2, \ldots$ be the noise random process, and use $X_{1,i}$ and $X_{2,i}$ to describe the channel inputs at time i. Then the channel output at time i is $Y_i = X_{1,i} + X_{2,i} + Z_i$. Separation fails when Z_i and $(X_{1,i}, X_{2,i})$ are statistically dependent random variables.*

Proof: Following the argument of Theorem 10, we demonstrate the failure of separation by showing that when the cooperative capacity of the network is different from the multiple access capacity, then there exist sources for which joint coding can reliably transmit sources that cannot be sent reliably using separate source and channel coding.

The maximal rate attainable in separate source and channel coding is bounded by the multiple access channel capacity's bound on the sum rate

$$R_1 + R_2 \leq \max_{P_1, P_2} I(X_1, X_2; Y),$$

where P_1 and P_2 are the marginal probability mass functions of X_1 and X_2, respectively. The cooperative capacity of the network provides the alternative bound

$$R_1 + R_2 \leq \max_{P_{12}} I(X_1, X_2; Y).$$

When

$$\max_{P_1, P_2} I(X_1, X_2; Y) < \max_{P_{12}} I(X_1, X_2; Y),$$

separation fails since the cooperative capacity is achievable through joint coding for the source with $p(u_1, u_2) = P_{12}(u_1, u_2)$.

For all $i, j \in \{0, 1\}$, let $\Pr(Z = 1|X_1 = i, X_1 = j) = q_{ij}$. For the multiple access capacity, let $p_1 = \Pr(X_1 = 1)$ and $p_2 = \Pr(X_2 = 1)$. Then

$$\begin{aligned}
&\max_{P_1,P_2} I(X_1, X_2; Y) \\
&\quad= \max_{p_1,p_2}[H((1-p_1)(1-p_2)(1-q_{00}) + (1-p_1)p_2q_{01} + p_1(1-p_2)q_{10} + p_1p_2(1-q_{11})) \\
&\qquad -(1-p_1)(1-p_2)H(q_{00}) - (1-p_1)p_2H(q_{01}) - p_1(1-p_2)H(q_{10}) - p_1p_2H(q_{11})].
\end{aligned}$$

For the cooperative capacity, let $\Pr(X_1 = 0, X_2 = 0) = p_{00}$, $\Pr(X_1 = 0, X_2 = 1) = p_{01}$, $\Pr(X_1 = 1, X_2 = 0) = p_{10}$, and $\Pr(X_1 = 1, X_2 = 1) = p_{11}$, where $p_{11} = 1 - p_{00} - p_{01} - p_{10}$. Then we similarly find

$$\begin{aligned}
\max_{P_{12}} I(X_1, X_2; Y) &= \max_{p_{00},p_{01},p_{10},p_{11}} [H(Y) - H(Y|U, V)] \\
&= \max_{p_{00},p_{01},p_{10},p_{11}} [H(p_{00}(1-q_{00}) + p_{01}q_{01} + p_{10}q_{10} + p_{11}(1-q_{11})) \\
&\qquad -p_{00}H(q_{00}) - p_{01}H(q_{01}) - p_{10}H(q_{10}) - p_{11}H(q_{11})].
\end{aligned}$$

The two equations are not equal in general. For example, let $q_{00} = q_{11} = 0$ while $q_{01} = q_{10} = 1/2$. Then

$$\begin{aligned}
\max_{P_1P_2} I(X_1, X_2; Y) &= H(3/4) - 1/2 = 0.311 \\
\max_{P_{12}} I(X_1, X_2; Y) &= 1.
\end{aligned}$$

(The maxima occur at $(p_1, p_2) = (1/2, 1/2)$ and $(p_{00}, p_{01}, p_{10}, p_{11}) = (1/2, 0, 0, 1/2)$.) Separation fails in this example since the source pair (U_1, U_2) with $\Pr(U_0 = 0, U_1 = 0) = \Pr(U_0 = 1, U_1 = 1) = 1/2$ can be reliably transmitted across the given channel, despite the fact that the achievable rate region for Slepian-Wolf source coding and the capacity region for the given channel do not overlap. (Slepian-Wolf source coding requires a rate $R_1 + R_2 \geq 1$ while the multiple access capacity region extends only as far as $R_1 + R_2 \leq .311$.)

In contrast, for input-independent noise $q_{00} = q_{01} = q_{10} = q_{11} = q$, giving

$$\begin{aligned}
&\max_{P_1P_2} I(X_1, X_2; Y) \\
&\quad= \max_{p_1,p_2}[H((1-p_1)(1-p_2)(1-q) + (1-p_1)p_2q + p_1(1-p_2)q + p_1p_2(1-q))
\end{aligned}$$

$$
\begin{aligned}
&\quad -(1-p_1)(1-p_2)H(q) - (1-p_1)p_2H(q) - p_1(1-p_2)H(q) - p_1p_2H(q)] \\
&= \max_{p_1,p_2}[H((1-(1-p_1)(1-p_2))(1-q) + (1-p_1)(1-p_2)q) - H(q)] \\
&= 1 - H(q).
\end{aligned}
$$

We achieve the maximum by setting $(1-p_1)(1-p_2) = 1/2$. Similarly, the cooperative capacity becomes

$$
\begin{aligned}
\max_{P_{12}} I(X_1, X_2; Y) &= \max_{p_{00},p_{01},p_{10},p_{11}} [H(p_{00}(1-q) + p_{01}q + p_{10}q + p_{11}(1-q)) \\
&\quad -p_{00}H(q) - p_{01}H(q) - p_{10}H(q) - p_{11}H(q)] \\
&= \max_{p_{00},p_{01},p_{10},p_{11}} [H((1-(p_{01}+p_{10}))(1-q) + (p_{01}+p_{10})q) - H(q)] \\
&= 1 - H(q).
\end{aligned}
$$

Again, we achieve the maximum when $p_{01} + p_{10} = 1/2$.

Comparing the two results in this case, we find that

$$\max_{P_1P_2} I(X_1, X_2; Y) = \max_{P_{12}} I(X_1, X_2; Y)$$

as expected from Theorem 10, and separation holds. □

Theorem 10 proves that separation fails for some multiple access adder channels with input-dependent noise and shows that the gap between the maximal joint entropy for sources that can be transmitted and the maximal achievable sum rate across the multiple access channel can be quite large (0.689 bits per symbol in the binary example provided in the proof). Figure 5 shows the corresponding difference for 100 randomly chosen sources. For these examples, we model the noise Z as the sum of two independent noise components, say $Z = Z_1 + Z_2$, where Z_1 is dependent on X_1 but not X_2, and Z_2 is dependent on X_2 but not X_1. In particular, we set $\Pr(Z_1 = 1|X_1 = 0) = \epsilon_{1,0}$, $\Pr(Z_1 = 1|X_1 = 1) = \epsilon_{1,1}$, $\Pr(Z_1 = 1|X_2 = 0) = \epsilon_{2,0}$, $\Pr(Z_1 = 1|X_2 = 1) = \epsilon_{2,1}$ and choose $\epsilon_{1,0}$, $\epsilon_{1,1}$, $\epsilon_{2,0}$, and $\epsilon_{2,1}$) independently at random according to the uniform distribution on $[0, 1]$. For each channel, we plot the difference between the maximal joint entropy for sources that can be transmitted and the maximal sum rate achievable across the multiple access channel. For examples where this value is non-zero, separation fails. While the example in the above proof suggests that the gap can be large, the gaps are far smaller for examples encountered in our experiments.

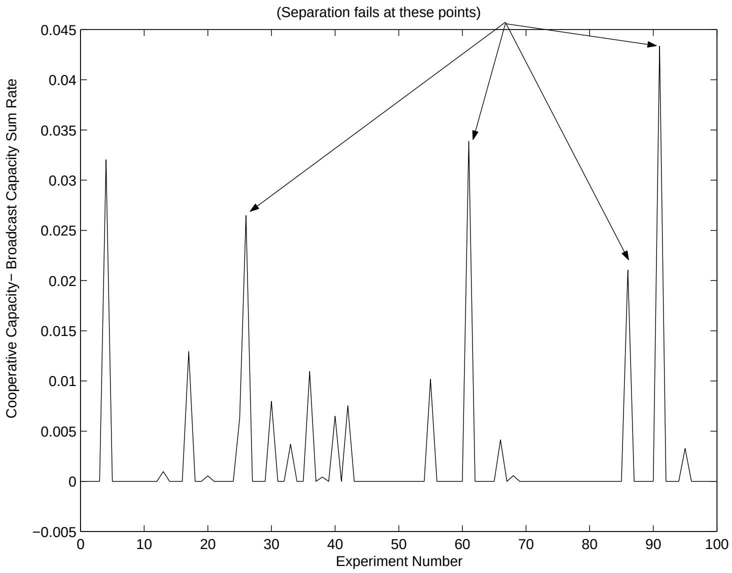


Figure 5: The difference between the maximal cooperative capacity and the maximal sum rate for the broadcast channel for a collection of multiple access channels with input-dependent additive noise.

VIII The Case for End-to-End Coding

The preceding sections treat the topics of source and channel coding using the tools of linear network coding, bringing previously disparate areas into a common framework. We end by demonstrating that this unification is not only useful in its combination of tasks once treated entirely separately but is in fact crucial to achieving optimal, reliable communication.

Traditional routing techniques rely entirely on repeat and forward strategies for getting a source from its point of origin to its desired destination. The network coding literature demonstrates the failure of that approach in achieving the optimal performance for some simple multi-cast examples [30]. We next demonstrate the failure of the network coding model.

The common network coding model assumes that all sources are independent and all links are noiseless. Implicit in the given model is the assumption that source and channel coding are performed separately from network coding at the edges of the network, so that the internal nodes need only pass along the information to the appropriate receivers. We

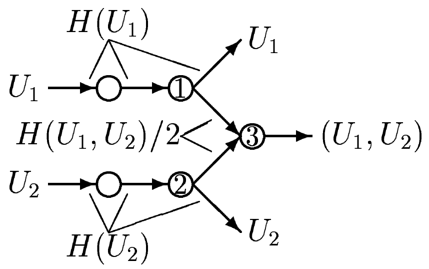


Figure 6: A network for which separation of source and network coding fails.

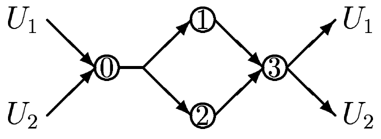


Figure 7: A network for which separation of channel and network coding fails.

next demonstrate that source-network separation and channel-network separation both fail. That is, there exist networks for which network coding and source coding must be performed jointly in order to achieve the optimal performance. Likewise, there exist networks for which network coding and channel coding must be performed jointly in order to achieve the optimal performance. We use a sequence of simple examples to prove these results.

Example 1: The network of Figure 6 comprises two transmitters and three receivers. Receiver node 1 wishes to receive U_1, receiver node 2 wishes to receive U_2, and receiver node 3 wishes to receive both U_1 and U_2. Sources (U_1, U_2) are dependent random variables, with $H(U_1, U_2) < H(U_1) + H(U_2)$. All network links are lossless, and the capacities are noted in the figure. Achieving reliable communication in this example requires the descriptions received by nodes 1 and 2 to be dependent random variables and requires sources U_1 and U_2 to be re-compressed at nodes 1 and 2, respectively. Thus separation of source coding and network coding fails. □

Example 2: Consider the network shown in Figure 7. The channel between node 0 and nodes 1 and 2 is a broadcast erasure channel with independent erasures of probabilities $q_1(1) = q_2(1) = q$. The network between nodes 1 and 2 and node 3 is a multiple access channel without interference. The network coding approach requires labeling each link with

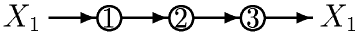


Figure 8: A network for which separation of source and network coding fails. The links between nodes 1 and 2 and nodes 2 and 3 are independent erasure channels with probabilities of erasure $q_1(1)$ and $q_2(1)$, respectively.

its corresponding link capacity. If R_1 and R_2 are the capacities of the edges to receivers 1 and 2, then R_1+R_2 must be less than $1-q$ by Theorem 14. The links from node 1 to node 3 and from node 2 to node 3 are both lossless, with capacity 1 bit per channel use. Optimal network coding on the given channel gives a maximal rate of $1-q$ from the encoder to the decoder. We contrast with the above separated channel and network coding approach an end-to-end coding strategy. In this case, we do not force zero error probability between node 0 and nodes 1 and 2 but instead simply forward the information received by those nodes to the decoder. The capacity of the resulting code is $1-q^2$ since receiver 3 suffers an erasure only if both node 1 and node 2 receive erasures. □

In addition to illustrating the failure of separate channel and network coding schemes, Example 2 serves as a reminder that general network capacities cannot be proven by breaking the network into canonical elements and solving them independently. Sadly, the strategy given for that example is not always optimal. In particular, the strategy discussed in Example 2 demonstrates that failure to decode at intermediate nodes of the network can yield performance superior to that achieved by decoding at intermediate nodes. Example 3 gives an example that teaches the opposite lesson.

Example 3: Consider the channel of Figure 8. The links between nodes 1 and 2 and nodes 2 and 3 are independent erasure channels with probabilities of erasure $q_1(1)$ and $q_2(1)$, respectively. If we do not decode at the intermediate node, then the maximal achievable rate from node 1 to node 3 is $(1-q_1(1))(1-q_2(1))$. Decoding at node 2 yields maximal achievable rate $\min\{1-q_1(1), 1-q_2(1)\} \geq (1-q_1(1))(1-q_2(1))$.

The failure of separation in Examples 1 and 2 and the contrasting lessons regarding decoding at intermediate nodes demonstrated by Examples 2 and 3 make the case for the need for end-to-end coding in network environments. The success of the linear coding technique in network coding, source coding, and channel coding suggests that a unified approach that obviates the need for separate routing, compression, and error correction codes may be within reach. In contrast, the failure of separation across canonical network systems seems to present a far greater challenge to optimal code design in networks.

References

[1] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, 1991.

[2] P. Elias. Coding for two noisy channels. In *Third London Symposium on Information Theory*, pages 61–74, London, 1956. Academic Press.

[3] D. J. C. MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory*, 45(2):399–431, March 1999.

[4] R. Ahlswede. Multi-way communication channels. In *Proc. 2nd. Int. Symp. Information Theory (Tsahkadsor, Armenian S.S.R.)*, pages 23–52, Prague, 1971. Publishing House of the Hungarian Academy of Sciences.

[5] H. Liao. *Multiple access channels.* Ph. D. Dissertation, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[6] S.-C. Chang and E. J. Weldon, Jr. Coding for t-user multiple access channels. *IEEE Transactions on Information Theory*, IT-25(6):684–691, November 1979.

[7] B. L. Hughes and A. B. Cooper. Nearly optimal multiuser codes for the binary adder channel. *IEEE Transactions on Information Theory*, 42(2):387–398, March 1996.

[8] G. H. Khachatrian and S. S. Martirossian. Code construction for the t-user noiseless adder channel. *IEEE Transactions on Information Theory*, 44(5):1953–1957, September 1998.

[9] J. Cheng and Y. Watanabe. t-user code with arbitrary code length for multiple-access adder channel. IEICE Transactions Fundamentals, E82-A(10), October 1999.

[10] P. Mathys. A class of codes for a t active users out of n multiple-access communication system. *IEEE Transactions on Information Theory*, IT-36(6):1206–1219, November 1990.

[11] B. Hughes. Capacity and coding for t active users out of m on the collision channel. In *Proceedings of the IEEE International Symposium on Information Theory*, page 323, San Antonio, Texas, January 1993.

[12] G. Poltyrev and J. Snyders. Linear codes for the sum mod-2 multiple-access channel with restricted access. *IEEE Transactions on Information Theory*, 41(3):794–799, May 1995.

[13] G. Caire, E. Leonardi, and E. Viterbo. Improving performance on wireless networks using collision resistant modculations. In *GLOBECOM 98*, volume 4, pages 2186–2191. IEEE, 1998.

[14] R. G. Gallager. Capacity and coding for degraded broadcast channels. Probl. Peredach. Inform., 10:3–14, July–September 1974.

[15] P. Bergmans. Coding theorems for broadcast channels with degraded components. *IEEE Transactions on Information Theory*, IT-19(2):197–207, March 1973.

[16] T. Ancheta, Jr. Bounds and techniques for linear source coding. *IEEE Transactions on Information Theory*, 24(2):276, March 1977.

[17] I. Csiszár. Linear codes for sources and source networks: Error exponents. *IEEE Transactions on Information Theory*, 28:585–592, July 1982.

[18] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, 1973.

[19] S. S. Pradhan and K. Ramchandran. Distributed source coding using syndromes (DISCUS) design and construction. In *Proceedings of the Data Compression Conference*, pages 158–167, Snowbird, UT, March 1999. IEEE.

[20] S. S. Pradhan and K. Ramchandran. Distributed source coding: symmetric rates and applications to sensor networks. In *Proceedings of the Data Compression Conference*, pages 363–372, Snowbird, UT, March 2000. IEEE.

[21] S. S. Pradhan and K. Ramchandran. Group-theoretic construction and analysis of generalized coset codes for symmetric / asymmetric distributed source coding. In *Proceedings of the Conference on Information Sciences and Systems*, Princeton, NJ, March 2000.

[22] S. S. Pradhan and K. Ramchandran. Distributed source coding using syndromes (DISCUS): design and construction. *IEEE Transactions on Information Theory*, 49(3):626–643, March 2003.

[23] K. Ramchandran, S. S. Pradhan, and R. Koetter. A constructive framework for distributed source coding with symmetric rates. In *Proceedings of the IEEE International Symposium on Information Theory*, Sorrento, Italy, June 2000.

[24] D. Schonberg, S. S. Pradhan, and K. Ramchandran. LDPC codes can approach the Slepian Wolf bound for general binary sources. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2002. IEEE.

[25] T. Uyematsu. An algebraic construction of codes for Slepian-Wolf source networks. *IEEE Transactions on Information Theory*, 47(7):3082–3088, November 2001.

[26] I. Csiszar and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.

[27] H. Loeliger. On the basic averaging arguments for linear codes. *Communications and Cryptography: Two Sides of One Tapestry*, pages 251–262, 1994.

[28] Q. Zhao and M. Effros. Broadcast system source codes: a new paradigm for data compression. In *Conference Record, Thirty-Third Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 337–341, Pacific Grove, CA, October 1999. IEEE. Invited paper.

[29] Q. Zhao and M. Effros. Lossless and lossy broadcast system source codes: theoretical limits, optimal design, and empirical performance. In *Proceedings of the Data Compression Conference*, pages 63–72, Snowbird, UT, March 2000. IEEE.

[30] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, IT-46(4):1204–1216, July 2000.

[31] S.-Y.R. Li, R.W. Yeung, and N. Cai. Linear network coding. *IEEE Transactions on Information Theory*, IT-49(2):371–381, February 2003.

[32] R. Koetter and M. Médard. Beyond routing: an algebraic approach to network coding. In *Proceedings of INFOCOM 2002*, volume 1, pages 122–130, 2002.

[33] T. Ho, R. Koetter, M. Médard, D. Karger, and M. Effros. The benefits of coding over routing in a randomized setting. In *Proceedings of the IEEE International Symposium on Information Theory*, Yokohama, Japan, June 2003. IEEE. To appear.

[34] S. Jaggi, P. A. Chou, and K. Jain. Low complexity algebraic multicast network codes. In *Proceedings of the IEEE International Symposium on Information Theory*, Yokohama, Japan, June 2003. IEEE. To appear.

[35] P. Sanders, S. Egner, and L. Tolhuizen. Polynomial time algorithms for network information flow. In *Proc. of the 15th ACM Symposium on Parallelism in Algorithms and Architectures*, 2003. To appear.

[36] R. G. Gallager. Low density parity check codes. *IRE Transactions on Information Theory*, IT-8:21–28, January 1962.

[37] G. M. Fel'dman. The Skitovich–Darmois theorem for discrete perdiodic Abelian groups. *Theory of Probability and Its Applications*, 42(4):611–617, 1998.

[38] R. Ahlswede and J. Körner. Source coding with side information and a converse for degraded broadcast channels. *IEEE Transactions on Information Theory*, 21(6):629–637, November 1975.

[39] E. C. van der Meulen. Random coding theorems for the general discrete memoryless broadcast channel. *IEEE Transactions on Information Theory*, IT-21(2):180–190, March 1975.

[40] A. D. Wyner. A theorem on the entropy of certain binary sequences and applications – ii. *IEEE Transactions on Information Theory*, IT-19:772–777, Nov 1973.