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

Michelle Effros, Muriel Médard, Tracey Ho, Siddharth Ray, David Karger, and Ralf Koetter

Abstract. We examine the issue of separation and code design for network data transmission environments. We demonstrate that source-channel separation holds for several canonical network channel models 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 broadcast channels. Moreover, we show that source-channel separation holds for these networks. The linearity of source, channel, and network coding thus blurs the delineation between these codes. This robustness of separation we show to be strongly predicated on the fact that noise and inputs are independent. Finally, we illustrate the fact that design for individual network modules may yield poor results when such modules are concatenated, demonstrating 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.

1. Introduction

The failure of source-channel separation in networks is often considered to be an impediment in applying information theoretic tools in network settings. A simple multiple access channel from [**CT91**] shows how separation can fail. The channel contains $m \geq 2$ transmitters and a single receiver. The receiver's channel output is the integer sum of the transmitters' binary channel inputs. 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.

Key words and phrases. Compression, error correction, multiuser information theory, network coding, routing.

This work was supported in part by NSF grant CCR-0220039, a grant from the Lee Center for Advanced Networking, Hewlett-Packard 008542-008, and University of Illinois subaward #02-194.

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 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 system design.

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

2. Background

The use of random linear transformations in coding receives considerable attention in the literature. For channel coding, Elias [**Eli56**] 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. MacKay [**Mac99**] 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, [**Ahl71, Lia72, CEJW79**]) 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 [**PS95**], 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, [**Gal74, Ber73**], we are likewise unaware of any prior work on linear broadcast channel codes.

On the source coding side, Ancheta [**TA77**] 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 [**Csi82**], Csiszar generalizes linear source coding techniques to allow linear multiple access source codes that achieve the optimal performance derived by Slepian and Wolf [**SW73**]. 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, [**PR99, PR00a, PR00b, PR03, RPK00**] introduce the Distributed Source Coding Using Syndromes (DISCUS) framework. Schonberg et al. [**SPR02**] 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;

[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.

they then demonstrate through simulation that belief propagation decoding works well in practice. Uyematsu proposes a deterministic construction for linear multiple access source codes in [**Uye01**].

Zhao and Effros introduce broadcast system source codes in [**ZE99, ZE00**], presenting design algorithms and performance bounds. 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 [**ACLY00**]. Koetter and Médard give an algebraic framework in [**KM02**]. Reference [**HKM+03**] considers a randomized approach for independent or linearly correlated sources, while [**JCJ03**] and [**SET03**] give polynomial-time deterministic and randomized network code constructions for independent sources.

3. 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 primarily 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 distribution 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 $n \to \infty$. Focusing on linear encoding and typical set decoding allows us to include simple proofs and illuminates the relationships between them.

While we choose to state and prove our results in their simplest form, 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. Erasures propagate as erasures when the output of one channel is fed into another channel.
- We state results for iid source and noise random processes; the results generalize to stationary, ergodic processes.
- We use non-systematic codes in channel coding; the results generalize to systematic codes.
- We use source-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 [**Csi82**].

- 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 those of [**Gal62, Mac99**].
- 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.

4. Single-Transmitter, Single-Receiver Networks

We begin by examining simple forms of some of the prior results described in Section 2. In particular, we give simple new proofs for the linear source and channel coding theorems for single-transmitter, single-receiver networks [**Eli56, TA77, Csi82**]. 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 approach 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.

Let a_n be an $\lceil nR \rceil \times n$ matrix with coefficients in the binary field $\mathbb{F}_2$. To use a_n as a linear source code, we define encoder

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

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

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

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 The error probability for source code a_n is

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

THEOREM 4.1. *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 with coefficients drawn iid Bernoulli$(1/2)$. For any $R > H(U)$, $EP_e(A_n) \to 0$ as $n \to \infty$.*

PROOF. Let $\mathbf{w}^t \in \mathbb{F}_2^n$ be an arbitrary nonzero vector, then

$$\begin{aligned} EP_e^{(n)} &= E\Pr(\text{Error} \wedge U^n \notin A_\epsilon^{(n)}) + E\Pr(\text{Error} \wedge U^n \in A_\epsilon^{(n)}) \\ &\leq \epsilon_n + \sum_{u^n, \hat{u}^n \in A_\epsilon^{(n)}} p(u^n) 1(\hat{\mathbf{u}} \neq \mathbf{u}) \Pr(A_n \hat{\mathbf{u}} = A_n \mathbf{u}) \quad (4.1) \\ &\leq \epsilon_n + \sum_{u^n \in A_\epsilon^{(n)}} p(u^n) 2^{n(H(U)+\epsilon)} \Pr(A_n \mathbf{w} = \mathbf{0}) \quad (4.2) \\ &\leq \epsilon_n + 2^{n(H(U)+\epsilon)} 2^{-\lceil nR \rceil} \quad (4.3) \end{aligned}$$

for some $\epsilon_n \to 0$. Equation (4.1) and the bound on the size of the typical set follow from the AEP. The symmetry represented by the introduction of $\mathbf{w}$ in (4.2) and the bound on the corresponding probability in (4.3) 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. Thus $EP_e^{(n)} \to 0$ as $n \to \infty$ if $\lceil nR \rceil > n(H(U) + \epsilon)$. □

Lemma 4.2 provides a form of converse to Theorem 4.1. While Theorem 4.1 shows that linear source codes are asymptotically optimal, Lemma 4.2 shows that any fixed linear code yields statistically dependent output symbols. An immediate consequence 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 fixed-rate, asymptotically lossless linear codes and variable-rate, truly lossless algorithms like Huffman and arithmetic codes. Variable-rate schemes can achieve lossless performance for any blocklength and precisely achieve the entropy for dyadic distributions. One advantage of fixed-rate codes becomes clear as we move to linear joint source-channel codes later in this section.

LEMMA 4.2. *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 [**Fel98**]. 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 [**KM02, HKM+03**]. 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 that 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.

For any $n \times \lfloor nR \rfloor$ matrix b_n, we can build a linear channel code with encoder $\gamma(v^{\lfloor nR \rfloor}) = b_n \mathbf{v}$. Let X^n denote the channel input and Y^n denote the corrupted channel output. For the erasure channel, $y^n = \mathbf{y}^t \in \{0, 1, E\}^n$, and we 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 $b_n\mathbf{v}$. Decoding to $\hat{V}^{\lfloor nR \rfloor}$ denotes a random decoder output.

THEOREM 4.3. *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. Let* $\{B_n\}_{n=1}^{\infty}$ *describe a sequence of* $n \times \lfloor nR \rfloor$ *linear channel encoders with elements chosen iid Bernoulli*$(1/2)$*. If* $R < 1 - q(1)$*, then* $EP_e(B_n) \to 0$ *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 $\mathcal{E}(z^n) = \{\mathbf{e} \in \mathbb{F}_2^n : e_i = z_i \ \forall i \text{ s.t. } z_i = 0\}$. 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 4.1), giving

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

for some $\epsilon_n \to 0$. 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 4.1 and 4.3 yields an optimal linear source-channel code.

As an alternative to separate design and decoding is joint design and decoding. While we call the resulting code a joint source-channel code for historical reason, 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

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

Here, $(c_n\mathbf{u})_i$ denotes the ith component of vector $c_n\mathbf{u}$.

THEOREM 4.4. *Consider the source of Theorem 4.1 and the channel of Theorem 4.3. Let $\{C_n\}_{n=1}^{\infty}$ describe a sequence of $n \times n$ linear joint source-channel codes 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. We decode Z^n from the received string Y^n. A decoding error occurs if there exists a $\hat{\mathbf{u}} \neq \mathbf{U}$ for which $B_n(\mathbf{U} - \hat{\mathbf{u}}) \in \mathcal{E}(Z^n)$. Thus

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

for some $\epsilon_n \to 0$. 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. Thus $EP_e^{(n)}(C_n) \to 0$ if $H(U) < 1 - q(1) - \epsilon - \epsilon'$. □

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.

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 [**Csi82**], 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.

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 4.5. *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 $n \to \infty$.*

For any $n \times n$ matrix c_n, we can build a joint source-channel code for the additive noise channel with encoder $\zeta(u^n) = c_n\mathbf{u}$ and decoder

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

Theorem 4.6 bounds the expected error probability for a randomly chosen linear code C_n.

THEOREM 4.6. *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 $n \times n$ linear joint source-channel codes 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 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

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

for some $\epsilon_n \to 0$. Here $EP_e^{(n)}(C_n) \to 0$ provided that $H(U) < 1 - H(Z) - 2\epsilon$. □

5. Multiple Access Systems

We next generalize to network systems. We begin with a simple re-derivation of the linear multiple access source codes first studied by Csiszar [**Csi82**].

Let each $\lceil nR_1 \rceil \times n$ matrix $a_{1,n}$ and $\lceil nR_2 \rceil \times n$ matrix $a_{2,n}$ define a two-transmitter, linear multiple access source code with encoders $\alpha_{1,n}(u_1^n) = a_{1,n}\mathbf{u}_1$ and $\alpha_{2,n}(u_2^n) = a_{2,n}\mathbf{u}_2$ and decoder

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

THEOREM 5.1. *Let $(U_{1,1}, U_{2,1}), (U_{1,2}, U_{2,2}), \ldots$ be drawn iid according to distribution $p(u_1, u_2)$ on $(\mathbb{F}_2)^2$. Choose the sequence $\{(A_{1,n}, A_{2,n})\}_{n=1}^\infty$ of rate-(R_1, R_2) linear multiple-access source codes iid uniform. Then for any rates*

$$R_1 > H(U_1|U_2), \quad R_2 > H(U_2|U_1), \quad R_1 + R_2 > H(U_1, U_2),$$

$EP_e(A_{1,n}, A_{2,n}) \to 0$ as $n \to \infty$.

PROOF. An error occurs if either or both of (U_1^n, U_2^n) is decoded in error. Thus,

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

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 if $\lceil nR_1 \rceil > n(H(U_1|U_2)+2\epsilon)$, $\lceil nR_2 \rceil > n(H(U_2|U_1)+2\epsilon)$, and $\lceil nR_1 \rceil + \lceil nR_2 \rceil > n(H(U_1, U_2)+\epsilon)$, then $EP_e(A_{1,n}, A_{2,n}) \to 0$ as n grows without bound. □

We next turn to linear channel coding on the two additive multiple access channels shown in Figure 1. 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.

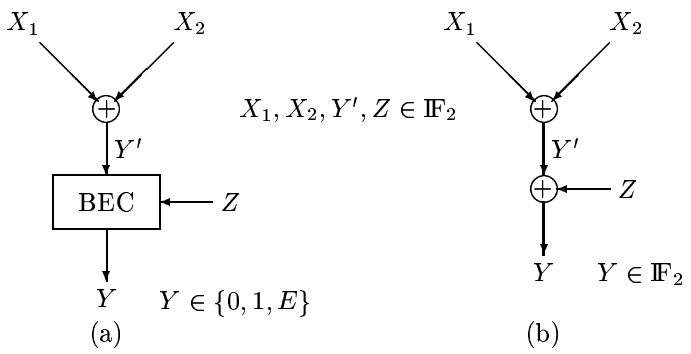

FIGURE 1. Binary additive multiple access channels with (a) erasures and (b) additive noise. In both cases, $Z_1, Z_2, \ldots$ are iid and independent of the channel inputs.

LEMMA 5.2. *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 channel equals the corresponding value of C. Since the multiple access capacity without cooperation cannot exceed the cooperative capacity and the time-sharing solution achieves the cooperative capacity, we have the desired result. □

Since time-sharing between two linear codes yields a linear code, all points in the set of achievable rates are achievable by linear multiple access channel codes.

THEOREM 5.3. *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. Let*

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

where $B_{\lambda n}$ *and* $B_{(1-\lambda)n}$ *are* $\lfloor \lambda n R \rfloor \times \lfloor \lambda n \rfloor$ *and* $(\lfloor nR \rfloor - \lfloor \lambda n R \rfloor) \times (n - \lfloor \lambda n \rfloor)$ *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 5.4. *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

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

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.

We next tackle the issue of source-channel separation.

THEOREM 5.5. *Given the source of Theorem 5.1 and the channel of Theorem 5.3, if $H(U_1, U_2) < 1 - q(1)$, 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 5.1, the Slepian-Wolf region 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 5.3, 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.

Since separation holds for the channel with vector input (X_1, X_2) and scalar output Y, no source pair (U_1, U_2) with $H(U_1, U_2) > 1 - q(1) = I(X_1, X_2; Y)$ can be reliably transmitted across the given communication system. □

THEOREM 5.6. *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. Parallels the proof of Theorem 5.5. □

We next turn to random linear joint source-channel coding.

THEOREM 5.7. *Consider the source of Theorem 5.1 and the channel of Theorem 5.3. Let $\{(C_{1,n}, C_{2,n})\}_{n=1}^{\infty}$ describe a sequence of $n \times n$ linear joint source-channel coding encoders with elements chosen iid Bernoulli$(1/2)$. 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. 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

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

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 5.8. *Consider the source of Theorem 5.1 and the channel of Theorem 5.4. Let $\{(C_{1,n}, C_{2,n})\}_{n=1}^{\infty}$ describe a sequence of linear joint source-channel codes 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. 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 $C_{1,n}(\mathbf{U}_1 - \hat{\mathbf{u}}_1) + C_{2,n}(\mathbf{U}_2 - \hat{\mathbf{u}}_2) \in \mathcal{F}(Z^n)$. From here, the proof parallels the proof of Theorem 5.7. In this case, $|\mathcal{F}(Z^n)| \leq 2^{n(H(Z)+\epsilon)} - 1$, giving

$$
\begin{aligned}
&EP_e^{(n)}(C_{1,n}, C_{2,n}) \\
&\quad \leq \quad 2\epsilon_n + \sum_{(u_1^n,u_2^n)\in A_\epsilon^{(n)}(p)} \sum_{z^n \in A_\epsilon^{(n)}(q)} p(u_1^n, u_2^n) q(z^n) \left[2^{n(H(U_1|U_2)+\epsilon)} 2^{-n} 2^{n(H(Z)+\epsilon)} \right. \\
&\qquad \left. + 2^{n(H(U_2|U_1)+\epsilon)} 2^{-n} 2^{n(H(Z)+\epsilon)} + 2^{n(H(U_1,U_2)+\epsilon)} 2^{-n} 2^{n(H(Z)+\epsilon)} \right]
\end{aligned}
$$

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

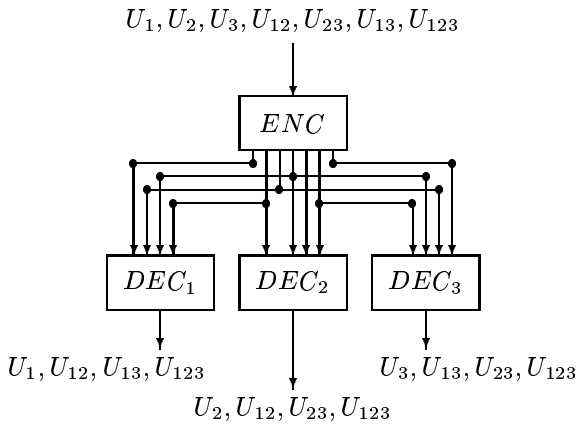

FIGURE 2. A broadcast system source code with three receivers.

6. Broadcast Systems

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 [**ZE99, ZE00**], we focus on the three-receiver system of Figure 2. 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 6.1 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, and so on. We again use typical set decoding.

THEOREM 6.1. *Let samples of source vector* $(U_1, U_2, U_3, U_{12}, U_{23}, U_{13}, U_{123})$ *be 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*

$$\begin{aligned}
(nR)_s &\geq H(U_s | U_{S_1 - s}) \quad \forall \quad s \subseteq S_1 = \{1, 12, 13, 123\}, s \neq \phi \\
(nR)_s &\geq H(U_s | U_{S_2 - s}) \quad \forall \quad s \subseteq S_2 = \{2, 12, 23, 123\}, s \neq \phi \\
(nR)_s &\geq H(U_s | U_{S_3 - s}) \quad \forall \quad s \subseteq S_3 = \{3, 13, 23, 123\}, s \neq \phi
\end{aligned}$$

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

PROOF. We break encoder matrix A_n into a collection of $\lceil nR_a \rceil \times n$ submatrices, $a \in \{1, 2, 3, 12, 23, 13, 123\}$, such that

$$A_n^t = \left[A_{1,n}^t A_{2,n}^t A_{3,n}^t A_{12,n}^t A_{23,n}^t A_{13,n}^t A_{123,n}^t\right].$$

Let $EP_e(A_{1*,n})$ denote the expected probability of decoding in error at receiver 1. The arguments for receivers 2 and 3 are similar, and the code error probability is bounded by the sum of the individual decoder error probabilities. An error occurs at receiver 1 if any subset of the desired sources is decoded in error. Thus,

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

for some $\epsilon_n \to 0$. □

We next consider two erasure broadcast channel models. In each, 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 second model, 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 [**CT91**, 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] Since we consider discrete channels, the degraded broadcast channel converses of [**AK75**] or of [**vdM75**], which allows no or partial common information, are applicable. Note that the elegant and simple converse for degraded BSC broadcast channels of [**Wyn73**], which relies on properties of binary sequences, might be readily extended to our model, albeit without the generality of [**AK75, vdM75**].

Lemma 6.2 proves time-sharing to be optimal for broadcast coding over the given family of channels. Theorem 6.3, is then immediate by the previous linearity of time-sharing argument.

LEMMA 6.2. *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 channel noise is independent of the channel input. The joint distribution $q(z_1, z_2)$ may be any distribution with*

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

the given marginals. The capacity region for sending independent information to the two receivers is described by

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

If independent rates (R_1, R_2) are achievable and $R_0 < R_2$, then $(R_1', R_2', R_{12}') = (R_1, R_2 - R_0, R_0)$ is achievable with common information rate R_{12}' and independent information rates R_1' and R_2'.

PROOF. By [**CT91**, 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 $p(w)p(x|w)p(y_1|x)p(y_2|y_1)$. Auxiliary random variable W is has alphabet size 2, and $p(y_2|y_1)$ is derived from the physically degraded channel model. By a 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

$$\begin{aligned} R_1 &\leq I(X; Y_1|W) = (1 - q_1(1))H(\beta) \\ R_2 &\leq I(W; Y_2) = (1 - q_2(1))(1 - H(\beta)). \end{aligned}$$

Varying $H(\beta)$ from 0 to 1 gives the independent coding result. The common information result comes from [**CT91**, Theorem14.6.4]. □

THEOREM 6.3. *Consider the channel from Lemma 6.2. Let $\{B_n\}_{n=1}^{\infty}$ describe a sequence of linear channel codes for the broadcast channel, where*

$$B_n = \begin{bmatrix} B_{\lambda n} & \mathbf{0} \\ \mathbf{0} & B_{(1-\lambda)n} \end{bmatrix}$$

Each $B_{\lambda n}$ has 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 $n \to \infty$.

For the additive noise broadcast channel model, linear codes can do at least as well as the time-sharing bound, but that bound is not the optimal solution [**CT91**].

7. 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. In the case of the single-transmitter, single-receiver channel, source-channel coding holds in general, but fails for linear codes. In the case of the additive multiple access channel with additive noise, separation fails more generally, as shown next. The same phenomena may be observed in erasure channels.

THEOREM 7.1. *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. 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. 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 $\Pr(Z = 1 | X_1 = i, X_1 = j) = q_{ij} = 1 - \bar{q}_{ij}$. For the multiple access capacity, let $p_i = \Pr(X_i = 1) = 1 - \bar{p}_i$. Then

$$\begin{aligned} \max_{P_1, P_2} I(X_1, X_2; Y) &= \max_{p_1, p_2} [H(\bar{p}_1\bar{p}_2\bar{q}_{00} + \bar{p}_1 p_2 q_{01} + p_1\bar{p}_2 q_{10} + p_1 p_2 \bar{q}_{11}) \\ &\quad -\bar{p}_1\bar{p}_2 H(q_{00}) - \bar{p}_1 p_2 H(q_{01}) - p_1\bar{p}_2 H(q_{10}) - p_1 p_2 H(q_{11})]. \end{aligned}$$

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

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

The two equations are not equal in general. For example, let $q_{00} = q_{11} = 0$ while $q_{01} = q_{10} = 1/2$. Then $\max_{P_1 P_2} I(X_1, X_2; Y) = 0.311$ while $\max_{P_{12}} I(X_1, X_2; Y) = 1$. (The maxima occur at $p_1 = p_2 = 1/2$ and $p_{00} = p_{11} = 1/2$.) 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 $R_1 + R_2 \leq .311$.) □

8. 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 [**ACLY00**]. 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

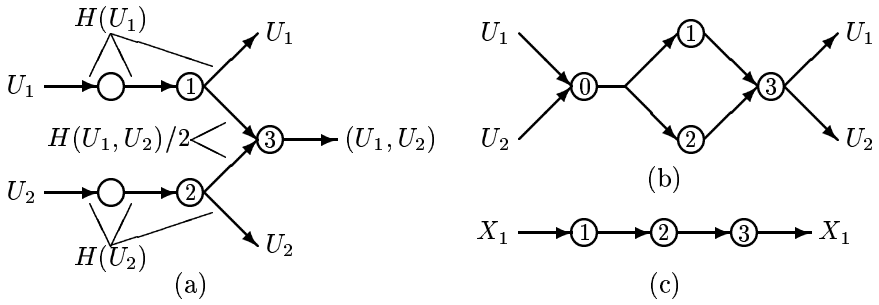

FIGURE 3. Networks for which (a) separation of source and network coding and (b) separation of channel and network coding fail. (c) A network for which decoding at intermediate nodes is required for optimal coding.

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 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 8.1. The network of Figure 3(a) comprises two transmitters and three receivers. Receiver nodes 1, 2, and 3 wish to receive U_1, U_2 and (U_1, U_2), respectively. 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 8.2. In the network shown in Figure 3(b), 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 channel between nodes 1 and 2 and node 3 is a multiple access channel without interference. The network coding approach requires labeling each link with 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 6.3. 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 $1 - p_1$ 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 $1 - p_1 p_2$ 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 8.2 serves as a reminder that while canonical network elements can be strung together to achieve codes for more complicated networks, the resulting

solutions are not optimal in general. While Example 8.2 demonstrates that failure to decode at intermediate nodes of the network can yield performance superior to that achieved by decoding at intermediate nodes, Example 8.3 teaches the opposite lesson.

EXAMPLE 8.3. In the channel of Figure 3(c), 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 $1-(q_1(1)+q_2(1))$. Decoding at node 2 yields maximal achievable rate $\min\{1-q_1(1), 1-q_2(1)\} > 1-(q_1(1)+q_2(1))$.

The failure of separation in Examples 8.1 and 8.2 and the contrasting lessons regarding decoding at intermediate nodes demonstrated by Examples 8.2 and 8.3 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

[ACLY00] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, *Network information flow*, IEEE Transactions on Information Theory **IT-46** (2000), no. 4, 1204–1216.

[Ahl71] R. Ahlswede, *Multi-way communication channels*, Proc. 2nd. Int. Symp. Information Theory (Tsahkadsor, Armenian S.S.R.) (Prague), Publishing House of the Hungarian Academy of Sciences, 1971, pp. 23–52.

[AK75] R. Ahlswede and J. Körner, *Source coding with side information and a converse for degraded broadcast channels*, IEEE Transactions on Information Theory **21** (1975), no. 6, 629–637.

[Ber73] P. Bergmans, *Coding theorems for broadcast channels with degraded components*, IEEE Transactions on Information Theory **IT-19** (1973), no. 2, 197–207.

[CEJW79] S.-C. Chang and Jr. E. J. Weldon, *Coding for t-user multiple access channels*, IEEE Transactions on Information Theory **IT-25** (1979), no. 6, 684–691.

[Csi82] I. Csiszár, *Linear codes for sources and source networks: Error exponents*, IEEE Transactions on Information Theory **28** (1982), 585–592.

[CT91] T. M. Cover and J. A. Thomas, *Elements of information theory*, Wiley, 1991.

[Eli56] P. Elias, *Coding for two noisy channels*, Third London Symposium on Information Theory (London), Academic Press, 1956, pp. 61–74.

[Fel98] G. M. Fel'dman, *The Skitovich–Darmois theorem for discrete perdiodic Abelian groups*, Theory of Probability and Its Applications **42** (1998), no. 4, 611–617.

[Gal62] R. G. Gallager, *Low density parity check codes*, IRE Transactions on Information Theory **IT-8** (1962), 21–28.

[Gal74] ———, *Capacity and coding for degraded broadcast channels*, *Probl. Peredach. Inform.* **10** (1974), 3–14.

[HKM+03] T. Ho, R. Koetter, M. Médard, D. Karger, and M. Effros, *The benefits of coding over routing in a randomized setting*, Proceedings of the IEEE International Symposium on Information Theory (Yokohama, Japan), IEEE, June 2003, To appear.

[JCJ03] S. Jaggi, P. A. Chou, and K. Jain, *Low complexity algebraic multicast network codes*, Proceedings of the IEEE International Symposium on Information Theory (Yokohama, Japan), IEEE, June 2003, To appear.

[KM02] R. Koetter and M. Médard, *Beyond routing: an algebraic approach to network coding*, Proceedings of INFOCOM 2002, vol. 1, 2002, pp. 122–130.

[Lia72] H. Liao, *Multiple access channels*, Ph. D. Dissertation, Department of Electrical Engineering, University of Hawaii, Honolulu, 1972.

[Mac99] D. J. C. MacKay, *Good error-correcting codes based on very sparse matrices*, IEEE Transactions on Information Theory **45** (1999), no. 2, 399–431.

[PR99] S. S. Pradhan and K. Ramchandran, *Distributed source coding using syndromes (DISCUS) design and construction*, Proceedings of the Data Compression Conference (Snowbird, UT), IEEE, March 1999, pp. 158–167.

[PR00a] ———, *Distributed source coding: symmetric rates and applications to sensor networks*, Proceedings of the Data Compression Conference (Snowbird, UT), IEEE, March 2000, pp. 363–372.

[PR00b] ———, *Group-theoretic construction and analysis of generalized coset codes for symmetric / asymmetric distributed source coding*, Proceedings of the Conference on Information Sciences and Systems (Princeton, NJ), March 2000.

[PR03] ———, *Distributed source coding using syndromes (DISCUS): design and construction*, IEEE Transactions on Information Theory **49** (2003), no. 3, 626–643.

[PS95] G. Poltyrev and J. Snyders, *Linear codes for the sum mod-2 multiple-access channel with restricted access*, IEEE Transactions on Information Theory **41** (1995), no. 3, 794–799.

[RPK00] K. Ramchandran, S. S. Pradhan, and R. Koetter, *A constructive framework for distributed source coding with symmetric rates*, Proceedings of the IEEE International Symposium on Information Theory (Sorrento, Italy), June 2000.

[SET03] P. Sanders, S. Egner, and L. Tolhuizen, *Polynomial time algorithms for network information flow*, Proc. of the 15th ACM Symposium on Parallelism in Algorithms and Architectures, 2003, To appear.

[SPR02] D. Schonberg, S. S. Pradhan, and K. Ramchandran, *LDPC codes can approach the Slepian Wolf bound for general binary sources*, Proceedings of the Allerton Conference on Communication, Control, and Computing (Monticello, IL), IEEE, October 2002.

[SW73] D. Slepian and J. K. Wolf, *Noiseless coding of correlated information sources*, IEEE Transactions on Information Theory **IT-19** (1973), 471–480.

[TA77] Jr. T. Anceta, *Bounds and techniques for linear source coding*, IEEE Transactions on Information Theory **24** (1977), no. 2.

[Uye01] T. Uyematsu, *An algebraic construction of codes for Slepian-Wolf source networks*, IEEE Transactions on Information Theory **47** (2001), no. 7, 3082–3088.

[vdM75] E. C. van der Meulen, *Random coding theorems for the general discrete memoryless broadcast channel*, IEEE Transactions on Information Theory **IT-21** (1975), no. 2, 180–190.

[Wyn73] A. D. Wyner, *A theorem on the entropy of certain binary sequences and applications – ii*, IEEE Transactions on Information Theory **IT-19** (1973), 772–777.

[ZE99] Q. Zhao and M. Effros, *Broadcast system source codes: a new paradigm for data compression*, Conference Record, Thirty-Third Asilomar Conference on Signals, Systems and Computers (Pacific Grove, CA), vol. 1, IEEE, October 1999, Invited paper, pp. 337–341.

[ZE00] ———, *Lossless and lossy broadcast system source codes: theoretical limits, optimal design, and empirical performance*, Proceedings of the Data Compression Conference (Snowbird, UT), IEEE, March 2000, pp. 63–72.

M. Effros: Department of Electrical Engineering, 136-93, California Institute of Technology, Pasadena, CA 91125.

E-mail address: effros@caltech.edu

M. Médard, T. Ho, and S. Ray: Laboratory for Information and Decision Systems (LIDS), Massachusetts Institute of Technology, Cambridge, MA 02139

E-mail address: medard@mit.edu, trace@mit.edu, sray@mit.edu

D. Karger: Laboratory for Computer Science (LCS), Massachusetts Institute of Technology, Cambridge, MA 02139

E-mail address: karger@lcs.mit.edu

R. Koetter: Coordinated Science Laboratory, University of Illinois at Urbana-Champain, Urbana, IL 61801

E-mail address: koetter@uiuc.edu