

Training Overhead for Decoding Random Linear Network Codes in Wireless Networks

Maximilian Riemensberger, Yalin Evren Sagduyu, Michael L. Honig,
and Wolfgang Utschick

Abstract

We consider multicast communications from a single source to multiple destinations through a wireless network with unreliable links. Random linear network coding achieves the min-cut flow rate; however, additional overhead is needed for end-to-end error protection and to communicate the network coding matrix to each destination. We present a *joint coding and training* scheme in which training bits are appended to each source packet, and the channel code is applied across both the training and data. This scheme allows each destination to decode jointly the network coding matrix along with the data without knowledge of the network topology. It also balances the reliability of communicating the network coding matrices with the reliability of data detection. The throughput for this scheme, accounting for overhead, is characterized as a function of the packet size, channel properties (error and erasure statistics), number of independent messages, and field size. We also compare the performance with that obtained by individual channel coding of training and data. Numerical results are presented for a grid network that illustrate the reduction in throughput due to overhead.

Index Terms

Channel coding, decoding error, network coding, overhead, random linear codes, training.

This work has been supported in part by the U.S. Army Research Office grant W911NF-06-1-0339 and the DARPA IT-MANET program grant W911NF-07-1-0028.

Part of this work has been presented in IEEE MILCOM 2008.

M. Riemensberger and W. Utschick are with the Associate Institute for Signal Processing, Technische Universität München, 80333 Munich, Germany (e-mail: riemensberger@tum.de; utschick@tum.de).

Y. E. Sagduyu and M. L. Honig are with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208, USA (e-mail: y-sagduyu@northwestern.edu; mh@eecs.northwestern.edu).

I. INTRODUCTION

Network coding has significant advantages relative to more conventional store-and-forward-based routing solutions [1]. For multicast communications linear network coding achieves the min-cut capacity, and eliminates the need for combinatorial-based routing optimization [2]. Different low-complexity algorithms have been proposed for deterministic and random network coding [3]–[5].

Here we consider *random* linear network coding for a wireless network. That is, each internal node transmits a random linear combination of the input packets onto the outgoing links [4]. Two challenges that arise are: (1) Wireless channels may be unreliable due to the presence of fading and interference; and (2) The network topology may change due to mobile users. To address the first issue, end-to-end channel coding can be applied to correct residual errors and erasures that are introduced on each link. The second issue implies that the network coding matrix, which maps the source symbols to received symbols, can change. Hence to enable decoding at all destinations, the network coding matrices may have to be communicated frequently to the destinations (along with associated error protection). Techniques that address both issues require additional overhead, which decreases the overall network throughput.

Communicating the network coding matrix to each destination can be accomplished by appending the combining coefficients to the data as part of the header at the source and updating them at each internal node [4]. Changes in the network topology and channel impairments may introduce errors in the packet overhead, which would likely prevent successful packet decoding. An issue then is how much end-to-end error protection should be used to communicate the network coding matrix to each destination.

To address the preceding issues, we present a *joint training and coding (JTC)* scheme in which the source appends training bits to the data, and then applies the end-to-end channel code across both the training and data. We focus on the scenario in which a single source is multicasting to multiple destinations, in which case the optimal number of training bits is simply the number of simultaneous messages to be transmitted by the source. This scheme allows each destination to decode jointly the network coding matrix along with the data. It therefore balances the reliability of communicating the network coding matrices with the reliability of data detection. The JTC scheme is compared with an *individual training and coding (ITC)* scheme in which the channel

code is applied separately to the training and data. In contrast to JTC, with ITC the amount of training must be explicitly optimized, i.e., adjusted to balance the reliabilities associated with training and data detection.¹

For both training schemes the training sequence is added at the source node. The intermediate nodes do not add additional overhead, but simply forward linear combinations of the input packets. Hence neither scheme assumes knowledge of the network topology at any node. Furthermore, any topology changes, or changes in network coding coefficients are automatically taken into account when estimating the network coding matrix at the destination.² The source must only ensure that sufficient training is provided to communicate the network coding matrix reliably.³

We consider a general network topology and analyze the effect of the training overhead on reliability (probability of decoding error) and throughput for JTC and ITC. (The throughput objective assumes a retransmission protocol for error control [8].) The links are modeled as discrete memoryless channels, where the error and/or erasure probabilities account for wireless impairments, such as interference and fading. We evaluate the throughput, accounting for overhead, as a function of the length of the training sequence (for ITC), packet length, network size (min-cut), number of packets transmitted N , error/erasure statistics, and field size. Our results can be used to optimize N as a function of the min-cut. A key assumption is that the elements of the network coding matrices are independent and identically distributed (*i.i.d.*). Although in general the elements of the network coding matrices are correlated, our analysis gives a *worst-case* estimate on the training overhead needed to achieve a particular throughput.

We illustrate our results for a 3×3 grid network with a single source transmitting multiple packets to a single destination. The analytical results for throughput and overhead (assuming *i.i.d.* network coding matrix elements) are compared with simulated results and show that the analytical results are accurate. The results also show that as the min-cut becomes large, the throughput becomes limited by overhead, and that JTC performs significantly better than ITC.

¹The training approach to communicating the network coding matrix presented here is motivated by similar types of training schemes for channel estimation (e.g., see [6]). A key difference here, however, is that the elements of the network coding matrix are in a finite field.

²These training schemes are therefore robust with respect to manipulations of coding coefficients by selfish or malicious agents [7].

³The destinations must know the number of training bits and their placement within the packet.

An alternative approach to the training method presented here are the *noncoherent* network coding schemes presented in [9], [10], which implicitly detect and/or correct errors in both the header and data. Here our emphasis is on determining the minimum overhead needed to compensate for unreliable links in addition to possible topology changes within the network. (The associated overhead is not evaluated in [9], [10].)

The paper is organized as follows. Section II presents the system model and describes the training-based detection of random network code. In Section III, we introduce the ITC and JTC schemes. Section IV formulates the problem of maximizing throughput as function of overhead, and also presents numerical results for the grid network. Conclusions are presented in Section V.

II. SYSTEM MODEL

We represent a wireless network as an equivalent acyclic/delay-free graph, where the links designate the desired connectivity (e.g., corresponding to a particular transmission schedule). Link rates are then determined by the particular assumptions concerning the physical layer and the configuration of multiple access and broadcast nodes, treating interference from other nodes as noise. A single source node in the network wishes to deliver the same messages to a set of destinations \mathcal{D} . Random linear network coding over a finite field \mathbb{F}_q with q elements is assumed. While packets are relayed through the network, some symbols in the packets may be changed to another symbol (i.e., an error occurs) or they may be erased. We define an extended set⁴ $\mathbb{F}_{q,e} = \mathbb{F}_q \cup \{e\}$ that includes the erasure symbol “ e ” to mark the positions of the erased symbols in each packet.

The source messages are random row vectors $\mathbf{s}^T \in \mathbb{F}_q^{1 \times D}$ of D symbols, where $\{\cdot\}^T$ denotes matrix transpose. The elements of all messages are independent and uniformly distributed random variables. The encoder \mathcal{E} maps these messages onto packets (row vectors) $\mathbf{x}^T = \mathcal{E}(\mathbf{s}^T) \in \mathbb{F}_q^{1 \times K}$ of K symbols. In each packet interval, the source transmits N independent multicast messages $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_N]^T \in \mathbb{F}_q^{N \times D}$, encoded as packets⁵ $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T = \mathcal{E}(\mathbf{S}) \in \mathbb{F}_q^{N \times K}$, over the network. The number of messages N must be less than the network cut capacity M of the

⁴This set is not a field, since addition and multiplication with erasures are not defined.

⁵ \mathcal{E} is applied row-wise to \mathbf{S} .

equivalent graph.⁶

With random linear network coding each matrix of source packets \mathbf{X} experiences a linear transformation $\mathbf{G}_d \in \mathbb{F}_q^{M_d \times N}$ from the source to destination $d \in \mathcal{D}$, where \mathbf{G}_d is the random coding matrix and M_d is the number of incoming links at d . We assume that the entries of the coding matrices are uniformly distributed and independent (both within each matrix and across destinations).

In the absence of channel errors and erasures, the received packets at destination d are given by the rows of $\mathbf{V}_d = \mathbf{G}_d \mathbf{X} \in \mathbb{F}_q^{M_d \times K}$. The effects of wireless impairments on each link are modeled as erasures and errors. Namely, although coding is applied on each link, fading and interference can lead to residual errors and erasures, which accumulate from source to destinations. This effect is modelled by a discrete memoryless channel \mathcal{H}_d between the source and each destination d . The matrix of received packets including errors and erasures is denoted as

$$\mathbf{Y}_d = \mathcal{H}_d(\mathbf{V}_d) = \mathcal{H}_d(\mathbf{G}_d \mathbf{X}), \quad (1)$$

where $\mathbf{Y}_d \in \mathbb{F}_{q,e}^{M_d \times K}$. The channel \mathcal{H}_d is specified by the conditional distribution $P_{\mathbf{Y}_d | \mathbf{V}_d}$ of the output \mathbf{Y}_d given the input \mathbf{V}_d . The destinations detect $\hat{\mathbf{S}}_d$ by applying the channel decoder \mathcal{E}_d^{-1} to the received packets, i.e., $\hat{\mathbf{S}}_d = \mathcal{E}_d^{-1}(\mathbf{Y}_d)$. The system model for destination d is shown in Fig. 1.

For the analysis that follows, we assume that the errors and erasures across all entries of the output matrix \mathbf{Y}_d are *i.i.d.* This serves as an end-to-end channel model between the source and destinations and does not account for statistical dependencies, which may be introduced by the network topology. (Alternatively, we can assume an interleaver and deinterleaver at the source and destinations, respectively.) Furthermore, we will assume that the elements of the network matrices are *i.i.d.* We therefore ignore the correlations that are likely to be introduced by a particular network topology. This facilitates tractability, and also corresponds to a practical scenario in which the source and destinations do not know the network topology. The *i.i.d.* assumption then leads to a *worst-cast* estimate of overhead needed to achieve a target throughput.

⁶For a wireless network with omnidirectional transmissions the cut capacity can be determined by introducing virtual nodes as in [11]. The graph we consider would be the result of combining the augmented network graph with the corresponding physical layer and multiple access/broadcast assumptions.

In principle, it may be possible to exploit the correlations ignored here to reduce the training overhead. However, that would require a substantial increase in overhead to learn the statistics, and also more complex decoders that take the correlations into account. Furthermore, numerical results to be presented in Section IV-B show that for the grid network considered, the analytical results based on the *i.i.d.* assumption accurately predict simulated performance.

III. ENCODING AND DECODING

For the network coding model described in Section II, there are three sources of errors:

- (i) Errors or erasures of data symbols;
- (ii) Errors or erasures of network coding matrix elements (overhead symbols);
- (iii) Rank-deficient network coding matrices (due to random network coding).

An error and/or erasure correcting code can protect the data against the first source of errors. Training is used to convey the network coding information reliably to the destinations. Both are components of the encoder \mathcal{E} . The probability that the coding matrix is rank-deficient depends on the field size, which must be the same for all network coding operations, and the dimensions of the matrix. Increasing the number of independent transmitted messages N increases the network data rate, but also increases the probability that at least one of the network coding matrices is rank-deficient. As the coding field size q increases, the probability that the network coding matrix is full rank increases, but at the expense of increasing overhead. In general, optimizing q requires centralized information about the full network topology, which may not be available.

A. Training Viewed as a Linear Code

Referring to (1), we assume a training phase, where the data (training) matrix \mathbf{X} is known to the destination, followed by transmission of the unknown data \mathbf{X} . Of course, detecting the unknown data given the coding matrix \mathbf{G}_d is a standard decoding problem. Likewise, in the training phase detecting the coding matrix \mathbf{G}_d given the training matrix \mathbf{X} is again a standard decoding problem, but with the roles of the transmitted symbols and channel being swapped. Namely, the transmitted symbol matrix \mathbf{X} is now the code, and the channel matrix \mathbf{G}_d is the “data”. In the training phase the source therefore implicitly transmits the generator matrix of the code through the network, which codes the data (\mathbf{G}_d) by performing the network coding operations.

The probability of incorrectly decoding the coding matrix depends on the choice of the training matrix, the distribution of channel errors and erasures, and the statistical properties of the coding matrix. Due to the equivalence of training and coding, choosing a “good” training matrix is equivalent to choosing a “good” code, which has a low complexity decoding algorithm. For general channels this problem can be difficult, so that for purposes of performance analysis, in what follows we use error probability bounds that are obtained by random coding arguments.

If the error control code for the data and the training are the same, then each destination requires only one decoder. This is also the case when the transmitter codes over both the training and data, as described in the next section.

B. Encoding

For reliability in the presence of errors or erasures, we encode the data using a linear $(K_D, D)_q$ code with generator matrix $\mathbf{F}_D \in \mathbb{F}_q^{D \times K_D}$ and append the known training symbols $\mathbf{X}_T \in \mathbb{F}_q^{N \times K_T}$, which can be selected as the generator matrix of a linear $(K_T, N)_q$ code, where $K_T + K_D = K$. This scheme is referred to as *individual training and channel coding (ITC)*. The encoding map $\mathcal{E}^{(i)}$ is defined by

$$\mathbf{X} = \mathcal{E}^{(i)}(\mathbf{S}) = \begin{bmatrix} \mathbf{X}_T & \mathbf{S}\mathbf{F}_D \end{bmatrix}, \quad (2)$$

which can be rewritten as

$$\mathbf{X} = \begin{bmatrix} \mathbf{I} & \mathbf{S} \end{bmatrix} \mathbf{F} = \begin{bmatrix} \mathbf{I} & \mathbf{S} \end{bmatrix} \begin{bmatrix} \mathbf{X}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_D \end{bmatrix}, \quad (3)$$

where \mathbf{F} is the coding matrix, which maps the extended source messages $[\mathbf{I} \ \mathbf{S}]$ onto the coded packets, and \mathbf{I} is the $N \times N$ identity matrix. The channel coding matrix \mathbf{F} for the extended message therefore has the specific form shown in (3).

ITC therefore consists of appending the identity matrix (uncoded training overhead) to the uncoded data, and separately encoding them with channel codes \mathbf{X}_T and \mathbf{F}_D . (Such a scheme is suggested in [4], [5] with reliable links (no errors/erasures) and represents a conventional scheme for forwarding the coding coefficients in a distinct packet header.)

The redundancy introduced for training (i.e., training length K_T) influences the reliability of data detection, and can be optimized. Namely, if K_T is too short, the coding matrix is unreliable,

whereas if K_T is too large, then not enough symbols are reserved for the data.⁷

Referring to (2), there is no need to constrain the generator matrix \mathbf{F} to have the block diagonal structure (3). Instead, we can apply *joint training and coding (JTC)* with the same encoding procedure

$$\mathbf{X} = \mathcal{E}^{(j)}(\mathbf{S}) = \begin{bmatrix} \mathbf{I} & \mathbf{S} \end{bmatrix} \mathbf{F}, \quad (4)$$

where \mathbf{F} is a *full* generator matrix that represents a $(K, N + D)_q$ linear code.

In contrast with ITC, the overhead and data are not separated in the transmitted packets \mathbf{x}_n^T , rather both are jointly encoded into the entire packet. Since only one channel code is used for both training and data, there is no explicit trade-off between the reliability of the training bits versus the data bits. Namely, the channel code \mathbf{F} implicitly balances the training and data protection.

In general, joint encoding of training and data achieves a lower probability of error than separate encoding for any choice of training length K_T . The reason is that coding the overhead and data jointly with one long code is better than coding them separately with two shorter codes.

C. Decoding

For the encoding procedures (2) and (4), the output at destination d is given by

$$\mathbf{Y}_d = \mathcal{H}_d(\mathbf{G}_d \mathbf{X}) = \mathcal{H}_d(\mathbf{G}_d [\mathbf{I} \ \mathbf{S}] \mathbf{F}) = \mathcal{H}_d([\mathbf{G}_d \ \mathbf{S}'_d] \mathbf{F}), \quad (5)$$

where we define the network-coded (NC) source message $\mathbf{S}'_d = \mathbf{G}_d \mathbf{S} \in \mathbb{F}_q^{M_d \times D}$.

The optimal (Maximum Likelihood (ML)) estimate of the data and coding matrices, given the observations \mathbf{Y}_d , is

$$(\hat{\mathbf{S}}_d, \hat{\mathbf{G}}_d) = \arg \max_{\mathbf{S}, \mathbf{G}} P_{\mathbf{Y}_d | \mathbf{V}_d}(\mathbf{Y}_d | \mathbf{G} [\mathbf{I} \ \mathbf{S}] \mathbf{F}), \quad (6)$$

where \mathbf{V}_d represents the inputs to the error or erasure channel model for the network. This joint detection problem is prohibitively complex to solve in practice. Furthermore, individual detection of the packets (rows) is not optimal. We therefore consider the following suboptimal detector, which also facilitates performance analysis.

⁷This is analogous to the optimization of training overhead for MIMO channel estimation considered in [6]. A key difference is that here we wish to *detect* the system matrix with acceptable probability of error, as opposed to *estimating* a complex-valued channel matrix. Also, the JTC scheme to be described is not directly applicable to the MIMO estimation problem.

First, the ML estimate for the NC source message and the coding matrix are computed as

$$[\hat{\mathbf{G}}_d \hat{\mathbf{S}}'_d] = \arg \max_{[\mathbf{G} \mathbf{S}']} P_{\mathbf{Y}_d | \mathbf{V}_d}(\mathbf{Y}_d | [\mathbf{G} \mathbf{S}'] \mathbf{F}). \quad (7)$$

Since errors and erasures are independent, the conditional probability $P_{\mathbf{Y}_d | \mathbf{V}_d}$ can be factored into M_d terms, one for each row of \mathbf{Y}_d . The decoder (7) therefore reduces to a decoder for each row m ,

$$[\hat{\mathbf{G}}_d \hat{\mathbf{S}}'_d]_m^T = \arg \max_{[\mathbf{G} \mathbf{S}']_m^T} P_{[\mathbf{Y}_d]_m^T | [\mathbf{V}_d]_m^T}([\mathbf{Y}_d]_m^T | [\mathbf{G} \mathbf{S}']_m^T \mathbf{F}), \quad (8)$$

where $[\cdot]_m$ denotes the m th row of the argument. Then the set \mathcal{S}_d of all valid source messages is found by solving a system of linear equations

$$\mathcal{S}_d = \{\mathbf{S} \in \mathbb{F}_q^{N \times D} : \hat{\mathbf{G}}_d \mathbf{S} = \hat{\mathbf{S}}'_d\}. \quad (9)$$

Although (8) is ML optimal with respect to determining \mathbf{G}_d and \mathbf{S}'_d , it is suboptimal with respect to detecting the actual data \mathbf{S} . The suboptimal decoder facilitates analysis since detection of a single row is the standard ML decoding problem for the linear code \mathbf{F} . The second step requires solving an overdetermined system of linear equations for $\hat{\mathbf{S}}'_d$, for which there could be no solution, a unique solution, or multiple solutions.

Theorem 1: If the set \mathcal{S}_d defined by (7) and (9) is nonempty, i.e., if there exists at least one $\hat{\mathbf{S}}'_d$ such that $\hat{\mathbf{G}}_d \hat{\mathbf{S}}'_d = \hat{\mathbf{S}}'_d$, then each element of \mathcal{S}_d satisfies the ML criterion (6). Otherwise, the estimate $(\hat{\mathbf{S}}'_d, \hat{\mathbf{G}}_d)$ given by (7) is not the ML estimate.

The proof is given in the appendix.

The suboptimal detector can therefore be viewed as a relaxed version of the ML detector. Furthermore, Theorem 1 implies that the destinations can identify which estimates differ from the ML estimate.

The suboptimal decoding procedure \mathcal{E}_d^{-1} (for both JTC and ITC) is given by the following algorithm, where $\mathbf{M}_{\mathcal{N}}$ denotes a matrix that contains only the rows of the matrix \mathbf{M} with indices in \mathcal{N} :

- 1: Solve (8) for each row.
- 2: Select m rows from $[\hat{\mathbf{G}}_d \hat{\mathbf{S}}'_d]$ according to a predefined rule and let \mathcal{N} contain the indices of these rows.
- 3: Determine solution set $\mathcal{S}_d = \{\mathbf{S} : [\hat{\mathbf{S}}'_d]_{\mathcal{N}} = [\hat{\mathbf{G}}_d]_{\mathcal{N}} \mathbf{S}\}$.
- 4: If $|\mathcal{S}_d| = 1$, return the estimate $\hat{\mathbf{S}}'_d \in \mathcal{S}_d$, otherwise, return a decoding error.

The selection of rows in Step 2 should depend on the specific channel and coding model. (Theorem 1 holds for detecting the subset of rows in \mathcal{N} , where the ML detector is based on $[\mathbf{Y}_d]_{\mathcal{N}}$ instead of \mathbf{Y}_d .) Of course, at least N rows must be selected to obtain a solution in Step 3. The selection of a subset \mathcal{N} of the decoded rows from Step 1 may improve performance if the decoder is able to identify incorrectly decoded rows, e.g., by comparing the subspaces spanned by different subsets of decoded rows, and dropping those rows in subsequent steps. In particular, for erasure channels the destination can identify rows, which have been decoded correctly, and can choose only those rows to decode \mathbf{S} .

D. Decoding Error Probability at a Single Destination

We assume that each destination d decodes the messages independently from other destinations, and give lower bounds on the associated probability of successful decoding. For JTC there exists a linear code such that destination d can decode any particular row of \mathbf{Y}_d independently with probability

$$Q_d^{(j)} \geq 1 - 2^{-K E_d(\frac{N+D}{K})}, \quad (10)$$

where E_d is the Gallager random coding error exponent [12, Ch. 5] for the end-to-end channel map \mathcal{H}_d . Since errors and erasures are assumed to be *i.i.d.* across all rows, the same code can be used for each row.

For ITC two linear codes must be independently decoded for each row. In that case there exist two codes such that the probability of successfully decoding a particular row of \mathbf{Y}_d satisfies

$$Q_d^{(i)} \geq \max_{\substack{K_T, K_D \\ K_T + K_D = K}} \left(1 - 2^{-K_T E_d(\frac{N}{K_T})} \right) \left(1 - 2^{-K_D E_d(\frac{D}{K_D})} \right). \quad (11)$$

The preceding bounds can be refined for particular coding schemes, e.g., bounds for erasure channels with MDS codes are presented in [13].

Assuming that $m \geq N$ rows of the estimates $[\hat{\mathbf{G}}_d \hat{\mathbf{S}}'_d]$ have been selected by the destination, the probability that the corresponding rows of \mathbf{G}_d have rank N is denoted as $\Phi(m, N)$. Using standard counting arguments [14], we have

$$\Phi(m, N) = \prod_{i=0}^{N-1} (1 - q^{m-i}), \quad (12)$$

where q denotes the field size.⁸

To derive bounds on the probability that the data are correctly decoded at a single destination, we consider two different ways for selecting the m rows of $[\hat{\mathbf{G}}_d \hat{\mathbf{S}}'_d]$, where $N \leq m \leq M_d$: (1) Select the m rows randomly; and (2) Select rows that have been correctly decoded. The first method gives a lower bound on the decoding probability that can be achieved by the suboptimal decoder, and the second (genie-aided) method gives an upper bound.

Theorem 2: The probability of correctly decoding the data at destination d using the suboptimal decoder \mathcal{E}_d^{-1} satisfies

$$\max_{N \leq m \leq M_d} Q_d^m \Phi(m, N) \leq P_d \leq \sum_{m=N}^{M_d} Q_d^m (1 - Q_d)^{M_d - m} \Phi(m, N), \quad (13)$$

where Q_d is $Q_d^{(j)}$ or $Q_d^{(i)}$ for ITC and JTC, respectively.

The proof is given in the appendix.

Choosing m for the lower bound trades off $\Phi(m, N)$ (the probability that the associated rows of \mathbf{G}_d have full rank) with $(Q_d)^m$ (the probability that the decoded rows are error free). Note that the upper bound can be achieved if the destinations are able to detect incorrectly decoded rows. That could occur in practice with erasure channels, or by adding redundancy (e.g., a cyclic-redundancy check) to enable this detection. For the special case of erasure channels with MDS codes the genie bound is tight. When combined with the lower bound on Q_d for JTC (10) or ITC (11), this gives a tight bound on the achievable performance of the suboptimal decoding algorithm.

IV. NETWORK CODING OVERHEAD

In this section, we use the preceding error probability bounds to determine the necessary overhead for communicating the network coding matrix to each destination. We do this analytically for a large network, and also present numerical results for finite K and N . For fixed packet length K and field size q , the source has two parameters to choose, namely, the number of data symbols per packet D and the number of independent packets per transmission N .

The performance criterion of interest is network throughput Λ , which is defined as the average amount of data that can be successfully communicated to all destinations in a single network

⁸With correlated matrix elements an alternative bound on $\Phi(m, N)$ in terms of the field size and the number of random coding coefficients is given in [4].

use.⁹ It depends on the network data rate, i.e., the number of data messages N the source transmits in one network use, and also on the average delay, i.e., the expected number of network uses (including retransmissions) needed to ensure that all destinations receive all packets correctly. We assume that a multi-receiver version of an automatic repeat request (ARQ) scheme is used to ensure reliable communications. Namely, packets are retransmitted until they are successfully received at all destinations. These ARQ schemes require one bit (ACK/NACK) feedback per transmitted packet to inform the source whether or not the transmitted packets have been successfully decoded.

The maximum achievable throughput is given by

$$\Lambda = \max_{D, N: N \leq M} \left[N \frac{D}{K} \min_{d \in \mathcal{D}} P_d \right], \quad (14)$$

which is equivalent to a max-flow min-cut bound, and can be achieved with a retransmission scheme that combines coding with queue management [8]. Here the source maximizes the objective over the number of data symbols per packet D as well as the number of independent source packets N . In addition, for the individual scheme the training length K_T is also optimized. We will use Λ as our performance objective.¹⁰

We define the total packet overhead due to both channel and network coding as $O_{\text{tot}} = \frac{K-D}{K}$. The total packet overhead is allocated across network coding and channel coding. With ITC K_T and K_D symbols are allocated for training and data, respectively, so that the network coding overhead is $O_T = \frac{K_T}{K}$ and the channel coding overhead is $O_D = \frac{K_D-D}{K}$. With JTC the network coding overhead is given by $O_T = \frac{N}{N+D}$ (fraction of appended training relative to number of source symbols per packet). Here the overhead due to network coding is not easily distinguished from the overhead for channel coding, although the total overhead can again be defined by O_{tot} .

A. Asymptotic Throughput and Overhead

Analytical results for the optimal throughput and corresponding overhead cannot be easily obtained with a finite packet and network size, since the optimization of the throughput with

⁹The definition of “network use” in general depends on assumptions concerning whether or not some nodes are half-duplex. In that situation a network use may require multiple time slots.

¹⁰Note that for the multicast application considered in this paper, ARQ schemes may require additional overhead, which should specify the intended destinations and possibly the coding coefficients for each retransmission. We do not consider this overhead in this paper.

respect to N and D is an integer program and has no closed form solution. We therefore consider asymptotic performance for a large network in which both $M \rightarrow \infty$ and $K \rightarrow \infty$. To relate these limits, let $M = M(K)$ be a function of K and define the relative min-cut $\bar{M} = \lim_{K \rightarrow \infty} \frac{M}{K} \in [0, \infty]$. It will be convenient to assume that $M_d = M$ for all d and that all destinations have the same statistics for the channel and network coding matrix. In this section we therefore drop the destination subscripts. The asymptotic growth rates for N and D for both schemes and K_T for the ITC scheme that maximize the throughput are defined as $\bar{N} = \lim_{K \rightarrow \infty} \frac{N}{K}$, $\bar{D} = \lim_{K \rightarrow \infty} \frac{D}{K}$, and $\bar{K}_T = \lim_{K \rightarrow \infty} \frac{K_T}{K}$.

Theorem 3: For a channel with capacity C the maximum throughput per symbol for both ITC and JTC is given by

$$\lim_{K \rightarrow \infty} \frac{\Lambda}{K} = \bar{N} \bar{D}, \quad (15)$$

where $\bar{N} = \min(\bar{M}, c/2)$, $\bar{D} = c - \bar{N}$, and $c = \frac{C}{\log_2 q}$ for both schemes and $\bar{K}_T = \frac{\bar{N}}{c}$ for the ITC scheme.

The proof is given in the appendix.

For the asymptotic analysis we do not need to specify a particular method for selecting rows in the suboptimal detection and decoding algorithm, since the lower bound (with randomly chosen rows) and the genie-aided bound (with correctly decoded rows) coincide asymptotically. Furthermore, the ITC and JTC have the same asymptotic performance since the codeword length approaches infinity. Hence the separation of training and coding does not decrease the throughput, provided that the the training length \bar{K}_T is also optimized.

Fig. 2 shows the asymptotic throughput given in Theorem 3 as a function of \bar{M} for a class of end-to-end q -ary symmetric error channels with different symbol error probabilities p and corresponding normalized capacities c . If the destinations know their network coding matrices, then the throughput grows linearly with \bar{M} . If, however, each destination $d \in \mathcal{D}$ needs to detect \mathbf{G}_d , then for $\bar{M} < c/2$ the asymptotic throughput is given by $\bar{M}(c - \bar{M})$, which achieves a maximum at $\bar{M} = c/2$. Hence for $\bar{M} < c/2$ the throughput is *network limited* in the sense that it increases with \bar{M} (min-cut capacity).

For $\bar{M} > c/2$, $\bar{N} = c/2$ is optimal and the throughput does not increase with \bar{M} but remains constant. This is due to the associated overhead with increasing \bar{N} , so that for $\bar{M} > c/2$ the throughput is *overhead limited*. If $\bar{N} = \bar{M}$ is chosen in this regime, the throughput decreases

and reaches zero at $\bar{M} = c$, at which point the overhead occupies the entire packet. The min-cut value where the throughput is maximized decreases as the error probability p increases. That is because increasing p increases the channel coding overhead, and also reduces the maximum number of parallel transmissions (\bar{N}) in the overhead-limited regime.

The asymptotic overhead is easily obtained by substituting the asymptotic values for \bar{N} , \bar{D} , and \bar{K}_T in Theorem 3, as stated in the next corollary.

Corollary 1: The optimized total overhead and the associated data protection and training overhead for ITC and JTC are given asymptotically by

$$\lim_{K \rightarrow \infty} O_{\text{tot}} = 1 - c + \bar{N}, \quad (16)$$

$$\lim_{K \rightarrow \infty} O_D = (1 - c) + (1 - c^{-1})\bar{N}, \quad (17)$$

$$\lim_{K \rightarrow \infty} O_T = c^{-1}\bar{N}. \quad (18)$$

Corollary 1 shows that the optimized overhead increases linearly with \bar{N} . Using the expression for \bar{N} in Theorem 3, we see that as \bar{M} increases from zero, the overhead increases until $\bar{M} = c/2$, at which point the overhead remains constant. That is, when the throughput is network limited, increasing \bar{M} increases the number of messages \bar{N} , which requires more training overhead. The overhead is maximized for $\bar{M} \geq c/2$, in which case as \bar{N} increases, the additional overhead outweighs the increase in packet transmission rate.

B. Numerical Evaluation of Throughput and Overhead

The grid network model considered in [4] and shown in Fig. 3 is used to illustrate the achievable throughput and packet overhead needed with network coding.¹¹ The source is located at the bottom left corner ((1, 1) position), and results are presented for the single destination at the upper right corner ((3, 3) position). In general there can be multiple destinations in the grid; however, here we consider only the worst-case destination with the largest distance to the source, which limits the throughput performance. The intermediate nodes perform random linear network coding of the inputs. Each link carries $\frac{M}{2}$ packets simultaneously so that the min-cut capacity of the network is M . Results are shown for $M = 2, 4, \dots, 60$.

¹¹Similar analytical results for a combination network with symmetric erasure channels on all links are presented in [13].

All links are modeled as q -ary symmetric error channels with symbol error probability δ . Note that the end-to-end uncoded symbol error probability p is needed to compute the random coding bounds, which in turn are needed to compute the throughput given in (14). Since it appears to be quite difficult to compute p analytically, we determine p numerically by generating network coding matrices corresponding to different sets of randomly chosen combining coefficients, simulating link errors (introduced independently across all links), and averaging the corresponding end-to-end error rates over different realizations. This is done for different values of M . Note that increasing M increases the end-to-end error probability, since the number of simultaneously transmitted packets on each link increases.

The estimated p is used to determine the training and code parameters, i.e., N and D for JTC, and K_T , N , and D for ITC, based on our analytical results. That is, we assume an end-to-end q -ary symmetric symbol error channel with symbol error probability p and *i.i.d.* error events and network coding matrix elements, and optimize the combined lower bound with random selection of decoded rows (13) and the random coding bounds on the probability of decoding a single row (11) and (10) with respect to the parameters. These values are used to obtain analytical estimates for the throughput. We then simulate ITC and JTC in the grid network with MDS codes and bounded distance decoding using the optimized parameters.¹² The simulated results therefore account for all statistical dependencies that are introduced by the grid network and random network coding. The simulation results (and estimates for p) are averaged over 100 network code realizations and 100 packet network uses for each code ($100 \times N$ encoded source messages of size K).

Fig. 4 shows throughput (Fig. 4(a)) and packet overhead (Fig. 4(b)) for the 3×3 grid network as a function of the min-cut capacity M with $\delta = 0.001$, $q = 2^8$, and $K = 60$. The throughputs given by the analysis (JTC/ITC lower) and by simulating the grid network (JTC/ITC sim.) are shown. Also shown is the throughput assuming that the destination knows the network coding matrix (REF), so that the overhead due to training is not included.

Fig. 4(a) shows that the analytical curves based on the estimated end-to-end error probability accurately approximate the performance of JTC and ITC for small to medium values of M , where the assumptions of statistical independence do not hold. The throughput is limited by

¹²Bounded distance decoding is suboptimal (not ML), so that the random coding bounds may not be achieved.

the packet overhead and tends to decrease in the overhead-limited regime, since the end-to-end error probability also increases as M increases. The approximation becomes less accurate in this regime since the gap between the performance of bounded distance and ML decoding increases. The throughput gain of JTC relative to ITC is significant for large M , and is accurately predicted by the analytical curves. The shape of the throughput versus min-cut is predicted by the asymptotic results. Namely, the throughput increases as a parabola for small M , but then decreases in the overhead-limited regime due to the increasing end-to-end error probability with M .

Fig. 4(b) shows the overhead for JTC and ITC based on the estimated end-to-end error probability.¹³ The overhead due to network coding increases linearly with M (network-limited region) until it reaches approximately 0.5 as M becomes large (overhead-limited region). This behavior is also predicted by the asymptotic results in Sec. IV-A. In addition, the results show that the total packet overhead and network coding overhead are larger for ITC than for JTC in the network-limited regime, giving the associated increase in JTC throughput.

V. CONCLUSIONS

Methods for communicating network coding matrices from a source to multiple destinations have been presented based on the use of training sequences. With these methods the destinations do not require knowledge of the network topology. The source only needs to know the end-to-end error probability and how many packets should be transmitted simultaneously (namely, the maximal rank of the coding matrix).¹⁴ In that sense training schemes are robust with respect to topology changes (e.g., due to mobility). Also, the “outer” channel code(s) used to code the training and data at the source (as opposed to the inner codes for each link) provides additional robustness with respect to residual errors, which may accumulate across unreliable links.¹⁵

With ITC optimization of the training balances the trade-off between the reliability of communicating the code and the packet overhead. This trade-off is not explicitly present with JTC since

¹³Note that the packet overhead is the same for the analytical and simulated results since the same code parameters N , D , and K_T are chosen.

¹⁴Both parameters could be estimated at the destinations by using an adaptive rate control scheme with ACK/NACK feedback. The source could successively increase or decrease N based on the feedback.

¹⁵In practice, the outer channel code may be combined with a CRC check.

the channel code automatically balances the reliability of training and data bits. JTC performs significantly better than ITC (although asymptotically there is no difference due to the longer code length) and should not be much more complex to decode. Furthermore, decoding a single code of fixed length, which depends on the packet length and field size, is likely to be preferred over decoding two shorter variable-length codes.

For a given network topology and packet length, the throughput can be maximized over the network parameters (e.g., number of packets simultaneously transmitted by the sources). Our results indicate regions in which the throughput is limited by training overhead, as opposed to the min-cut capacity of the network. Results for the grid network show that as the min-cut M becomes large, the optimized network coding overhead is approximately half of the uncoded packet size, and that this limits the overall throughput.

Our system model is based on the assumption that the coding coefficients are independent and randomly chosen for each packet transmission. This is primarily for analytical convenience, and also models the scenario where knowledge of the network topology is not readily available or cannot be exploited due to the associated increase in overhead and complexity. A possibility for future work is to consider an alternative model, where the coding matrices over successive transmissions are statistically dependent. The overhead could then be reduced by applying joint source-channel coding, which compresses the sequence of coding matrices. Finally, other types of retransmission schemes could also be considered, which further improve throughput at the expense of additional overhead.

APPENDIX A

PROOFS OF THEOREMS 1, 2, AND 3.

Proof of Theorem 1: From the definition of the auxiliary data matrix $\mathbf{S}'_d = \mathbf{G}_d \mathbf{S}$, the conditional probability mass function $P_{\mathbf{S}'_d | \mathbf{G}_d \mathbf{S}}(\mathbf{S}' | \mathbf{G} \mathbf{S}) = 1$ when \mathbf{S}' satisfies the definition. Introducing the auxiliary matrix in the optimal ML detector (6), we obtain

$$\begin{aligned}
 (\hat{\mathbf{S}}_d, \hat{\mathbf{G}}_d) &= \\
 &= \arg \max_{\mathbf{S}, \mathbf{G}} \sum_{\mathbf{S}'} P_{\mathbf{S}'_d | \mathbf{G}_d \mathbf{S}}(\mathbf{S}' | \mathbf{G} \mathbf{S}) P_{\mathbf{Y}_d | \mathbf{V}_d}(\mathbf{Y}_d | [\mathbf{G} \ \mathbf{S}'] \mathbf{F}) \\
 &= \arg \max_{\mathbf{S}, \mathbf{G}} \max_{\mathbf{S}'} P_{\mathbf{S}'_d | \mathbf{G}_d \mathbf{S}}(\mathbf{S}' | \mathbf{G} \mathbf{S}) P_{\mathbf{Y}_d | \mathbf{V}_d}(\mathbf{Y}_d | [\mathbf{G} \ \mathbf{S}'] \mathbf{F})
 \end{aligned} \tag{19}$$

From the definition of \mathbf{S}'_d , $\text{span}(\mathbf{S}') \subset \text{span}(\mathbf{G})$ is necessary for a nonzero likelihood of every triple $(\mathbf{S}, \mathbf{G}, \mathbf{S}')$. Furthermore, if a pair $(\mathbf{G}, \mathbf{S}')$ satisfies this condition, then the likelihood of all data \mathbf{S} such that $\mathbf{S}' = \mathbf{G}\mathbf{S}$ is constant, since the data are uniformly distributed and \mathbf{V}_d is independent of \mathbf{S} given \mathbf{G}_d and \mathbf{S}'_d . Therefore, the detector

$$\begin{aligned} [\hat{\mathbf{G}}_d \hat{\mathbf{S}}'_d] &= \arg \max_{[\mathbf{G} \ \mathbf{S}']} P_{\mathbf{Y}_d | \mathbf{V}_d}(\mathbf{Y}_d | [\mathbf{G} \ \mathbf{S}'] \mathbf{F}) \\ &\text{s.t. } \text{span}(\mathbf{S}') \subset \text{span}(\mathbf{G}), \end{aligned} \quad (20)$$

and where $\hat{\mathbf{S}}'_d$ satisfies $\hat{\mathbf{S}}'_d = \hat{\mathbf{G}}_d \hat{\mathbf{S}}'_d$ is equivalent to ML detection. The suboptimal detector (7) and (9) is obtained by relaxing the subspace constraint ($\text{span}(\hat{\mathbf{S}}'_d) \subset \text{span}(\hat{\mathbf{G}}_d)$). Consequently, if the decoding result of the suboptimal detector satisfies the constraint, it is also optimal with respect to the ML criterion. ■

Proof of Theorem 2: For the lower bound, suppose the decoder chooses a random set \mathcal{N} of m rows with $N \leq m \leq M_d$. The probability that these rows are correctly decoded is $(Q_d)^m$. Furthermore, the probability that the part of \mathbf{G}_d corresponding to these rows has rank N given that they have been decoded correctly is $\Phi(m, N)$. Finally, the probability of correct decoding (if both prior conditions are satisfied) is one. Multiplying these three probabilities for each m and optimizing over the number of randomly selected rows yields the desired lower bound.

For the upper bound, suppose that a genie tells the decoder after the first step of the decoding algorithm which rows have been decoded correctly. The decoder uses only the correctly decoded rows in succeeding decoding steps. If these rows have rank N , then the data $\hat{\mathbf{S}}'_d$ are calculated from the decoded rows and must be error-free. If the rank is smaller than N , then unique decoding is not possible and a decoding error is returned. Since the rows are decoded independently, this probability is given by

$$\binom{M_d}{m} (Q_d)^m (1 - Q_d)^{M_d - m} \Phi(m, N) \quad (21)$$

for each m . The probability that correct decoding of the data is possible follows directly from the disjoint union over all $m \geq N$ of the events that m rows are correctly decoded and have rank N , and is given by the sum of (21) over $m \geq N$. ■

Proof of Theorem 3: Since the optimal throughput of ITC can not exceed the optimal throughput of JTC, we prove the asymptotic lower bound for ITC which is also a lower bound on JTC. Finally, we derive an upper bound on the throughput of JTC that coincides with said lower bound.

For ITC and large K the throughput is asymptotically lower bounded by

$$\lim_{K \rightarrow \infty} \max_{\substack{N, D, m \\ 0 \leq \bar{K}_T \leq 1 \\ N \leq m \leq M}} \frac{ND}{K^2} \left(1 - 2^{-\bar{K}_T K E(\frac{N}{\bar{K}_T K})}\right)^m \times \quad (22)$$

$$\left(1 - 2^{-(1-\bar{K}_T) K E(\frac{N}{(1-\bar{K}_T) K})}\right)^m \Phi(m, N),$$

where we fix $K_T = \bar{K}_T K$ and $K_D = (1 - \bar{K}_T) K$. We introduce the constraints $\frac{N}{K_T K} \leq c - \mu$ and $\frac{D}{(1-\bar{K}_T) K} \leq c - \mu$ for any $\mu > 0$ such that $E(\frac{N}{K_T K}) \geq \nu$ and $E(\frac{D}{(1-\bar{K}_T) K}) \geq \nu$ for some $\nu > 0$ (see [12]). We restrict the maximization with respect to N and m with the constraint $0 \leq N \leq M - \log_2 K$ and choose $m = N + \log_2 K$. This yields a new lower bound

$$\lim_{K \rightarrow \infty} \max_{\substack{N, D \geq 0, 0 \leq \bar{K}_T \leq 1 \\ N \leq M - \log_2 K \\ N \leq \bar{K}_T K (c - \mu) \\ D \leq (1 - \bar{K}_T) K (c - \mu)}} \left\{ \frac{ND}{K^2} \Phi(N + \log_2 K, N) \times \quad (23)$$

$$\left[\left(1 - 2^{-\nu \bar{K}_T K}\right) \left(1 - 2^{-\nu (1 - \bar{K}_T) K}\right) \right]^{N + \log_2 K} \right\}.$$

L'Hospital rule implies that $(1 - 2^{-\nu \bar{K}_T K})^{N + \log_2 K}$ and $(1 - 2^{-\nu (1 - \bar{K}_T) K})^{N + \log_2 K}$ approach 1 as $K \rightarrow \infty$ for $\nu > 0$ and $0 < \bar{K}_T < 1$. Additionally, Eq. (12) implies $\lim_{K \rightarrow \infty} \Phi(N + \log_2 K, N) = 1$. For $\bar{K}_T = 0$ and $\bar{K}_T = 1$ we have $\bar{N} = 0$ and $\bar{D} = 0$, respectively, where those limits are irrelevant. Consequently, taking the limit and optimizing the bound with respect to μ yields the desired result.

The throughput of JTC is asymptotically upper bounded by

$$\lim_{K \rightarrow \infty} \max_{\substack{N, D \geq 0 \\ N \leq M}} \frac{ND}{K^2} \sum_{m=N}^M \binom{M}{m} Q^m (1 - Q)^{M-m} \Phi(m, N) \quad (24)$$

with row decoding probability Q . From the sphere packing lower bound on the decoding error probability [12], it follows that Q for any row approaches 0 if $\lim_{K \rightarrow \infty} \frac{N+D}{K} > c$. Otherwise, the sum in (24) is upper bounded by 1 since it is the tail of a weighted binomial distribution with weights $\Phi(m, N) \leq 1$. Therefore, taking the limit in (24) yields

$$\max_{\substack{\bar{D} \geq 0 \\ 0 \leq \bar{N} \leq \bar{M}}} \bar{N} \bar{D} \quad \text{s.t.} \quad \bar{N} + \bar{D} \leq c, \quad (25)$$

which coincides with the lower bound. ■

REFERENCES

- [1] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network Information Flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Apr. 2000.
- [2] S.-Y. R. Li, R. W. Yeung, and N. Cai, "Linear Network Coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.
- [3] S. Jaggi, P. Sanders, P. A. Chou, M. Effros, S. Egner, K. Jain, and L. M. G. M. Tolhuizen, "Polynomial Time Algorithms for Multicast Network Code Construction," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1973–1982, Jun. 2005.
- [4] T. Ho, M. Medard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A Random Linear Network Coding Approach to Multicast," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4413–4430, Oct. 2006.
- [5] P. A. Chou and Y. Wu, "Network Coding for the Internet and Wireless Networks," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 77–85, Sep. 2007.
- [6] B. Hassibi and B. M. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
- [7] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, M. Medard, and M. Effros, "Resilient Network Coding in the Presence of Byzantine Adversaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2596–2603, Jun. 2008.
- [8] Y. E. Sagduyu and A. Ephremides, "On Network Coding for Stable Multicast Communication," in *Proc. IEEE MILCOM*, Oct. 2007.
- [9] Z. Zhang, "Linear Network Error Correction Codes in Packet Networks," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 209–218, Jan. 2008.
- [10] R. Koetter and F. R. Kschischang, "Coding for Errors and Erasures in Random Network Coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3579–3591, Aug. 2008.
- [11] Y. Wu, P. A. Chou, Q. Zhang, K. Jain, W. Zhu, and S.-Y. Kung, "Network Planning in Wireless Ad Hoc Networks: A Cross-Layer Approach," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 136–150, Jan. 2005.
- [12] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley and Sons, 1968.
- [13] M. Riemensberger, Y. E. Sagduyu, M. L. Honig, and W. Utschick, "Training Overhead for Decoding Random Linear Network Codes," in *Proc. IEEE MILCOM*, San Diego, CA, USA, Nov. 2008.
- [14] T. Migler, K. E. Morrison, and M. Ogle, "How Much Does a Matrix of Rank k Weigh?" *Mathematics Magazine*, vol. 79, no. 4, pp. 262–271, Oct. 2006.

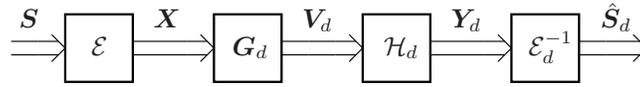


Figure 1. System model for destination d .

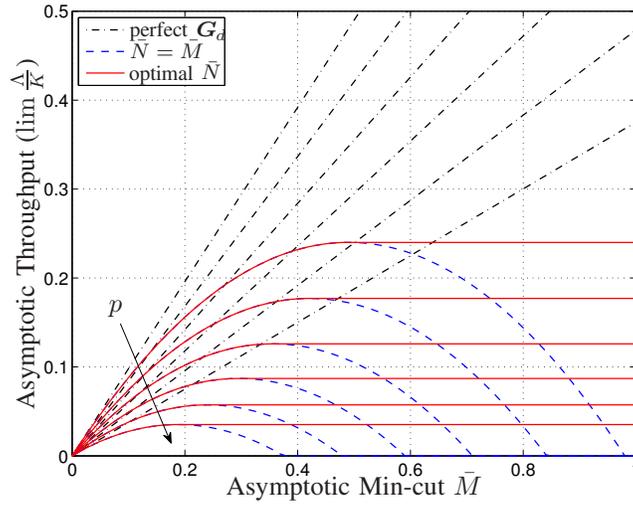


Figure 2. Asymptotic throughput $\lim_{K \rightarrow \infty} \frac{\lambda}{K}$ versus \bar{M} for a q -ary symmetric error channel with symbol error probability p ; Parameters $q = 2^8$ and $p = [0.01, 0.1, 0.2, 0.3, 0.4, 0.5]$.

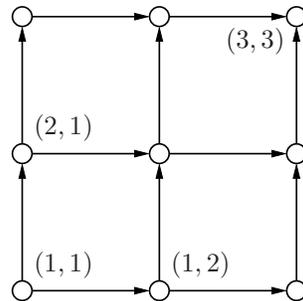
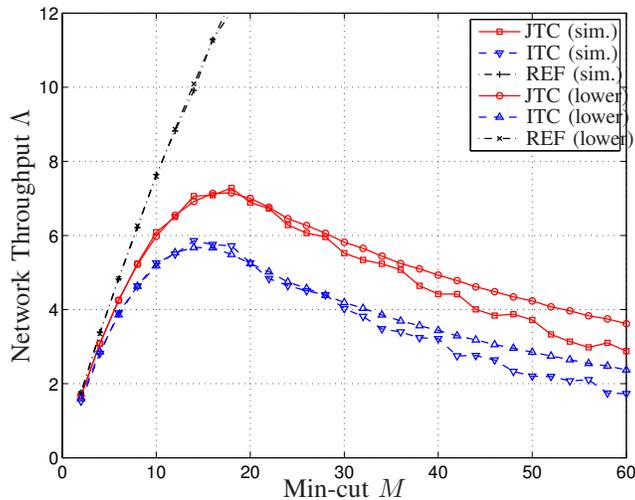
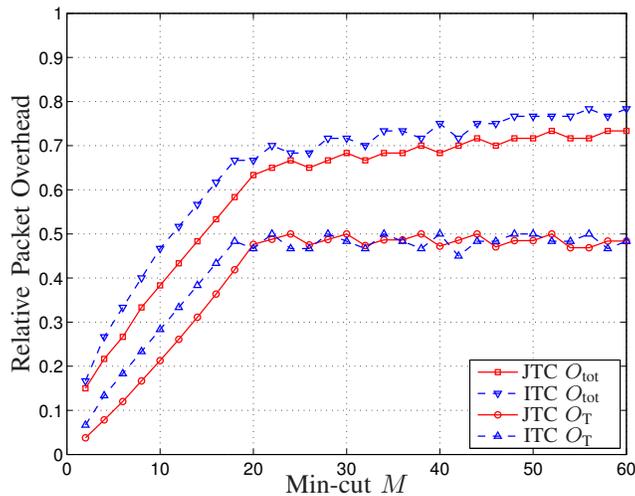


Figure 3. 3×3 Grid Network. Source at bottom left position (1, 1) and destination at top right position (3, 3).



(a) Network Throughput



(b) Packet Overhead

Figure 4. (a) Network throughput Λ vs. cut-set capacity M with ITC and JTC. Simulated and analytical results are shown with random row selection for the 3×3 grid network. Also shown is the throughput assuming the destination knows the network coding matrix; (b) Total overhead O_{tot} and network coding overhead O_{T} for JTC and ITC with the grid network; Curves are shown for optimized D and N . Parameters are $q = 2^8$, $K = 60$, and the link error probability $\delta = 0.001$.