SCHOOL OF ENGINEERING - STI
SIGNAL PROCESSING LABORATORY - LTS4
Hyunggon Park, Nikolaos Thomos and Pascal Frossard

CH-1015 LAUSANNE

Telephone: +4121 6935655
Telefax: +4121 6937600
e-mail: hyunggon.park@ewha.ac.kr, nikolaos.thomos@epfl.ch, pascal.frossard@epfl.ch

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

NETWORK CODING OF CORRELATED DATA WITH APPROXIMATE DECODING

Hyunggon Park, Nikolaos Thomos and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)

Signal Processing Laboratory LTS4 Technical Report

TR-LTS-2009-013

December, 2010

Part of this work has been submitted to the IEEE Transactions on Communications.

This work has been supported by the Swiss National Science Foundation grants 200021-118230 and PZ00P2-121906.

Network Coding of Correlated Data with Approximate Decoding

Hyunggon Park, Nikolaos Thomos and Pascal Frossard
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory (LTS4), Lausanne, 1015 - Switzerland.
hyunggon.park@ewha.ac.kr, {nikolaos.thomos, pascal.frossard}@epfl.ch
Fax: +41 21 693 7600, Phone: +41 21 693 5655

Abstract

This paper considers a framework where data from correlated sources are transmitted with help of network coding in ad hoc network topologies. The correlated data are encoded independently at sensors and network coding is employed in the network nodes for improved data delivery performance. In such settings, we focus on the problem of reconstructing the sources at decoder when perfect decoding is not possible due to losses or capacity limits. We first show that the correlation between the data sources can be exploited at decoder by the design of a novel approximate decoding scheme. We analyze the influence of the network coding parameters and in particular the size of finite coding fields on the decoding performance. We further determine the optimal field size that maximizes the expected decoding performance as a trade-off between information loss incurred by limiting the resolution of the source data and the error probability in the reconstructed data. Moreover, we show that the approximate decoding performance improves when the accuracy of the correlation estimation increases even with simple approximate decoding techniques. We illustrate our algorithms in sensor networks and distributed video transmitting applications. In both cases, the experimental results confirm the validity of our analysis and demonstrate the benefits of our solution for delivery of correlated data.

Index Terms

Network coding, approximate decoding, correlated data, distributed transmission, ad hoc networks.

I. Introduction

The rapid deployment of sensor networks has motivated a plethora of researches that study the design of low complexity sensing strategies and efficient solutions for information delivery. Since the coordination among sensors is often difficult to achieve, the transmission of information from the sensors has typically to be performed in a distributed manner on ad-hoc or overlay mesh network topologies. Network coding [1] has been recently proposed as a method to build efficient distributed delivery algorithms in networks with path and source diversity. It is based on a paradigm where the network nodes are allowed to perform basic processing operations on information streams. The network nodes can combine information packets and transmit the resulting data to the next network nodes. When the decoder receives enough data, it can recover the original source information by performing inverse operations (e.g., with Gaussian elimination). Such a strategy permits to improve the throughput of the system and to approach better the max-flow min-cut limit of networks [2], [3].

These advantages motivate the deployment of network coding in various scenarios where the network diversity is significant (e.g., [4]–[9]). Many of these solutions are based on random linear network coding (RLNC) [10] that permits to implement distributed solutions with low communication costs. It leads to small performance degradation compared to globally optimized solutions (e.g., solutions based on linear network coding [3]) when the size of the coding filed is large enough [8]. RLNC thus certainly represents an interesting solution for the deployment of practical systems where it can work in conjunction with data dissemination protocols such as gossiping algorithms [8]. The resulting systems are robust against link failures, do not require reconciliation between the network node, and can significantly improve the performance of storing and forwarding data. Most of research so far has however focused either on theoretical aspects of network coding such as achievable capacity and coding gain, or on its practical aspects such as robustness when the number of innovative packets is sufficient for perfect decoding. However, it generally does consider the problematic cases where the clients receive an insufficient number of innovative packets for perfect decoding, due to losses or timing constraints for example. This is the main problem addressed in this paper.

We propose to use network coding for the transmission of correlated data sources that are discretized and independently encoded at the sensors. The information streams are delivered with help of RLNC in lossy ad hoc networks. When an insufficient number of symbols at decoder prevent exact data recovery, we design a novel low complexity *approximate decoding* algorithm that uses the correlation information for signal reconstruction. We first show the benefits of using data correlation such as *external* correlation (e.g., data measured from different locations in sensor networks) or *intrinsic* redundancy (e.g., images in a video sequence) for decoding. The information about correlation typically provides additional constraints in the decoding process, such that well-known approaches for matrix inversion (e.g., Gaussian elimination) can be efficiently even in the case where the decoding system is a priori underdetermined. We show analytically that the use of correlation leads to an improved

data recovery, or equivalently, that the proposed approximate decoding solution results in improved decoding performance. Then, we analyze the impact of the accuracy of the correlation information on the decoding performance, where higher precision leads to better performance in the approximate decoding. We further analyze the influence of the choice of the Galois Field (GF) size in the coding operations on the performance of the approximate decoding. We demonstrate that the field size should be selected by considering the tradeoff between resolution in representing the source and approximate decoding performance. Specifically, when the GF size increases, the quantization error of the source data decreases while the decoding error probability increases with the GF size. We show that there is an optimal value for the GF size when the approximate decoding is enabled at the receivers. Finally, we illustrate the performance of the network coding algorithm with the approximate decoding on two types of correlated data, i.e., seismic data (external correlation) and video sequences (intrinsic correlation). The simulation results confirms the validity of the GF size analysis and show that the approximate decoding leads to efficient reconstruction when the correlation information is used during decoding. In summary, the main contributions of our paper are (i) a new framework for the distributed delivery of correlated data with network coding, (ii) a novel approximate decoding strategy that exploits the data correlation, (iii) an analysis of the influence of the accuracy of the correlation information and the GF size on the decoding performance, and (iv) the implementation of illustrative examples with external or intrinsic source correlation.

In general, the transmission of correlated sources is studied in the framework of distributed coding [11] (i.e., in the context of Slepian-Wolf problem), where sources are typically encoded by systematic channel encoders and eventually decoded jointly [12], [13]. The DSC (distributed source coding) based network coding schemes [12]–[17] also assume that the source data is encoded by systematic channel codes. Our focus is however not on the design of a distributed compression scheme, which generally assumes that sensors have information about the correlation between the data sources. Rather, we focus on the transmission of correlated data that are encoded independently, transmitted with help of network coding and jointly decoded. Network coding is an efficient solution to the transmission of correlated data over networks with diversity [18]. However, due to the network dynamics, there is no guarantee that each node receives enough useful packets for successful data recovery. Hence, it is essential to have a methodology that enables the recovery of the original data with a good accuracy, when the number of useful packets is not sufficient for perfect decoding. When RLNC is implemented in the network, the encoding and decoding processes of each node are based on linear operations (e.g., linear combinations, inverse of linear matrix, etc.) in a finite algebraic field. Hence, if the number of innovative packets is not sufficient for perfect decoding, one can implement a simple regularization technique, which minimizes the norm of the errors in decoding with the help of the pseudo-inverse of the encoding matrix. While regularization techniques can provide a closed form solution and can be deployed in a general case, they may result in significantly unreasonable approximation [19]. Tikhonov regularization provides an improved performance by slightly modifying the standard least square formula. However, this technique requires to determine additional optimization parameters, which is nontrivial in practice. Sparsity assumptions might also be used [20] for decoding in underdetermined systems in case where a sparse model of the signal of interest is known a priori. However, these regularization techniques have been designed and developed for the field of real numbers, but not for the finite fields. Thus, they may show significantly poor performance if they are blindly applied in our framework, as they cannot consider several properties (e.g., cyclic) of finite field operations. Furthermore, we would like to note that, to the best of our knowledge, there is no regularization techniques that works on finite fields.

The paper is organized as follows. In Section II, we present our framework and describe the approximate decoding algorithm. We discuss the influence of the correlation information in the approximate decoding process in Section III. In Section IV, we analyze the relation between the decoding performance and the GF size, and then determine an optimal GF size that achieves the smallest expected decoding error. Section V and Section VI provide illustrative examples that show how the proposed approach can be implemented in sensor network or video delivery applications.

II. Approximate Decoding Framework

We describe here the general framework considered in this paper and present the distributed delivery strategy for correlated data sources. We also discuss the concept of approximate decoding that enables receivers to estimate the source information when the number of data packets is not sufficient for perfect decoding.

A. RLNC Encoding

We consider an overlay network with sources, intermediate nodes, and clients distributed over a network (e.g., ad hoc network). We denote by $x_1, \ldots, x_N$ the symbols generated by N discrete and correlated sources, where $x_n \in \mathcal{X} (\subset \mathbb{R})$ for $1 \leq n \leq N$. $\mathcal{X}$ is an alphabet set of x_n and $|\mathcal{X}|$ denotes a size of $\mathcal{X}$. These source data are transmitted to the clients via the intermediate nodes that are able to perform network coding (i.e., RLNC). Hence, each x_n needs to be considered as an element in GF. In order to be consistent with the operations defined in different fields, we introduce an *identity function*, which maps a value in a field into the corresponding same value in another field. Specifically, we consider an identity function $1_{\mathbb{R}G}$: $\mathbb{R} \rightarrow$ GF, defined as $1_{\mathbb{R}G}(x) = x^{\mathcal{F}}$, where the superscript $\mathcal{F}$ denotes that the element is in a finite field, which maps a real value in source data into a finite field value for coding operations. Similarly, the identity function $1_{G\mathbb{R}}$: GF $\rightarrow \mathbb{R}$ can be defined as

$1_{G\mathbb{R}}(x^{\mathcal{F}}) = x$. Thus, a node k in RLNC can transmit

$$y(k)^{\mathcal{F}} = \sum_{n=1}^{N} \bigoplus \left\{ c_n(k) \otimes x_n^{\mathcal{F}} \right\} \triangleq \left\{ c_1(k) \otimes x_1^{\mathcal{F}} \right\} \oplus \left\{ c_2(k) \otimes x_2^{\mathcal{F}} \right\} \oplus \cdots \oplus \left\{ c_N(k) \otimes x_N^{\mathcal{F}} \right\}$$

which is a linear combination of $x_n^{\mathcal{F}}$ and coding coefficients $c_n(k)$. $\oplus$ and $\otimes$ denote an additive operation and multiplicative operation defined in GF, respectively. The coding coefficients are uniformly and randomly chosen from GF with size 2^r, denoted by GF(2^r). This implies that the GF size is determined by r and $c_n(k) \in$ GF(2^r). The symbols generated in each node are transmitted to its neighboring nodes towards the sink or client nodes. The size of the field determines the set of coding operations that can be performed on source symbols. We thus assume that the size of the input set is $|\mathcal{X}| \leq 2^r$, which means that coding operations do not need to discard information of input data in GF(2^r). If $|\mathcal{X}| > 2^r$ a priori, the input set is reduced (using e.g., source binning or quantization), such that the resized input set does not exceed the GF size (i.e., 2^r).

If K innovative (i.e., linearly independent) symbols $y(1)^{\mathcal{F}}, \ldots, y(K)^{\mathcal{F}}$ are available at a decoder, a linear system $\mathbf{y}^{\mathcal{F}} = \mathbf{C} \odot \mathbf{x}^{\mathcal{F}}$ can be formed as[1]

$$\left[\begin{array}{ccc} y(1)^{\mathcal{F}} & \cdots & y(K)^{\mathcal{F}} \end{array} \right]^T = \left[\begin{array}{ccc} \mathbf{c}_1 & \cdots & \mathbf{c}_N \end{array} \right] \odot \left[\begin{array}{ccc} x_1^{\mathcal{F}} & \cdots & x_N^{\mathcal{F}} \end{array} \right]^T \triangleq \sum_{n=1}^{N} \bigoplus \{ \mathbf{c}_n \otimes x_n^{\mathcal{F}} \} \tag{1}$$

where $\odot$ denotes a multiplication between matrices in a finite field. The $K \times N$ matrix $\mathbf{C}$ is referred to as the coding coefficient matrix, which consists of column vectors $\mathbf{c}_n = [c_n(1), c_n(2), \cdots, c_n(K)]^T$. An illustrative example for $N = 3$ is shown in Fig. 1, where the symbols x_1, x_2, and x_3 from three sources are encoded by several nodes with different coding coefficients.

B. Approximate Decoding

Upon receiving a set of symbols $\mathbf{y}^{\mathcal{F}}$, the decoder attempts to recover the source data. If $K = N$, i.e., the coding coefficient matrix $\mathbf{C}$ is full-rank, then $\mathbf{x}$ is uniquely determined as $\mathbf{x} = 1_{G\mathbb{R}} \left(\mathbf{C}^{-1} \odot \mathbf{y}^{\mathcal{F}} \right)$ from the linear system in (1). Note that $\mathbf{C}^{-1}$ represents the inverse of the coding coefficient matrix $\mathbf{C}$. The decoding can thus be implemented with well-known approaches such as the Gaussian elimination method over a finite field.

However, if the number of received symbols is insufficient (i.e., $K < N$), there may be an infinite number of solutions $\hat{\mathbf{x}}^{\mathcal{F}} = \left[\hat{x}_1^{\mathcal{F}}, \ldots, \hat{x}_N^{\mathcal{F}} \right]^T$ to the system in (1), as $\mathbf{C}$ is not full-rank. Hence, additional constraints should be imposed in order for the coding coefficient matrix to become a full-rank matrix. This leads to approximate decoding, where the correlation of the input data is exploited to set additional constraints $\mathbf{D}$ in the decoding process so that the system becomes solvable. The additional constraints $\mathbf{D}$ typically depend on the problems under consideration, and on the correlation model between the input data. With the additional constraints, an approximation decoding solution can be found using decoding algorithms such as the Gaussian elimination method in a finite field, i.e.,

$$\hat{\mathbf{x}}^{\mathcal{F}} = \left[\begin{array}{cc} \mathbf{C} & \mathbf{D} \end{array} \right]^{-T} \odot \left[\begin{array}{cc} \mathbf{y}^{\mathcal{F}} & \mathbf{0}_{(N-K)}^{\mathcal{F}} \end{array} \right]^T \tag{2}$$

where $\mathbf{0}_{(N-K)}^{\mathcal{F}}$ is a vector with $(N-K)$ zeros. An approximation $\hat{\mathbf{x}}$ of the original data can then be obtained by the identity mapping defined previously, i.e., $\hat{\mathbf{x}} = 1_{G\mathbb{R}} \left(\hat{\mathbf{x}}^{\mathcal{F}} \right)$. The distortion introduced by information loss and approximate decoding can be written as $\|\mathbf{x} - \hat{\mathbf{x}}\|$.

While the approximate decoding framework is rather generic, we focus in this paper on a simple algorithm where the matrix $\mathbf{D}$ is a $(N-K) \times N$ matrix of coefficients that only depends on the correlation between data. In particular, we construct $\mathbf{D}$ with each row consisting of zeros (i.e., additive identity of GF(2^r)) except two elements of value "1" and "1" [2] that correspond to the positions of the highly correlated data $x_i, x_j \in \mathcal{X}$. Moreover, zeros are also appended to $\mathbf{y}^{\mathcal{F}}$ and represent the results of the additional conditions set in $\mathbf{D}$. In our simple model, approximate decoding relies on the assumption that highly correlated data have a very similar value. This results into encouraging the decoder to reconstruct similar values for correlated data whenever the number of symbols is not sufficient for perfect decoding. This particular instance of approximate decoding is described in Algorithm 1.

Note that the coding coefficient matrix in (2) is assumed here be non-singular, which happens with high probability if the size of the GF is large enough. However, the probability that the coding coefficient matrix becomes singular increases as the size of $\mathbf{D}$ is enlarged. In this case, the system includes a large number of correlation-driven coefficient rows with respect to the random coefficients of the original coding matrix. The impact of the singularity of the coding coefficient matrix on the performance of the approximated decoding is quantified in Section VI-B. Finally, we generally consider that there exists a solution to the decoding problem formed by the augmented coefficient matrix. Otherwise the decoder implementation detects problematic situations and outputs a decoding error signal.

[1] In this paper, vectors and matrices are represented by boldfaced lowercase and boldfaced capital letters, respectively.

[2] "1" is an additive inverse of "1" in GF(2^r).

Algorithm 1 Approximate Decoding

Set: received symbols $\mathbf{y}^{\mathcal{F}}$, coefficient matrix $\mathbf{C}$, correlation model, data size N, GF size 2^r.

1: **if** $rank(\mathbf{C}) = N$, **then**
2: $\quad \mathbf{x} = 1_{G\mathbb{R}}\left(\mathbf{C}^{-1} \odot \mathbf{y}^{\mathcal{F}}\right)$
3: **else** // $rank(\mathbf{C}) < N$ and use approximate decoding
4: $\quad$ Construct $\mathbf{D}$ based on correlation model
5: $\quad$ Solve approximate system and compute $\hat{\mathbf{x}} = 1_{G\mathbb{R}}\left(\left[\begin{array}{cc} \mathbf{C} & \mathbf{D} \end{array}\right]^{-T} \odot \left[\begin{array}{cc} \mathbf{y}^{\mathcal{F}} & \mathbf{0}^{\mathcal{F}}_{(N-K)} \end{array}\right]^{T}\right)$
6: **end if**

We study in the next section the influence of the finite field size (GF size) in the proposed approximate decoding algorithm. Specific implementations of the approximate decoding are later discussed in detail in Section V and Section VI with illustrative examples.

III. Approximate Decoding based on Correlation Information

We discuss in this section the performance of the proposed approximate decoding for recovering the original data from an insufficient number of innovative packets. In particular, we analyze and quantify the impact of the data correlation on the decoding performance when the augmented system at decoder enforces that correlated data should have similar values after decoding. We further study the influence of the accuracy of the correlation estimation.

We first show that the decoding error in our approximate decoding algorithm decreases as data correlation increases. This is described in Property 1.

Property 1: Suppose that $[\mathbf{C}\ \mathbf{D}]^T$ *in* (2) *of the approximate decoding is not singular. Then, the maximum expected errors between the original and recovered data decrease as the data that are used for constructing* $\mathbf{D}$ *in* (2) *are more correlated.*

Proof: Let $\mathbf{y}^{\mathcal{F}}$ be a set of K received innovative packets (with K smaller than the number of original symbols N, i.e., $K < N$). Let further $\mathbf{C}$ and $\mathbf{x}$ be the corresponding coding coefficient matrix and original data as in (1). Since $K(< N)$ innovative packets are available, $(N-K)$ additional constraints are imposed into the coding coefficient matrix as $\mathbf{D}$. Based on the correlation model of the original data, we try to enforce the most correlated sources to be decoded into similar symbols. This leads to the approximated decoding solution $\mathbf{x}^{\mathcal{F}}$ in (2).

We now analyze the error incurred by approximate decoding compared to the perfect decoding with a full-rank system. The approximate solution $\hat{\mathbf{x}} = 1_{G\mathbb{R}}\left(\hat{\mathbf{x}}^{\mathcal{F}}\right)$ is compared to the exact solution $\mathbf{x}$. This exact solution is reconstructed based on the set of coding coefficients $\mathbf{C}$ and the coefficients $\mathbf{D}$, but with the exact constraints $\mathbf{d}^{\mathcal{F}}$ and not their approximation by a zero vector as done in (2). We denote these actual constraints by the vector $\mathbf{d}^{\mathcal{F}} = \mathbf{D} \odot \mathbf{x}^{\mathcal{F}} = \left[d(1)^{\mathcal{F}}, \ldots, d(N-K)^{\mathcal{F}}\right]^T$ that is computed by applying the additional coefficients in $\mathbf{D}$ on the original vector $\mathbf{x}$. Equivalently, the source $\mathbf{x}$ can be computed by

$$\mathbf{x} = 1_{G\mathbb{R}}\left(\mathbf{x}^{\mathcal{F}}\right) = 1_{G\mathbb{R}}\left(\left[\begin{array}{cc} \mathbf{C} & \mathbf{D} \end{array}\right]^{-T} \odot \left[\begin{array}{cc} \mathbf{y}^{\mathcal{F}} & \mathbf{d}^{\mathcal{F}} \end{array}\right]^{T}\right). \tag{3}$$

Note that $\hat{\mathbf{x}}^{\mathcal{F}}$ in (2) and $\mathbf{x}^{\mathcal{F}}$ in (3) are obtained based on the operations defined in GF(2^r), and thus, the resulting elements in $\mathbf{x}^{\mathcal{F}}$ or $\hat{\mathbf{x}}^{\mathcal{F}}$ are indeed in GF(2^r). However, they originally represent data in $\mathbb{R}$ (e.g., source data). Hence, in order to quantify the performance of the proposed algorithm, we are interested in the error between the exact and approximate solutions, i.e., $\|\mathbf{x} - \hat{\mathbf{x}}\|$, where the norm operation $\|\cdot\|$ is defined in a vector space over a subfield of the complex numbers.

Using (2) and (3), the error between the exact and approximate solutions can be expressed as

$$\|\mathbf{x} - \hat{\mathbf{x}}\| = \left\|1_{G\mathbb{R}}\left(\mathbf{x}^{\mathcal{F}}\right) - 1_{G\mathbb{R}}\left(\hat{\mathbf{x}}^{\mathcal{F}}\right)\right\| = \left\|1_{G\mathbb{R}}\left(\left[\begin{array}{cc} \mathbf{C} & \mathbf{D} \end{array}\right]^{-T} \odot \left[\begin{array}{cc} \mathbf{0}^{\mathcal{F}}_{K} & \mathbf{d}^{\mathcal{F}} \end{array}\right]^{T}\right)\right\|. \tag{4}$$

From the assumption that $[\mathbf{C}\ \mathbf{D}]^T$ in (2) is not singular, its inverse can be written as $[\mathbf{C}\ \mathbf{D}]^{-T} = \left[\mathbf{q}_1^{\mathcal{F}}, \ldots, \mathbf{q}_K^{\mathcal{F}}, \mathbf{q}_{K+1}^{\mathcal{F}}, \ldots, \mathbf{q}_N^{\mathcal{F}}\right]$, where $\mathbf{q}_n^{\mathcal{F}}$ is the nth column vector. Thus,

$$\begin{aligned}
\|\mathbf{x} - \hat{\mathbf{x}}\| &= \left\|1_{G\mathbb{R}}\left(\left[\mathbf{q}_1^{\mathcal{F}}, \ldots, \mathbf{q}_K^{\mathcal{F}}, \mathbf{q}_{K+1}^{\mathcal{F}}, \ldots, \mathbf{q}_N^{\mathcal{F}}\right] \odot \left[\begin{array}{cc} \mathbf{0}_K^{\mathcal{F}} & \mathbf{d}^{\mathcal{F}} \end{array}\right]^{T}\right)\right\| \\
&= \left\|1_{G\mathbb{R}}\left(\sum_{l=1}^{N-K} \bigoplus \left\{\mathbf{q}_{K+l}^{\mathcal{F}} \otimes d(l)^{\mathcal{F}}\right\}\right)\right\| \\
&\leq \left\|\sum_{l=1}^{N-K} 1_{G\mathbb{R}}\left(\mathbf{q}_{K+l}^{\mathcal{F}} \otimes \{x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}\}\right)\right\| \\
&\leq \sum_{l=1}^{N-K} \|\mathbf{q}_{K+l}\|\, 1_{G\mathbb{R}}\left(x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}\right)
\end{aligned} \tag{5}$$

where $d(l)^{\mathcal{F}} = x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}$, $0 \leq i, j \leq N$, as each element in $\mathbf{d}^{\mathcal{F}}$ depends on the two non-zero elements in $\mathbf{D}$ due to our choice of the additional constraints. Note that the relations in the last lines of (5) are valid due to the identity mapping defined in Section II and the unique correspondences between elements in the finite field and the elements on the real field. Since the coding coefficient matrix is fixed, the error $\|\mathbf{x} - \hat{\mathbf{x}}\|$ only depends on $1_{G\mathbb{R}}\left(x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}\right)$. Therefore,

$$E\{\|\mathbf{x} - \hat{\mathbf{x}}\|\} \leq \sum_{l=1}^{N-K} \|\mathbf{q}_{K+l}\| \, E\left\{1_{G\mathbb{R}}\left(x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}\right)\right\}. \tag{6}$$

As x_i and x_j are highly correlated, i.e., higher value of correlation coefficient $\rho_{i,j}$, the difference between them (i.e., $|x_i - x_j|$) decreases, and vice versa. Moreover, as shown experimentally in Appendix I, smaller values of $|x_i - x_j|$ then result in smaller values of $1_{G\mathbb{R}}\left(x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}\right)$ with very high probability. Therefore, the maximum expected errors, i.e., $E\{\|\mathbf{x} - \hat{\mathbf{x}}\|\}$, between the exact and approximate solutions decreases when the data correlation increases. ■

Property 1 confirms that the expected decoding error can be reduced if the information about correlation between the original data is exploited at the decoder. It also implies that the maximum decoding error is bounded, and that this bound can be minimized if the original data is highly correlated. This means that the best way to construct $\mathbf{D}$ consists in building additional constraints with sources that have the highest correlation (i.e., highest correlation coefficient ρ). Interestingly, we do not need to know the exact correlation values for approximate decoding; it is sufficient to identify the most correlated sources.

We further analyze the influence of the accuracy of the correlation model. Recall that, as shown in Property 1, the decoding error can be minimized by exploiting the correlation (that can be measured by correlation coefficient $\rho_{i,j}$) among data and by constructing $\mathbf{D}$ based on the correlation information. In practice, the exact correlation information may not be available; it often needs to be estimated from (possibly imperfect) observations. The observations may include estimation noise, thereby leading to inaccurate estimation of the correlation model. We now consider the observation noise in the process of estimating the correlation model and investigate its impact on the decoding error.

Corollary 1: If the correlation model is more accurately estimated, the performance of the approximate decoding improves.

Proof: Let x_i' and x_j' be noisy observations of x_i and x_j with noise n_i and n_j, respectively, i.e., $x_i' = x_i + n_i$ and $x_j' = x_j + n_j$. We assume that n_i and n_j are independent and zero mean random variables with the variances of $\sigma_{n_i}^2$ and $\sigma_{n_j}^2$, respectively. The corresponding correlation coefficient $\rho_{i,j}'$ for x_i' and x_j' can be expressed as

$$\begin{aligned}
\rho_{i,j}' &= \frac{E(x_i'x_j') - E(x_i')E(x_j')}{\sqrt{E(x_i'^2) - E(x_i')^2}\sqrt{E(x_j'^2) - E(x_j')^2}} \\
&= \frac{E((x_i + n_i)(x_j + n_j)) - E(x_i + n_i)E(x_j + n_j)}{\sqrt{E((x_i + n_i)^2) - E(x_i + n_i)^2}\sqrt{E((x_j + n_j)^2) - E(x_j + n_j)^2}} \\
&= \frac{E(x_ix_j) - E(x_i)E(x_j)}{\sqrt{E(x_i^2) - E(x_i)^2 + \sigma_{n_i}^2}\sqrt{E(x_j^2) - E(x_j)^2 + \sigma_{n_j}^2}} \leq \rho_{i,j}
\end{aligned} \tag{7}$$

because $\sigma_{n_k}^2 \geq 0$ for $k \in \{i, j\}$. This implies that the noise in the data observations results in a smaller correlation coefficient than the true correlation coefficient. This may further result in a failure to find the best matched data and in reduced performance for the approximate decoding algorithm, as shown in Property 1. Hence, more accurate estimation of the correlation model can provide better performance. ■

In summary, we observe that the efficiency of approximate decoding increases with the data correlation and with the accuracy of the correlation information that is used to derive additional constraints for decoding.

IV. Optimal Finite Field Size

We study here the importance of designing the coding coefficient matrix, and in particular, of the size of the finite field on the performance of the approximate decoding framework. This size has an influence on the reconstruction error, when the number of symbols is insufficient for perfect decoding. At the same time, the field size drives the resolution of the source encoding since only a finite number of symbols can be uniquely represented by the identity mapping defined in Section II. We therefore determine the optimal GF size that minimizes the expected decoding error by trading off source approximation error and decoding error probability.

We first prove the following property, which states that the probability of large decoding errors increases with the GF size. In the analysis, we assume that the decoding errors between the transmitted data and the data recovered from the approximate decoding are uniformly distributed over $\mathcal{X}$.

Property 2: Given a finite set of data $\mathcal{X}$, the expected decoding error increases as the size of the finite field for the coding operations increases.

Proof: Let $x \in \mathcal{X}$ be an original data, where the size of the input space is given by $|\mathcal{X}| = 2^r$. Let further $\hat{x}_r$ and $\hat{x}_R$ be the decoded symbols when coding is performed in respectively GF(2^r) and GF(2^R) with $R > r$, for $r, R \in \mathbb{N}$, i.e., GF(2^R) is

an extended GF from GF(2^r). Since the decoding errors are uniformly distributed over $\mathcal{X}$, the probability mass function of $\hat{x}_k$ is given by

$$p_k(\hat{x}_k) = \begin{cases} 1/2^k, & \text{if } x_k \in [0, 2^k - 1] \\ 0, & \text{otherwise} \end{cases}$$

for $k \in \{r, R\}$. To prove that a larger GF size results in a higher decoding error, we have to show that

$$\Pr\left(|x - \hat{x}_R| \geq |x - \hat{x}_r|\right) > \frac{1}{2}. \tag{8}$$

If this condition is satisfied, the expected distortion is larger for x_R than x_r or equivalently for the larger GF size. The left hand side of (8) can be expressed as

$$\begin{aligned} &\Pr\left(\hat{x}_R \geq \hat{x}_r, x \leq \frac{\hat{x}_R + \hat{x}_r}{2}\right) + \Pr\left(\hat{x}_R < \hat{x}_r, x > \frac{\hat{x}_R + \hat{x}_r}{2}\right) \\ &= \Pr\left(\hat{x}_R \geq \hat{x}_r\right) \Pr\left(x \leq \frac{\hat{x}_R + \hat{x}_r}{2} \middle| \hat{x}_R \geq \hat{x}_r\right) + \Pr\left(\hat{x}_R < \hat{x}_r\right) \Pr\left(x > \frac{\hat{x}_R + \hat{x}_r}{2} \middle| \hat{x}_R < \hat{x}_r\right) \\ &= \left(1 - 2^{r-R-1}\right) \hat{P} + 2^{r-R-1}\left(1 - \hat{P}\right) = 2^{r-R-1} + \left(1 - 2^{r-R}\right) \hat{P} \end{aligned}$$

because $\hat{x}_R$ and $\hat{x}_r$ are both uniformly distributed. In the previous equations, we have posed $\hat{P} \triangleq \Pr\left(x \leq \frac{\hat{x}_R + \hat{x}_r}{2} \middle| \hat{x}_R \geq \hat{x}_r\right)$. We further show in Appendix II that $\hat{P} > \frac{1}{2}$. Therefore, we have

$$2^{r-R-1} + \left(1 - 2^{r-R}\right) \hat{P} > 2^{r-R-1} + \left(1 - 2^{r-R}\right) \cdot \frac{1}{2} = \frac{1}{2}, \tag{9}$$

which completes the proof. ■

Property 2 implies that a small GF size is preferable in terms of expected decoding error. In particular, it is preferred not to enlarge the GF size more than the size of the input space since approximate decoding performs worse in very large field where uncertainty becomes to large in case of loss.

Alternatively, if the GF size becomes smaller than the size of the input alphabet size, the maximum number of source symbols that can be distinctively represented decreases correspondingly. Specifically, if we choose a GF size of $2^{r'}$ such that $|\mathcal{X}| > 2^{r'}$ for $r' < r$, part of the data in $\mathcal{X}$ needs to be discarded to form a subset $\mathcal{X}'$ such that $|\mathcal{X}'| \leq 2^{r'}$. In this case, we assume that if the GF size is reduced from GF(2^r) to GF(2^{r-z}), where $0 \leq z (\in \mathbb{Z}) \leq r - 1$, the least significant z bits are first discarded from $x \in \mathcal{X}$. Then, all the data in $\mathcal{X}'$ can be distinctly encoded in GF($2^{r'}$).

In summary, while reducing the GF size may result in lower decoding error, it may induce larger information loss in the source data. Based on this clear tradeoff, we propose Property 3 that shows the existence of an optimal GF size. Note that discarding part of source data information results in errors at the source, similar to data quantization. Thus, we assume that the corresponding source information loss is uniformly distributed and that the decoded data is also uniformly distributed in the following analysis.

Property 3: There exists an optimal GF size that minimizes the overall expected error in data reconstruction at decoder. The optimal GF size is given by GF(2^{r-z^*}), where $z^* = \lceil (r-1)/2 \rceil$ and $z^* = \lfloor (r-1)/2 \rfloor$.

Proof: Suppose that the number of original source symbols is $|\mathcal{X}| = 2^r$ and that the coding field is GF(2^r). As discussed in Property 2, the GF size does not need to be enlarged more than 2^r, as this only increases the probability of the expected decoding error. If the GF size is reduced from GF(2^r) to GF(2^{r-z}), the approximate decoding is more efficient and the decoding errors are uniformly distributed over $[-r_D, r_D]$, where $r_D = 2^{r-1-z} - 1$, i.e.,

$$p_{e_D}(e_D) = \begin{cases} 1/(2r_D + 1), & \text{if } e_D \in [-r_D, r_D] \\ 0, & \text{otherwise} \end{cases}. \tag{10}$$

At the same time, if the GF size is reduced, the input data set $\mathcal{X}$ is reduced to $\mathcal{X}'$ and the number of input symbols is decreased. By discarding the z least significant bits, the number of input symbols becomes $|\mathcal{X}'| = 2^{r-z}$. Such an information loss also results in errors over $[-r_I, r_I]$, where $r_I = 2^z - 1$, i.e.,

$$p_{e_I}(e_I) = \begin{cases} 1/(2r_I + 1), & \text{if } e_I \in [-r_I, r_I] \\ 0, & \text{otherwise} \end{cases}. \tag{11}$$

Based on these independent distortions, the distribution of the total error, $p_{e_T}(e_T) = p_{e_D}(e_D) + p_{e_I}(e_I)$, is given by [21]

$$p_{e_T}(e_T) = \frac{H}{2}\left\{|e_T + r_I + r_D + 1| - |e_T + r_I - r_D| - |e_T - r_I + r_D| + |e_T - r_I - r_D - 1|\right\}$$

for $|e_T| \leq r_I + r_D \triangleq e_T^{max}$ and $H = (2r_I + 1)^{-1}(2r_D + 1)^{-1}$. Since $e_T + r_I + r_D + 1 \geq 0$ and $e_T - r_I - r_D - 1 \leq 0$ for all $|e_T| \leq e_T^{max} (= r_I + r_D)$, by substituting r_I and r_D, we have

$$p_{e_T}(e_T) = \frac{H}{2}\left\{2\left(2^z + 2^{r-1-z} - 1\right) - \left|e_T + 2^z - 2^{r-1-z}\right| - \left|e_T - 2^z + 2^{r-1-z}\right|\right\}. \tag{12}$$

By denoting $a(z) \triangleq 2^z - 2^{r-1-z}$ and $b(z) \triangleq 2^z + 2^{r-1-z}$, the expected decoding error can be expressed as

$$E\left[|e_T|\right] = \sum_{e_T=-\infty}^{\infty} |e_T| \cdot p_{e_T}(e_T) = \sum_{e_T=-e_T^{max}}^{e_T^{max}} \frac{H}{2} |e_T| \cdot \left[2(b(z)-1) - |e_T + a(z)| - |e_T - a(z)|\right]. \tag{13}$$

Since both $|e_T|$ and $[2(b(z)-1) - |e_T + a(z)| - |e_T - a(z)|]$ are symmetric on $z = \lceil (r-1)/2 \rceil$ and $z = \lfloor (r-1)/2 \rfloor$ (see Appendix III), $E[|e_T|]$ is also symmetric. Thus,

$$\begin{aligned} E[|e_T|] &= H \sum_{e_T=1}^{e_T^{max}} e_T \cdot \{2(b(z)-1) - |e_T + a(z)| - |e_T - a(z)|\} \\ &= H \sum_{e_T=1}^{e_T^{max}} e_T \cdot \{2(b(z)-1)] - H \sum_{e_T=1}^{e_T^{max}} e_T \cdot \{|e_T + a(z)| + |e_T - a(z)|\} \\ &= H \cdot (b(z)-1) e_T^{max}(e_T^{max}+1) - H \sum_{e_T=1}^{e_T^{max}} e_T \cdot \{|e_T + a(z)| + |e_T - a(z)|\}. \end{aligned} \tag{14}$$

If we consider the case where $a(z) > 0$, which corresponds to $r/2 < z \leq r-1$, we have

$$\begin{aligned} \sum_{e_T=1}^{e_T^{max}} e_T \cdot \{|e_T + a(z)| + |e_T - a(z)|\} &= \sum_{e_T=1}^{a(z)-1} e_T \cdot 2a(z) + \sum_{e_T=a(z)}^{e_T^{max}} e_T \cdot 2e_T \\ &= \frac{1}{3} e_T^{max}(e_T^{max}+1)(2e_T^{max}+1) + \frac{1}{3} a(z)(a(z)^2 - 1). \end{aligned}$$

Note that $e_T^{max} = b(z) - 2$. Therefore, for the case where $a(z) > 0$, $E[e_T]$ can be expressed as

$$\begin{aligned} E[e_T] &= H \cdot \left[(b(z)-1)^2(b(z)-2) - \frac{1}{3}(b(z)-1)(b(z)-2)(2b(z)-3) - \frac{1}{3}a(z)(a(z)^2-1) \right] \\ &= H \cdot \left[\frac{1}{3} b(z)(b(z)-1)(b(z)-2) - \frac{1}{3} a(z)(a(z)^2-1) \right] \end{aligned} \tag{15}$$

which is an increasing function for $r/2 < z \leq r-1$ (see Appendix IV). Since $E[e_T]$ is symmetric on $z = \lceil (r-1)/2 \rceil$ and $z = \lfloor (r-1)/2 \rfloor$, and is an increasing function over $r/2 < z \leq r-1$, $E[e_T]$ is convex over $0 \leq z \leq r-1$. Therefore, there exists an optimal z^* that minimizes the expected decoding error.

Finally, since $E[e_T]$ is symmetric on $\lceil (r-1)/2 \rceil$ and $\lfloor (r-1)/2 \rfloor$, the minimum $E[e_T]$ can be achieved if $z^* = \lceil (r-1)/2 \rceil$ and $z^* = \lfloor (r-1)/2 \rfloor$. These two optimum points can be identical if r is odd. ■

V. Approximate Decoding in Sensor Networks

A. System Description

We illustrate in this section an example, where the approximated decoding framework is used to recover the data transmitted by sensors that capture a source signal from different spatial locations. We consider a sensor network, where sensors transmit RLNC encoded data.

Specifically, each sensor measures its own observations and receives the other observations from its neighbor sensors. Then, each sensor combines the received data with its own data using RLNC. It transmits the resulting data to its relay nodes or receivers. In the considered scenario, the sensors measure seismic signals and are placed at a distance of 100m by each other.

A sensor h captures a signal X_h that represents a series of sampled values in a window of size w, i.e., $X_h = [x_h^1, \ldots, x_h^w]^T$. We assume that the data measured at each sensor are in the range of $[-x_{min}, x_{max}]$, i.e., $x_h^l \in \mathcal{X} = [-x_{min}, x_{max}]$ for all $1 \leq l \leq w$, and that they are quantized and mapped to the nearest integer values, i.e., $x_h^l \in \mathbb{Z}$. Thus, if the measured data exceed the range of $[-x_{min}, x_{max}]$, then they are clipped to the minimum or maximum values of the range (i.e., $x_h^l = -x_{min}$ or $x_h^l = x_{max}$).

The data captured by the different sensors are correlated, as the signals at different neighboring positions are time-shifted and energy-scaled versions of each other. The captured data have lower correlation as the distance between sensors becomes larger. An illustrative example is shown in Fig. 4(a) that presents seismic data recorded by 3 different sensors. The data measured by sensor 1 has much higher temporal correlation with the data measured by sensor 2 in terms of time shift and signal energy than the data measured by sensor 30. This is because sensor 2 is significantly closer to sensor 1 than sensor 30.

We consider that the nodes perform network coding for data delivery. We denote by $\mathbf{S}_n (\subseteq \mathbf{S})$ a set of sensors that are in the proximity of a sensor $n \in \mathbf{S}$. The number of sensors in $\mathbf{S}_n$ is $|\mathbf{S}_n| = N_n$. A sensor n receives data X_h from all the sensors $h \in \mathbf{S}_n$ in its proximity and encodes the received data with RLNC. The coding coefficients $c_h(k)$ are randomly selected from a GF whose size 2^r is determined such that $|\mathcal{X}| \leq 2^r$. The encoded symbols are then transmitted to the neighboring nodes

or to the receiver. The kth encoded data packets for a window are denoted by $Y(k)^{\mathcal{F}} = \sum_{h \in \mathbf{S}_n} \bigoplus \{c_h(k) \otimes X_h^{\mathcal{F}}\}$, where $X_h^{\mathcal{F}} = 1_{\mathbb{R}G}(X_h)$. An illustrative example is shown in Fig. 2. This example presents a set of four sensors denoted by $\mathbf{S}$ that consists in two subsets of neighbors, i.e., $\mathbf{S}_1 = \{1, 3, 4\}$ and $\mathbf{S}_2 = \{2, 4\}$. The encoded data packets that the receiver collects from sensor 2 and sensor 4 are denoted by $Y(k_1)^{\mathcal{F}}$ and $Y(k_2)^{\mathcal{F}}$.

When a receiver collects enough innovative packets, it can solve the linear system given as in (3) and it can recover the original data. However, if the number of packets is not sufficient, the receiver applies our proposed approximate decoding strategy that exploits the correlation between the different signals. With such a strategy, the decoding performance can be improved as discussed in Property 1. We assume that the system setup is approximately known by the sensors. In other words, a simple correlation model can be computed, which include the relative temporal shifts and energy scaling between the signals from the different sensors. In particular, since the sensor positions are known, one can simply assume that the correlation depends on the distance between sensors.

B. Simulation Results

We analyze an illustrative scenario, where the receiver collects encoded packets from sensors 1, 2 and 30 and tries to reconstruct the original signals from these three sensors. We consider temporal windows of size $w = 300$ for data representation. The captured data is in the range of $[0, 1023]$. Thus, the maximum GF size is 2^{10}, i.e., GF(2^{10}). We assume that $2/3$ of the linear equations required for perfect decoding are received with no error, and that the rest of $1/3$ of equations are not received. Thus, the $1/3$ of constraints are imposed into the coding coefficient matrix based on the assumption that the signals from sensor 1 and sensor 2 are highly correlated.

We study the influence of the size of the coding field on the decoding performance. Fig. 3 shows the MSE (Mean Square Error) distortion for the decoded signals for different number of discarded bits z, or equivalently for different GF sizes 2^{10-z}. The conclusion drawn from Property 3 is confirmed from these results, as the decoding error is minimized at $z^* = \lceil (10-1)/2 \rceil = 5$.

An instantiation of seismic data recovered by the approximate decoding is further shown in Fig. 4, where a GF(2^{10-z^*}) = GF(2^5) is used. Since the additional constraints are imposed into the coding coefficient matrix based on the assumption of high correlation between the data measured by sensors 1 and 2, the recovered data of sensors 1 and 2 in Fig. 4(b) are very similar, but at the same time, the data are quite accurately recovered. We observe that the error in correlation estimation results in higher distortion in the signal recovered by sensor 30. However, the first part of the signal is correctly recovered, as the signals captured by sensors 1 and 2 are highly correlated.

VI. Approximate Decoding of Image Sequences

A. System Description

In this section, we illustrate the application of approximate decoding to the recovery of image sequences. We consider a system, where information from successive frames is combined with network coding. Encoded packets are transmitted to a common receiver. Packets may, however, be lost or delayed, which prevents perfect reconstruction of the images. Thus, for improved decoding performance, we exploit the correlation between successive frames.

We consider a group of successive images in a video sequence. Each image X_n is divided into N patches $X_{n,p}$, i.e., $X_n = [X_{n,1}, \ldots, X_{n,N}]$. A patch $X_{n,p}$ contains $L \times L$ pixels $x_{n,p}^b$, $1 \leq b \leq L \times L$, i.e., $X_{n,p} = [x_{n,p}^1, \ldots, x_{n,p}^{L \times L}]$. Such a representation is illustrated on Fig. 5. The system implements RLNC and combines patches at similar positions in different frames to produce encoded symbols. In others words, it produces a series of symbols $Y_p(k)^{\mathcal{F}} = \sum_{n=1}^{N} \bigoplus c_{n,p}(k) \otimes X_{n,p}^{\mathcal{F}}$, where $X_{n,p}^{\mathcal{F}} = 1_{\mathbb{R}G}(X_{n,p})$, for a location of patch p. The coding coefficients $c_{n,p}(k)$ are randomly chosen in GF(2^r). We assume that the original data (i.e., pixels) can take values in $[0, 255]$, and thus, we choose the maximal size of the coding field to be $|\mathcal{X}| = 256 = 2^8$.

When the receiver collects enough innovative symbols per patch, it can recover the corresponding sub-images in each patch, and eventually the group of images. If, however, the number of encoded symbols is insufficient, additional constraints are added to the decoding system in order to enable approximate decoding. These constraints typically depend on the correlation between the successive images. In our case, the constraints are based on similarities between blocks of pixels in successive frames. We impose conditions for blocks in patches at the same position in successive frames to be identical. Formally, we have conditions of the form $x_{n,p}^{b_1} = x_{n+1,p}^{b_2}$, where $1 \leq b_1, b_2 \leq L \times L$. These conditions depend on the motion in the image sequence. The motion information permits to add additional constraints to the decoding system so that estimations of the original blocks of data can be obtained by Gaussian elimination techniques. Due to our design choices, we can finally note that the decoding system can be decomposed into smaller independent sub-systems that correspond to patches.

B. Performance of Approximate Decoding

In our experiments, we consider three consecutive frames extracted from the *Silent* standard MPEG sequence with QCIF format (174×144). The patches are constructed with four blocks of 8×8 pixels. We assume that only $2/3$ of the linear equations required for perfect decoding are received. The decoder implements approximate decoding assuming that the correlation

information is known at the decoder. The missing constraints are added to the decoding system based on the best matched pairs of blocks in consecutive frames, in the sense of the smallest MSE between the pixel values.

In the first set of experiments, we analyze the influence of the size of the coding field, by changing the GF sizes from GF(2^8) to GF(2^{8-z}). We reduce the size of the field by discarding the z least significant bits for each pixel value. Fig. 6 shows the PSNR (Peak Signal to Noise Ratio) quality achieved from the decoded frames for different number of discarded bits z. As discussed in Property 3, the expected decoding error can be minimized if $z^* = \lceil (r-1)/2 \rceil$ and $z^* = \lfloor (r-1)/2 \rfloor$, which corresponds to $z^* = 3$ and $z^* = 4$. This can be verified from this illustrative example, where the maximum PSNR is achieved at $z = 4$ for frame 1 and frame 2, and at $z = 3$ for frame 3. The corresponding decoded images for two different GF sizes are presented in Fig. 7. From the decoded images, we can observe that several patches are completely black or white. This is because the coding coefficient matrices are singular, leading to the failure of Gaussian elimination during the decoding process.

We also illustrate the influence of the accuracy of the correlation information by considering zero motion at the decoder. In other words, additional constraints for approximate decoding simply impose that the consecutive frames are identical. Fig. 8 shows the frames decoded with no motion over GF(32). We can see that the first three frames still provides an acceptable quality since the motion between these frames is actually very small. However, in frames 208, 209, and 210, where motion is higher, we clearly observe significant performance degradation, especially in the positions where high motion exists.

We study next the influence of the size of the group of images (i.e., window size) considered for encoding. It has been discussed that the coding coefficient matrices can be singular, as the coefficients are randomly selected in a finite field. This results in the performance degradation of the approximate decoding. Moreover, it is shown that the probability that random matrices over finite fields are singular becomes smaller as the size of matrices becomes larger [22]. Thus, if the group of images (i.e., window size) becomes larger, the coding coefficient matrix becomes larger. As a result, the probability that Gaussian elimination fails is correspondingly smaller. This is quantitatively investigated from the following experiment.

We design an experiment, where we consider 24 frames extracted from the *Silent* sequence and a set of different window sizes that contains 3, 4, 6, 8, and 12 frames. For example, if window size is 3, then there are 24/3=8 windows that are used in this experiment. The average PSNR achieved in the lossless case, where the decoder receives enough packets for decoding, is presented in Fig. 9. The PSNR increases as the window sizes are enlarged. The only reason why all the frames are not perfectly recovered is the failure of the Gaussian elimination, when the coding coefficient matrices becomes singular. This confirms the above-mentioned discussion, i.e., if window size increases, the size of coding coefficient matrix also increases. Since the probability that the enlarged coding coefficient matrices are singular becomes smaller, higher average PSNRs can correspondingly be achieved for larger size of window.

Finally, we study the influence of the window size in the lossy case. We assume that we have a loss rate of $1/24$ in all the configurations and the approximate decoding is implemented. Fig. 9 shows the achieved average PSNR across the recovered frames for different window sizes. Since the decoding errors incurred by the approximate decoding are limited to a window and do not influence the decoding of the other windows, a small window size is desirable for limited error propagation. However, as discussed, a smaller window size can result in higher probability that the coding coefficient matrices become singular, and correspondingly, the failure of the Gaussian elimination. Due to this tradeoff, we can observe that the achieved PSNR becomes low when window size is 4 in our example. Note that the computational complexity for decoding (i.e., Gaussian elimination) also increases as the window size increases. Hence, the proper window size needs to be determined based on several design tradeoffs.

C. Performance in various network conditions

We thus far considered a network having a fixed packet loss rate (i.e., a dedicated final node receives 2/3 of the required linear equations and does not receive 1/3 of the required linear equations), and investigated several properties of the approximate decoding. We now examine more general network scenarios, which may result in different packet loss rates for the final decoder. As an illustration, we consider a network which consists of three pairs of sources and destinations, and there are several nodes performing network coding operations. We assume that there are no loss in the sources and destinations and they are properly dimensioned. However, the links between nodes performing network coding operations are lossy with different packet loss rates. We study the achieved performance (PSNR) that corresponds to different packet loss rates. The results are shown in Fig. 10. These results show the averaged PSNRs that the final node achieves when it experiences a variety of packet loss rates and decodes the received data with the proposed approximate decoding method for binary symmetric channel (BSC) and Gilbert Elliot channels (GEC) [23], respectively. The source images are from *Container* sample MPEG sequences with QCIF resolution. In all cases, the data is encoded with RLNC and a window of four packets is considered. We simulate loss with a GEC model [23] that consists in a two-state Markov chain where the good and bad states represent the correct reception or the loss of a packet, respectively. We choose the average length of burst of errors to 9 packets, and we vary the average loss rate in order to study the performance of our approximate reconstruction algorithm in different channel conditions. As expected, performance worsens as the packet loss rate increases. Moreover, these results show that the approximate decoding enables the decoder to achieve a noticeable gain in terms of decoded quality compared to traditional network coding based systems, which may completely fail to recover data. This implies that the approximate decoding may require less network loads than traditional decoding algorithms in order to achieve the same decoding quality.

VII. Conclusions

In this paper, we have described a framework for the delivery of correlated information sources with help of network coding and a novel approximate decoding algorithm. The approximate decoding algorithm permits to reconstruct an approximation of the source signals even when an insufficient number of innovative packets are available for perfect decoding. We have analyzed the tradeoffs between the decoding performance and the size of the coding fields. We have determined an optimal field size that leads to the highest approximated decoding performance. We also have investigated the impact of the accuracy of the data correlation information on the decoding performance. The proposed approach is implemented in illustrative examples of sensor network and image transmission applications, where the experimental results confirm our analytical study as well as the benefits of approximate decoding solutions.

VIII. Acknowledgments

The authors would like to thank Dr. Laurent Duval for providing the seismic data used in the sensor network example.

Appendix I

In this appendix, we provide illustrative examples that verify the arguments, where smaller values of $|x_i - x_j|$ can indeed lead to smaller values of $1_{G\mathbb{R}}(x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}})$. In this example, we consider GF(512), and study several examples of $|x_i - x_j| = 0, 1, 2, 50, 100, 150, 256$. In the cases where small differences between x_i and x_j (e.g., $|x_i - x_j| = 0, 1, 2$), we can observe that the most of the values of $1_{G\mathbb{R}}\left(x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}\right)$ are concentrated around 0. In the cases where large differences between x_i and x_j (e.g., $|x_i - x_j| = 50, 100, 150, 256$), however, the values of $1_{G\mathbb{R}}\left(x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}\right)$ are spread over 512. Therefore, it is obviously confirmed that smaller values of $|x_i - x_j|$ indeed result in $1_{G\mathbb{R}}\left(x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}\right)$. These are depicted in Fig. 11.

Appendix II

In this appendix, we show that $\hat{P} \geq \frac{1}{2}$, defined in (9). Note that both $\hat{x}_r$ and $\hat{x}_R$ are reconstructed data, and thus, they are real values. Using Bayes' rule,

$$\begin{aligned}\hat{P} = \Pr\left(x \leq \frac{\hat{x}_R + \hat{x}_r}{2} \middle| \hat{x}_R \geq \hat{x}_r\right) &= \sum_{z=0}^{2^r-1} \Pr\left(z \leq \frac{\hat{x}_R + \hat{x}_r}{2} \middle| \hat{x}_R \geq \hat{x}_r, x = z\right) \Pr\left(x = z\right) \\ &= \frac{1}{2^r} \sum_{z=0}^{2^r-1} \Pr\left(z \leq \frac{\hat{x}_R + \hat{x}_r}{2} \middle| \hat{x}_R \geq \hat{x}_r, x = z\right).\end{aligned}$$

Hence, we have

$$\begin{aligned}\sum_{z=0}^{2^r-1} \Pr\left(z \leq \frac{\hat{x}_R + \hat{x}_r}{2} \middle| \hat{x}_R \geq \hat{x}_r, x = z\right) &= \frac{1}{2^{r+R}} \sum_{z=0}^{2^r-1} \left[2^{r+R} - \left\{2^{r-1}(2^r - 1) + 2\sum_{l=0}^{z} l\right\}\right] \\ &= \frac{1}{2^{r+R}} \left\{2^{2r+R} - \frac{1}{6}\left(5 \cdot 2^{3r} - 3 \cdot 2^{2r} - 2 \cdot 2^r\right)\right\}.\end{aligned}$$

Thus, $\hat{P}$ can be expressed as

$$\hat{P} = \frac{1}{2^r}\left[\frac{1}{2^{r+R}}\left\{2^{2r+R} - \frac{1}{6}\left(5 \cdot 2^{3r} - 3 \cdot 2^{2r} - 2 \cdot 2^r\right)\right\}\right] = 1 - \frac{1}{6}\left[5 \cdot \frac{2^r}{2^R} - \frac{3}{2^R} - \frac{2}{2^{r+R}}\right].$$

Since $r, R \in \mathbb{N}$ and $R > r$, R can be expressed as $R = r + \alpha$, where $\alpha \in \mathbb{N}$. Thus,

$$\hat{P} = 1 - \frac{1}{6}\left[5 \cdot \frac{1}{2^\alpha} - \frac{3}{2^{r+\alpha}} - \frac{2}{2^{2r+\alpha}}\right].$$

Since $\lim_{r\to\infty} \hat{P} = 1 - \frac{5}{6} \cdot \frac{1}{2^\alpha} > \frac{1}{2}$ for all $\alpha \in \mathbb{N}$ and $\hat{P}$ is a non-increasing function of r, $\hat{P} > \frac{1}{2}$ for all r, R.

Appendix III

In this appendix, we prove that the function $g(z) = 2(b(z)-1) - |e_T + a(z)| - |e_T - a(z)|$ is symmetric on $\lceil (r-1)/2 \rceil$. To show this, we need to prove that $g(z) = g(r-1-z)$ for all $0 \leq z (\in \mathbb{Z}) \leq r-1$. Note that $a(r-1-z) = 2^{r-1-z} - 2^{r-1-(r-1-z)} = -(2^z - 2^{r-1-z}) = -a(z)$ and $b(r-1-z) = 2^{r-1-z} + 2^{r-1-(r-1-z)} = 2^z + 2^{r-1-z} = b(z)$. Thus,

$$\begin{aligned}g(r-1-z) &= 2(b(r-1-z)-1) - |e_T + a(r-1-z)| - |e_T - a(r-1-z)| \\ &= 2(b(z)-1) - |e_T - a(z)| - |e_T + a(z)| = g(z)\end{aligned}$$

which completes the proof.

APPENDIX IV

In this appendix, we show that

$$h(z) = \frac{1}{3}b(z)(b(z)-1)(b(z)-2) - \frac{1}{3}a(z)(a(z)^2-1) \tag{16}$$

is an increasing function for $z \in \mathbb{Z}$ where $r/2 < z \leq r-1$. Note that (16) is equivalent to function $h(z)$ with $z \in \mathbb{R}$ where $r/2 < z \leq r-1$, sampled at every $z \in \mathbb{Z}$. Thus, we focus on showing that $h(z)$ is an increasing function over $z \in \mathbb{R}$ where $r/2 < z \leq r-1$. To show that $h(z)$ is an increasing function, we may show that $\frac{d}{dz}h(z) > 0$ for $r/2 < z \leq r-1$. Note that

$$\frac{d}{dz}a(z) = \ln 2 \cdot (2^z + 2^{r-1-z}) = b(z)\ln 2$$

and

$$\frac{d}{dz}b(z) = \ln 2 \cdot (2^z - 2^{r-1-z}) = a(z)\ln 2.$$

Therefore,

$$\begin{aligned}\frac{d}{dz}h(z) &= \frac{\ln 2}{3}\left\{\left(3b(z)^2\frac{db(z)}{dz} - 6b(z)\frac{db(z)}{dz} + 2\frac{db(z)}{dz}\right) - \left(3a(z)^2\frac{da(z)}{dz} - \frac{da(z)}{dz}\right)\right\} \\ &= \frac{\ln 2}{3}\left\{3a(z)b(z)(b(z)-a(z)-2) + 2a(z) + b(z)\right\}.\end{aligned}$$

Since $a(z)b(z) = 2^{2z} - 2^{2(r-1-z)} > 0$ and $b(z) - a(z) = 2 \cdot 2^{r-1-z} \geq 2$ for $r/2 < z \leq r-1$,

$$\frac{d}{dz}h(z) = \frac{\ln 2}{3}\left\{3a(z)b(z)(b(z)-a(z)-2) + 2a(z) + b(z)\right\} > 0$$

which implies that $h(z)$ is an increasing function over $r/2 < z \leq r-1$.

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, Jul. 2000.

[2] 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.

[3] 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.

[4] C. Gkantsidis and P. R. Rodriguez, "Network coding for large scale content distribution," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM 2005)*, vol. 4, Mar. 2005, pp. 2235–2245.

[5] C. Gkantsidis, J. Miller, and P. Rodriguez, "Comprehensive view of a live network coding P2P system," in *Proc. ACM SIGCOMM/USENIX IMC'06*, Brasil, Oct. 2006.

[6] A. G. Dimakis, P. B. Godfrey, M. J. Wainwright, and K. Ramchandran, "Network coding for distributed storage systems," in *IEEE Int. Conf. Comput. Commun. (INFOCOM 2007)*, May 2007.

[7] S. Acedański, S. Deb, M. Médard, and R. Koetter, "How good is random linear coding based distributed networked storage?" in *Proc. Workshop Network Coding, Theory, and Applications*, Apr. 2005.

[8] S. Deb, M. Médard, and C. Choute, "On random network coding based information dissemination," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT '05)*, Adelaide, Australia, Sep. 2005, pp. 278–282.

[9] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. Allerton Conf. Commun., Control and Comput.*, Monticell, IL, USA, Oct. 2003.

[10] T. Ho, M. Médard, J. Shi, M. Effros, and D. R. Karger, "On randomized network coding," in *Proc. Allerton Annual Conf. Commun., Control, and Comput.*, Monticello, IL, USA, Oct. 2003.

[11] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[12] T. P. Coleman, E. Martinian, and E. Ordentlich, "Joint source-channel coding for transmitting correlated sources over broadcast networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3864–3868, Aug. 2009.

[13] S. L. Howard and P. G. Flikkema, "Integrated source-channel decoding for correlated data-gathering sensor networks," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC '08)*, Las Vegas, NV, USA, Apr. 2008, pp. 1261–1266.

[14] T. Ho, M. Médard, M. Effros, and R. Koetter, "Network coding for correlated sources," in *IEEE Int. Conf. Inf. Sciences and Syst. (CISS '04)*, Princeton, NJ, USA, Mar. 2004.

[15] J. Barros and S. D. Servetto, "Network information flow with correlated sources," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 155–170, Jan. 2006.

[16] A. Ramamoorthy, K. Jain, P. A. Chou, and M. Effros, "Separating distributed source coding from network coding," vol. 52, no. 6, pp. 2785–2795, Jun. 2006.

[17] Y. Wu, V. Stankovic, Z. Xiong, and S. Y. Kung, "On practical design for joint distributed source and network coding," in *Proc. First Workshop on Netw. Coding, Theory, and Applications, NetCod-05*, Riva del Garda, Italy, Apr. 2005.

[18] M. Yang and Y. Yang, "A linear inter-session network coding scheme for multicast," in *Proc. IEEE Int. Symp. Netw. Comput. and Applications*, 2008, pp. 177–184.

[19] A. Neumaier, "Solving ill-conditioned and singular linear systems: A tutorial on regularization," *SIAM Review*, vol. 40, no. 3, pp. 636–666, Sep. 1998.

[20] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[21] D. M. Bradley and R. C. Gupta, "On the distribution of the sum of n non-identically distributed uniform random variables," *J. Annals of the Institute of Statistical Mathematics*, vol. 54, no. 3, pp. 689–700, Sep. 2002.

[22] J. Kahn and J. Komlós, "Singularity probabilities for random matrices over finite fields," *Combinatorics, Probability and Computing*, vol. 10, no. 2, pp. 137–157, Mar. 2001.

[23] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell Syst. Tech. J.*, vol. 42, pp. 1977–1997, Sep. 1963.

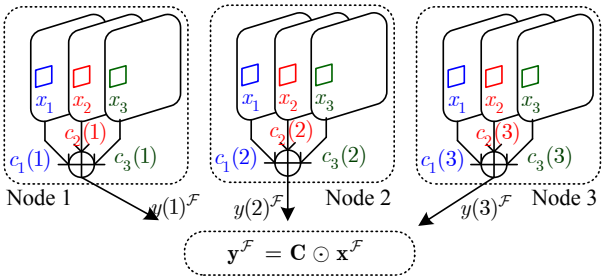


Fig. 1. Illustrative example of network coding with $N = 3$ sources and three network coding nodes. The input data x_n are linearly combined with random coefficients in each network coding node, to generate vector $\mathbf{y}^{\mathcal{F}}$.

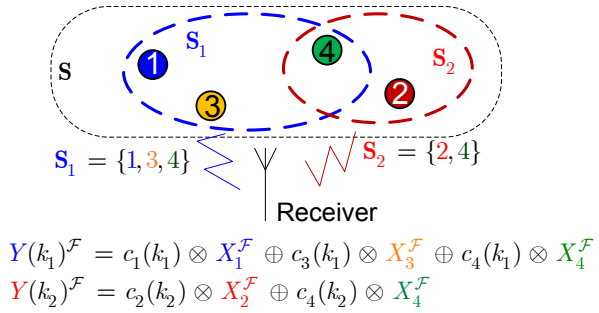


Fig. 2. Illustrative example of network coding in sensor networks.

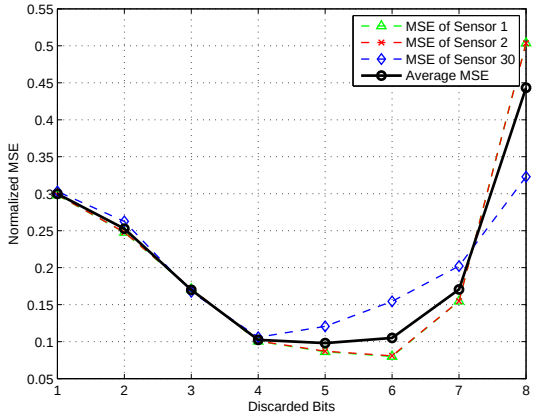
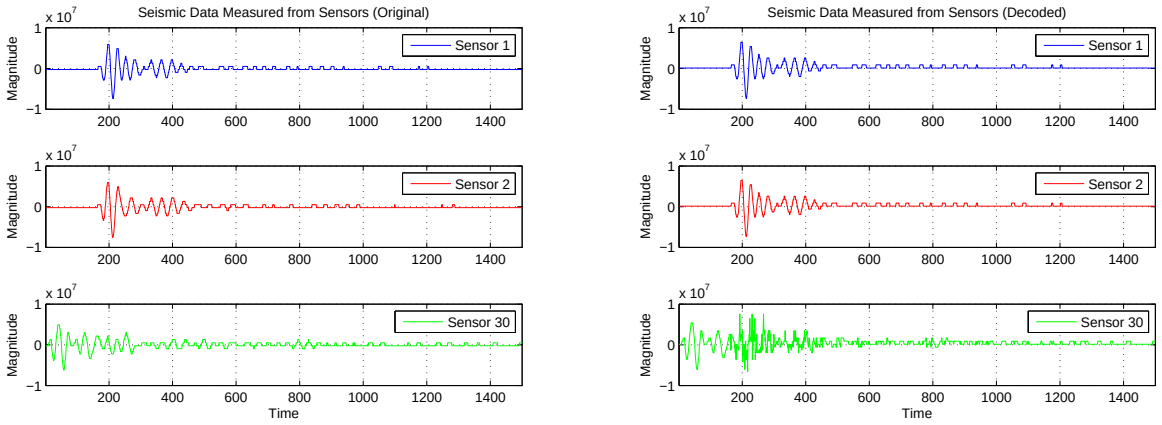


Fig. 3. Normalized average MSE for different GF sizes (i.e., GF(2^{10-z})).



(a) (b)

Fig. 4. Measured original seismic data (a) and decoded seismic data based on approximate decoding (b).

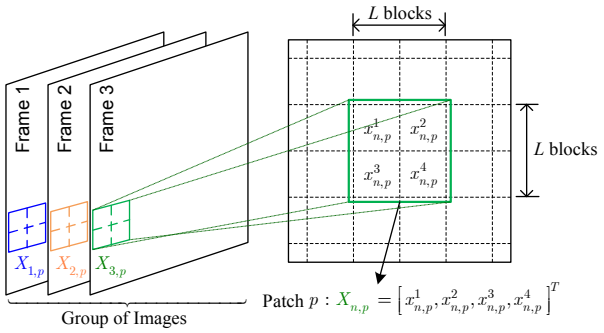


Fig. 5. Illustrative examples of patches in a group of images ($L = 2$).

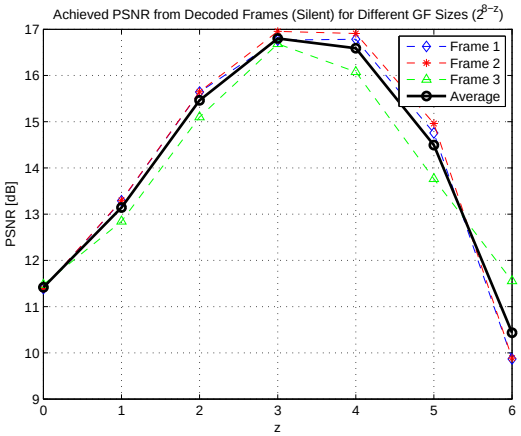


Fig. 6. Achieved PSNR for different GF sizes (i.e., GF(2^{8-z})) in the approximate decoding of the *Silent* sequence.

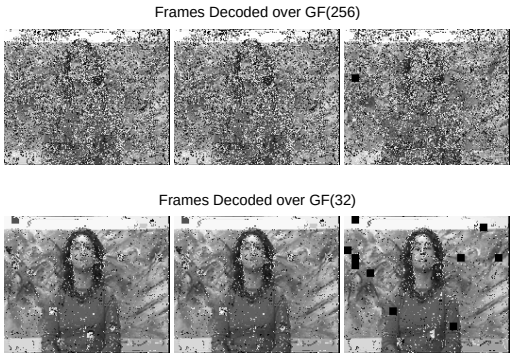


Fig. 7. Decoded frames for the *Silent* sequence for 2 different sizes of the coding field.

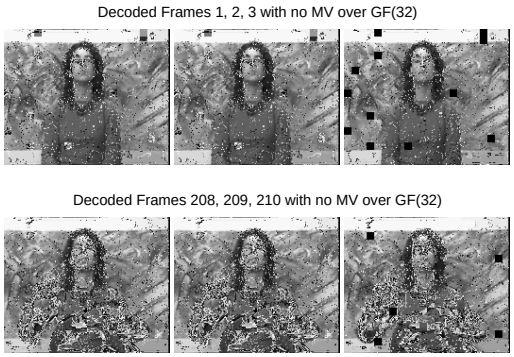


Fig. 8. Decoded frames with no information about motion estimation.

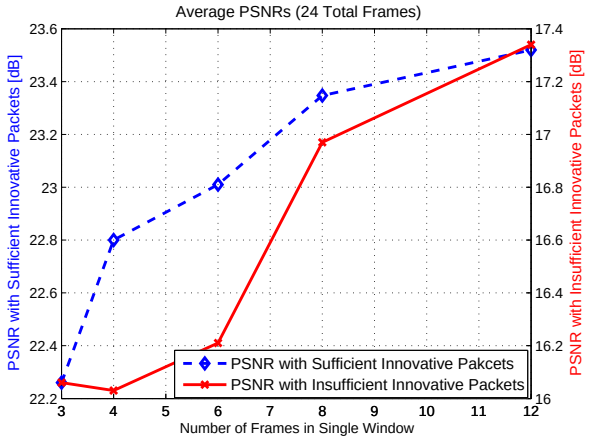


Fig. 9. Decoding PSNR for different window sizes in the encoding of the sequence *Silent*.

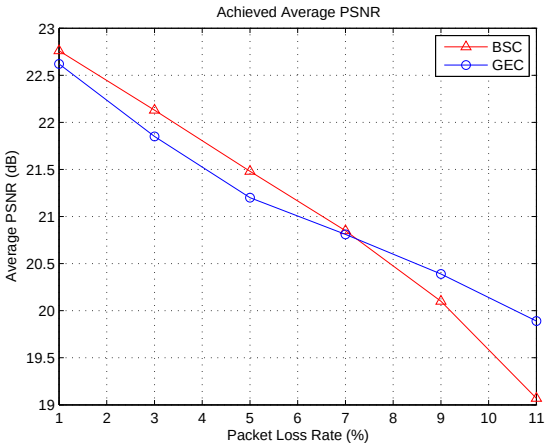
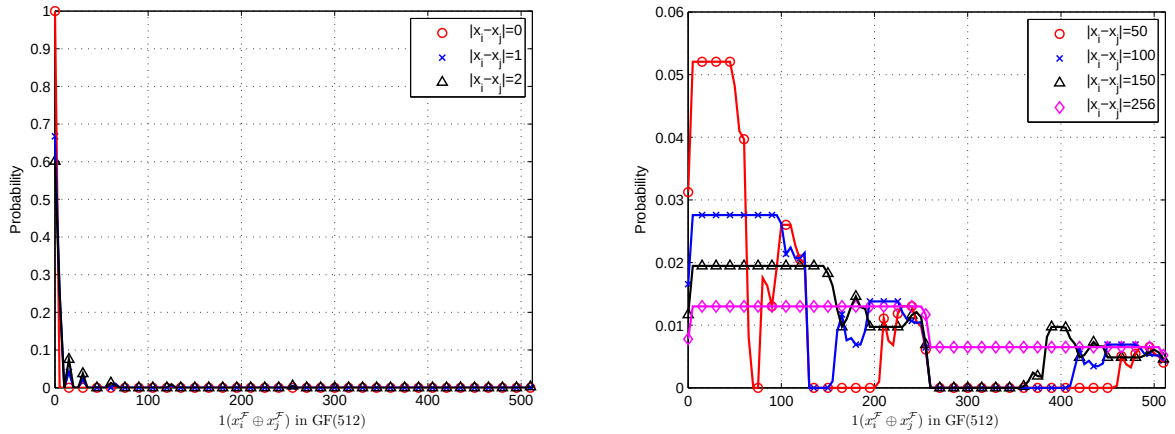


Fig. 10. Average PSNR for the transmission of *Container* QCIF sequence with respect to various packet loss rates in the BSC and in GEC. Network coding is performed on windows of four frames.



(a) (b)

Fig. 11. The probability mass function for different values of $1_{G\mathbb{R}}\left(x_i^{\mathcal{F}} \oplus x_j^{\mathcal{F}}\right)$ corresponding to various values of $|x_i - x_j|$ in GF(512).