

---

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

---

CH-1015 LAUSANNE

Telephone: +4121 6935652

Telefax: +4121 6937600

e-mail: [hyunggon.park@epfl.ch](mailto:hyunggon.park@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

November, 2009

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

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, nikolaos.thomos, pascal.frossard}@epfl.ch  
Fax: +41 21 693 7600, Phone: +41 21 693 5652

## Abstract

We consider the problem of distributed delivery of correlated data from sensors in ad hoc network topologies. We propose to use network coding in order to exploit the path diversity in the network for efficient delivery of the sensor information. We further show that the correlation between the data sources can be exploited at receivers for efficient approximate decoding when the number of received data packets is not sufficient for perfect decoding. We analyze how the decoding performance is influenced by the choice of the network coding parameters and in particular by the size of finite fields. We determine the optimal field size that maximizes the expected decoding performance, which actually represents a trade-off between information loss incurred by quantizing the source data and the error probability in the reconstructed data. Moreover, we show that the decoding performance improves when the accuracy of the correlation estimation increases. We have illustrated our network coding based algorithms with approximate decoding in sensor networks and video coding applications. In both cases, the experimental results confirm the validity of our analysis and demonstrate the benefits of our solution for distributed delivery of correlated information in ad hoc networks.

## 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 to 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 the 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., Gaussian elimination for linear combinations). Such a strategy permits to improve the throughput of the system and to approach better max-flow min-cut limit of networks [2], [3]. It provides enhanced robustness to data loss and reduces the need for coordination in the transmission of data in overlay networks compared to classical routing and scheduling algorithms. In practice, random linear network coding (RLNC) [4] is often the preferred network coding solution for the distributed delivery of time-sensitive multimedia information [5].

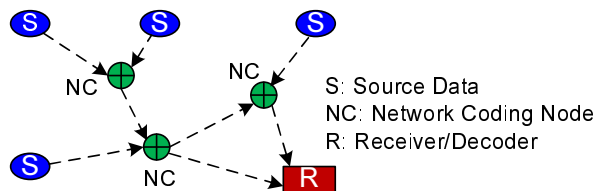


Fig. 1. A distributed data transmission system.

In this paper, we consider the distributed transmission of correlated data sources with network coding techniques, which is illustrated in Fig. 1. The transmission of correlated sources is generally studied in the framework of distributed coding [6], where sources are encoded by systematic channel encoders and eventually jointly decoded [7]–[10]. This choice, however, does not fully exploit the network diversity. Unlike the above approach, network coding is a natural solution to the transmission of correlated data over networks with diversity [11], where it leads to efficient distributed algorithms. However, due to the source and network dynamics, there is no guarantee that each node can receive enough useful packets for successful data recovery.

This becomes even more critical if applications are delay-sensitive, as the timing constraints lead to discard delayed packets. Thus, it is essential to have a methodology that enables the recovery of the original data with a good accuracy, when the number of innovative packets is not sufficient for perfect decoding.

Since the encoding and decoding processes in each node are based on linear operations (e.g., weighted linear combinations, inverse of linear matrix, etc.) in a finite algebraic field for the RLNC algorithm, the original data can be approximately recovered with help of regularization techniques such as Tikhonov regularization [12]. While regularization techniques provide a closed form solution and can be used in a general case, it may result in significantly unreasonable approximation. In this paper, we rather propose to use the correlation between the sources to design *approximate decoding* algorithms when the number of packets is insufficient. We show that the use of 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), at decoding can lead to an efficient solution for data recovery.

The information about correlation provides additional constraints in the decoding process, such that well-known approaches for matrix inversion (e.g., Gaussian elimination) can be efficiently used. We show analytically that the use of correlation leads to a better data recovery, or equivalently, that the proposed approximate decoding solution results in improved decoding performance. Moreover, we analyze the impact of the accuracy of the correlation information on the decoding performance, since the correlation information is usually obtained from estimation in practice. Our analysis shows that more accurate correlation information leads to better performance in the approximate decoding. We then analyze the influence of the network coding strategy, and in particular, of the choice of the finite field size (i.e., Galois Field (GF) size) on the performance of the approximate decoding. We demonstrate that the GF size should be selected by considering the tradeoff between source approximation and decoding performance. Specifically, the quantization error of the source data decreases with the coding GF size, while the decoding error probability increases with the field size. We show that there is an optimal value for the GF size when 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). We demonstrate the results of the finite field 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.

The paper is organized as follows. In Section II, we present our framework and describe the approximate decoding problem. 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. The related work is discussed in Section VII and the conclusions are drawn in Section VIII.

## II. PROPOSED FRAMEWORK

We describe here the framework considered in this paper and present the distributed encoding strategy for correlated data sources. We also discuss the approximate decoding algorithm that enables receivers to reconstruct the original information when the number of data packets is not sufficient for perfect decoding.

We consider an overlay network with source, intermediate and client nodes. The correlated information in the sources is transmitted to the clients via the intermediate nodes that are able to perform network coding. Let  $x_1, \dots, x_N$  be  $N$  non-negative correlated data, where  $x_n \in \mathcal{X}$  for  $1 \leq n \leq N$ . A node  $k$  produces  $y(k) = \sum_{n=1}^N c_n(k)x_n$ , which is a linear combination of its input data  $x_n$ , based on RLNC. The weights  $c_n(k)$  are randomly chosen from GF with size  $2^r$ , denoted by  $\text{GF}(2^r)$ . These coefficients are generated by each node  $k$  and transmitted to neighboring nodes. Note that, even in the case of successive coding operations on symbols, the resulting linear combinations can always be expressed in terms of the original data  $x_n$ .

We assume that the size of the input set is  $|\mathcal{X}| \leq 2^r$ . If  $K$  innovative (i.e., linearly independent) symbols or packets  $y(1), \dots, y(K)$  are available at a decoder, a linear system  $\mathbf{y} = \mathbf{C}\mathbf{x}$  can be formed as

$$\begin{bmatrix} y(1) \\ \vdots \\ y(K) \end{bmatrix} = \begin{bmatrix} c_1(1) & \dots & c_N(1) \\ \vdots & \ddots & \vdots \\ c_1(K) & \dots & c_N(K) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \quad (1)$$

where  $\mathbf{C}$  is referred as the coding coefficients matrix in this paper. An illustrative example for  $N = 3$  is shown in Fig. 2, where the symbols from different sources  $x_1$ ,  $x_2$ , and  $x_3$  are encoded by three nodes with different coefficients.

The decoder reconstructs the original data from the received symbols  $\mathbf{y}$ . If  $K = N$ ,  $\mathbf{x}$  can be uniquely determined as  $\mathbf{x} = \mathbf{C}^{-1}\mathbf{y}$  from the linear system in (1). The decoding could be performed with well-known approaches such as the Gaussian

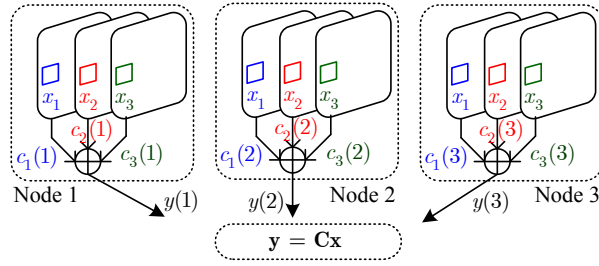


Fig. 2. Illustrative example of network coding with  $N = 3$  sources and three network coding nodes. The input data  $x_n$  are linearly combined with random coefficients in each network coding node, to generate vector  $\mathbf{y}$ .

elimination method. However, if the number of received symbols is insufficient (i.e.,  $K < N$ ), there may be an infinite number of solutions  $\hat{\mathbf{x}}$  to the system in (1), as the coding coefficient matrix  $\mathbf{C}$  is not full-rank. Hence, additional constraints need to be imposed such that the coefficient matrix becomes a full-rank matrix. This leads to approximate decoding, where the correlation of the input data can be exploited to set additional constraints  $\mathbf{D}$  in the decoding process. Thus, the new system can be formed as

$$\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{(N-K)} \end{bmatrix} \quad (2)$$

where  $\mathbf{0}_{(N-K)}$  is a vector with  $(N-K)$  zeros. The additional constraints  $\mathbf{D}$  typically depend on the nature of the problem under consideration, and on the correlation model between the input data. With the additional constraints, the linear system given in (2) can be decoded with classical algorithms such as the Gaussian elimination method in order to produce an approximation  $\hat{\mathbf{x}}$  of the original data.

We study in the next section the influence of the finite field size (GF size) in the proposed approximate decoding, when the correlation information is used to impose additional constraints  $\mathbf{D}$  in the decoding process. 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 insufficient number of innovative packets. In particular, we analyze and quantify the impact of the data correlation on the decoding performance and later 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 stated in Theorem 1.

*Theorem 1: If the correlation of the original data is used to impose additional constraints of a coding coefficient matrix in the proposed approximate decoding, the upper bound of the expected decoding error decreases as the original data is more correlated.*

*Proof:* Let  $\mathbf{y}$  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 need to be imposed into the coding coefficient matrix based on the correlation among the original data, in order for the system be decodable.

Let  $\hat{\mathbf{x}} = [\hat{x}_1, \dots, \hat{x}_N]^T$  denote the approximation of the input data when  $(N-K)$  additional constraints are imposed as  $\mathbf{D}$  to the system. Thus,  $\hat{\mathbf{x}}$  can be obtained as

$$\hat{\mathbf{x}} = \begin{bmatrix} \mathbf{C} \\ \mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{(N-K)} \end{bmatrix}. \quad (3)$$

The correlation information is used to build additional constraints, represented by  $\mathbf{D}$ , which is a  $(N-K) \times N$  matrix of coefficients that depends on the correlation between data. Without loss of generality, we construct such a matrix with each row consisting of zeros (i.e., additive identity of  $\text{GF}(2^r)$ ) except two elements of value '1' and '-1' that correspond to the positions of the best matched data  $x_i, x_j \in \mathcal{X}$ .<sup>1</sup>

We now analyze the error incurred by approximate decoding compared to the perfect decoding with a full rank system. The approximate solution  $\hat{\mathbf{x}}$  in (3) is compared to the exact solution  $\mathbf{x}$  that would be produced by the same coefficient matrix, but with the actual vector  $\mathbf{d}$  created with the set of additional constraints, i.e.,  $\mathbf{d} = \mathbf{D}\mathbf{x} = [d(1), \dots, d(N-K)]^T$ . Therefore,  $\mathbf{x}$  can be expressed as

$$\mathbf{x} = \begin{bmatrix} \mathbf{C} \\ \mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{y} \\ \mathbf{d} \end{bmatrix}. \quad (4)$$

<sup>1</sup>Note that the element '-1' denotes the additive inverse in  $\text{GF}(2^r)$ . Hence, '-1' can be replaced with '1' in actual coding operations.

Using (3) and (4), the error between the exact and approximate solutions can thus be expressed as

$$\|\mathbf{x} - \hat{\mathbf{x}}\| = \left\| \begin{bmatrix} \mathbf{C} \\ \mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_K \\ \mathbf{d} \end{bmatrix} \right\|. \quad (5)$$

We assume that the coding coefficient matrix  $[\mathbf{C} \ \mathbf{D}]^T$  in (3) is not singular, and hence, its inverse can be written as  $[\mathbf{C} \ \mathbf{D}]^{-T} = [\mathbf{q}_1, \dots, \mathbf{q}_K, \mathbf{q}_{K+1}, \dots, \mathbf{q}_N]$ , where  $\mathbf{q}_n$  is  $n$ th column vector. Thus,

$$\begin{aligned} E(\|\mathbf{x} - \hat{\mathbf{x}}\|) &= E\left(\left\| [\mathbf{q}_1, \dots, \mathbf{q}_K, \mathbf{q}_{K+1}, \dots, \mathbf{q}_N] \begin{bmatrix} \mathbf{0}_K \\ \mathbf{d} \end{bmatrix} \right\|\right) \\ &= E\left(\left\| \sum_{l=1}^{N-K} \mathbf{q}_{K+l} d(l) \right\|\right) \\ &\leq E\left(\sum_{l=1}^{N-K} \|\mathbf{q}_{K+l} d(l)\|\right) \\ &= \sum_{l=1}^{N-K} E(|d(l)|) \|\mathbf{q}_{K+l}\|. \end{aligned} \quad (6)$$

We can now analyze the influence of the correlation in the decoding performance. Recall that each element in  $\mathbf{d}$  depends on the two non-zero elements in  $\mathbf{D}$  due to our choice of the additional constraints. Thus,  $d(l) = x_i - x_j$ , where  $0 \leq i, j \leq N$ . Let  $\rho_{i,j}$  denote the correlation coefficient between  $x_i$  and  $x_j$ , where  $0 \leq \rho_{i,j} \leq 1$ . By definition,

$$\begin{aligned} \rho_{i,j} &= \frac{E(x_i x_j) - E(x_i)E(x_j)}{\sigma_i \sigma_j} \\ &= \frac{E(x_i x_j) - E(x_i)E(x_j)}{\sqrt{E(x_i^2) - E(x_i)^2} \sqrt{E(x_j^2) - E(x_j)^2}} \end{aligned} \quad (7)$$

which implies that

$$E(x_i x_j) = \rho_{i,j} \sigma_i \sigma_j + E(x_i)E(x_j). \quad (8)$$

The expectation  $E(d(l)^2)$  can be expressed as a function of  $\rho_{i,j}$  using (8):

$$\begin{aligned} E(d(l)^2) &= E(|d(l)|^2) = E(|x_i - x_j|^2) = E(x_i^2 - 2x_i x_j + x_j^2) \\ &= E(x_i^2) + E(x_j^2) - 2E(x_i x_j) \\ &= \{E(x_i) - E(x_j)\}^2 + \sigma_i^2 + \sigma_j^2 - 2\rho_{i,j} \sigma_i \sigma_j \end{aligned} \quad (9)$$

which implies that  $E(d(l)^2)$  is a decreasing function of  $\rho_{i,j}$ . Since  $\sigma_{|d(l)|}^2 = E(d(l)^2) - E(|d(l)|)^2 \geq 0$ , we have

$$E(|d(l)|) \leq \sqrt{E(d(l)^2)} \quad (10)$$

which is also a decreasing function of  $\rho_{i,j}$ . Based on (6) and (10), we finally have

$$E(\|\mathbf{x} - \hat{\mathbf{x}}\|) \leq \sum_{l=1}^{N-K} \sqrt{E(d(l)^2)} \cdot \|\mathbf{q}_{K+l}\|. \quad (11)$$

Since the coding coefficient matrix is given (i.e.,  $\mathbf{q}_{K+l}$  is fixed) and higher correlation between  $x_i$  and  $x_j$  results in a higher value of  $\rho_{i,j}$ , the upper bound of the expected error shown in (6) decreases when the data correlation increases. ■

Theorem 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.

Note that the coding coefficient matrix in (3) is assumed to be non-singular in order to quantify how the correlation can influence the decoding performance. 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.

We further analyze the influence of the accuracy of the data correlation estimation. The analysis in Theorem 1 is based on the assumption that the true correlation  $\rho_{i,j}$  is available at the decoder. In practice, however, the correlation coefficients are often estimated from (possibly imperfect) observations or correlation models. Thus, the true correlation between the data is not exactly known. We therefore introduce an estimation noise in the correlation information and investigate its impact on the decoding error.

*Corollary 1: If the correlation coefficients are more accurately estimated, the performance of the approximate decoding improves.*

*Proof:* Let  $x'_i$  and  $x'_j$  be estimations of  $x_i$  and  $x_j$  with an estimation 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

$$\begin{aligned} \rho'_{i,j} &= \frac{E(x'_i x'_j) - E(x'_i)E(x'_j)}{\sqrt{E(x_i'^2) - E(x_i')^2} \sqrt{E(x_j'^2) - E(x_j')^2}} \\ &= \frac{E((x_i + n_i)(x_j + n_j)) - E(x_i + n_i)E(x_j + n_j)}{\sqrt{E((x_i + n_i)^2) - E(x_i + n_i)^2} \sqrt{E((x_j + n_j)^2) - E(x_j + n_j)^2}} \\ &= \frac{E(x_i x_j) - E(x_i)E(x_j)}{\sqrt{E(x_i^2) - E(x_i)^2 + \sigma_{n_i}^2} \sqrt{E(x_j^2) - E(x_j)^2 + \sigma_{n_j}^2}} \leq \rho_{i,j} \end{aligned} \quad (12)$$

because  $\sigma_{n_k}^2 \geq 0$  for  $k \in \{i, j\}$ . This implies that the estimation noise results in a smaller correlation coefficient than the true correlation coefficient. Since higher correlation coefficient provides a better performance (i.e., a smaller upper-bound for the expected decoding error) as shown in Theorem 1, more accurate correlation coefficients provide better performance. ■

Finally, 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. Then, we 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 proposition, which states that the probability of large decoding errors increases with the GF size.

*Proposition 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 vectors when coding is performed in respectively  $\text{GF}(2^r)$  and  $\text{GF}(2^R)$  with  $R > r$  for  $r, R \in \mathbb{N}$ , i.e.,  $\text{GF}(2^R)$  is an extended GF from  $\text{GF}(2^r)$ . We assume that approximate decoding generate solutions that are uniformly distributed over  $\mathcal{X}$ , i.e., 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 show that

$$\Pr(|x - \hat{x}_R| \geq |x - \hat{x}_r|) > \frac{1}{2}. \quad (13)$$

The left hand side of (13) can be expressed as

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

because  $\hat{x}_R$  and  $\hat{x}_r$  are both uniformly distributed. Note that

$$\hat{P} \triangleq \Pr\left(x \leq \frac{\hat{x}_R + \hat{x}_r}{2} \mid \hat{x}_R \geq \hat{x}_r\right) > \frac{1}{2} \quad (14)$$

as shown in Appendix I. Therefore,

$$2^{r-R-1} + (1 - 2^{r-R}) \hat{P} > 2^{r-R-1} + (1 - 2^{r-R}) \cdot \frac{1}{2} = \frac{1}{2} \quad (15)$$

which completes the proof. ■

Proposition 2 implies that a small GF size is preferable in terms of expected decoding error, i.e., it is not preferred to enlarge the GF size more than the size of the input space. However, if the GF size becomes smaller than the size of the input space,

the maximum number of data 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  $|\mathcal{X}'| \leq 2^{r'}$ . Then, all the data in  $\mathcal{X}'$  can be distinctly encoded in  $\text{GF}(2^{r'})$ . Such a choice introduces some error at the source, similarly to data quantization. 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 Theorem 3 that shows the existence of an optimal GF size.

*Theorem 3: There exists an optimal GF size that minimizes the expected decoding error. Moreover, the optimal GF size,  $\text{GF}(2^{r-z^*})$ , is determined by  $z^* = \lceil (r-1)/2 \rceil$  and  $z^* = \lfloor (r-1)/2 \rfloor$ .*

*Proof:* Suppose that  $|\mathcal{X}| = 2^r$  and  $\text{GF}(2^r)$ . As discussed in Proposition 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. We assume that if the GF size is reduced from  $\text{GF}(2^r)$  to  $\text{GF}(2^{r-z})$ , where  $0 \leq z \in \mathbb{Z} \leq r-1$ , the least significant  $z$  bits are first discarded from  $x \in \mathcal{X}$ . Moreover, we assume that the corresponding source information loss is uniformly distributed and that the decoded data is also uniformly distributed.

If GF size is reduced from  $\text{GF}(2^r)$  to  $\text{GF}(2^{r-z})$  the decoding errors are uniformly distributed over  $[-r_D, r_D]$ , where  $r_D = 2^{r-1-z} - 1$ , i.e.,

$$p_{e_D}(e_D) = \begin{cases} 1/(2r_D + 1), & \text{if } e_D \in [-r_D, r_D] \\ 0, & \text{otherwise} \end{cases}. \quad (16)$$

Correspondingly,  $\mathcal{X}$  is reduced to  $\mathcal{X}'$ , where  $|\mathcal{X}'| = 2^{r-z}$  by discarding the  $z$  least significant bits from all  $x \in \mathcal{X}$ . This information loss also results in errors over  $[-r_I, r_I]$ , where  $r_I = 2^z - 1$ , i.e.,

$$p_{e_I}(e_I) = \begin{cases} 1/(2r_I + 1), & \text{if } e_I \in [-r_I, r_I] \\ 0, & \text{otherwise} \end{cases}. \quad (17)$$

Based on these 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 [13]

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

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

$$p_{e_T}(e_T) = \frac{H}{2} \{ 2(2^z + 2^{r-1-z} - 1) - |e_T + 2^z - 2^{r-1-z}| - |e_T - 2^z + 2^{r-1-z}| \}. \quad (18)$$

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

$$E[|e_T|] = \sum_{e_T=-\infty}^{\infty} |e_T| \cdot p_{e_T}(e_T) = \sum_{e_T=-e_T^{max}}^{e_T^{max}} \frac{H}{2} |e_T| \cdot [2(b(z) - 1) - |e_T + a(z)| - |e_T - a(z)|]. \quad (19)$$

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 II),  $E[|e_T|]$  is also symmetric. Thus,

$$\begin{aligned} E[|e_T|] &= H \sum_{e_T=1}^{e_T^{max}} e_T \cdot \{ 2(b(z) - 1) - |e_T + a(z)| - |e_T - a(z)| \} \\ &= H \sum_{e_T=1}^{e_T^{max}} e_T \cdot \{ 2(b(z) - 1) \} - H \sum_{e_T=1}^{e_T^{max}} e_T \cdot \{ |e_T + a(z)| + |e_T - a(z)| \} \\ &= H \cdot (b(z) - 1) e_T^{max} (e_T^{max} + 1) - H \sum_{e_T=1}^{e_T^{max}} e_T \cdot \{ |e_T + a(z)| + |e_T - a(z)| \}. \end{aligned} \quad (20)$$

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

$$\begin{aligned} E[|e_T|] &= H \cdot \left[ (b(z) - 1)^2 (b(z) - 2) - \frac{1}{3} (b(z) - 1) (b(z) - 2) (2b(z) - 3) - \frac{1}{3} a(z) (a(z)^2 - 1) \right] \\ &= H \cdot \left[ \frac{1}{3} b(z) (b(z) - 1) (b(z) - 2) - \frac{1}{3} a(z) (a(z)^2 - 1) \right] \end{aligned} \quad (21)$$

which is an increasing function for  $r/2 < z \leq r-1$  (see Appendix III). Since  $E[e_T]$  is a 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.

Moreover, 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 when  $z^* = \lceil (r-1)/2 \rceil$  and  $z^* = \lfloor (r-1)/2 \rfloor$ . ■

An illustrative example is given in Fig. 3, which confirms that the optimal value of  $z^*$  is found at  $z = \lfloor (r-1)/2 \rfloor = 3$  and  $z^* = \lceil (r-1)/2 \rceil = 4$ , as discussed above.

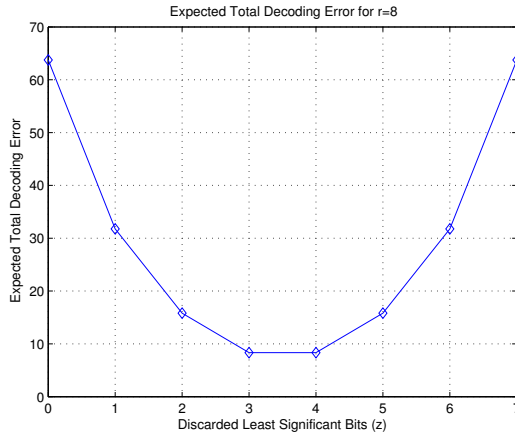


Fig. 3. Total decoding error for  $r = 8$ , or  $\text{GF}(2^8)=\text{GF}(256)$ .

## 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 that adopts the RLNC scheme for data delivery over an ad hoc network. Each sensor measures its own observations and receives the observations from its neighbor sensors. Then, each sensor combines the received data and its own data with RLNC and transmits the resulting information to receivers.

Specifically, we analyze a scenario, where seismic signals captured by sensors placed at a distance of 100m by each other are encoded based on RLNC and transmitted to relay nodes or receivers. We consider that a sensor  $h$  captures a signal  $X_h$  that represents a series of sampled values in a window of size  $w$ , i.e.,  $X_h = [x_h^1, \dots, x_h^w]^T$ . 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 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  $\mathbf{S}_n (\subseteq \mathbf{S})$  be 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 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  $k$ th encoded data packets for a window are denoted by  $Y(k) = \sum_{h \in \mathbf{S}_n} c_h(k) X_h$ . An illustrative example is shown in Fig. 5. 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  $Y(k_1)$  and  $Y(k_2)$ .

When a receiver collects enough innovative packets, it can solve the linear system given as in (4) and recover the original data. If the number of packets is not sufficient, however, the receiver implements an approximate decoding strategy that exploits



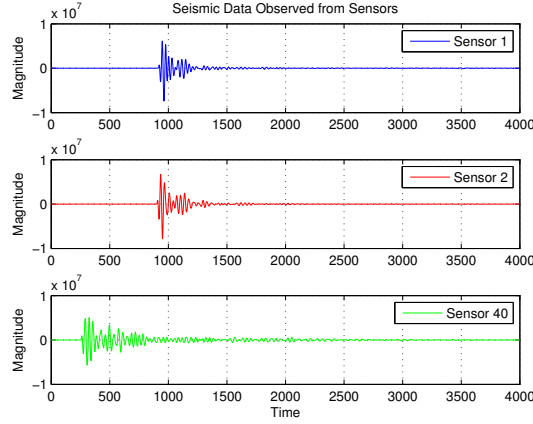


Fig. 4. Seismic signals captured at different spatial locations.

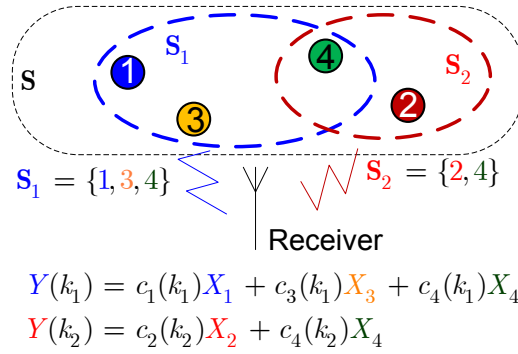


Fig. 5. Illustrative example of network coding in sensor networks.

the correlation between the different signals. With such a strategy, the decoding performance can be improved as discussed in Theorem 1. We assume that the system setup is approximately known by the sensors. In other words, the sensors can estimate correlation models that include the relative temporal shifts and energy scaling between the signals from the different sensors.<sup>2</sup>

### B. Experiment 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.,  $\text{GF}(2^{10})$ . 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. 6 shows the MSE 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 Theorem 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. 7, where a  $\text{GF}(2^{10-z^*}) = \text{GF}(2^5)$  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. 7(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 more 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 correlate.

<sup>2</sup>Note that The data can also be communicated with each other sensor.

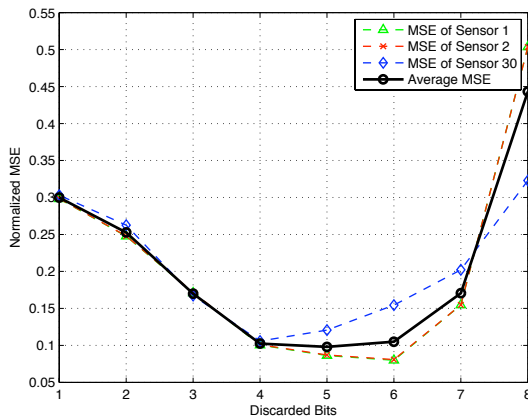


Fig. 6. Normalized average mean square error (MSE) for different GF sizes (i.e.,  $\text{GF}(2^{10-z})$ ).

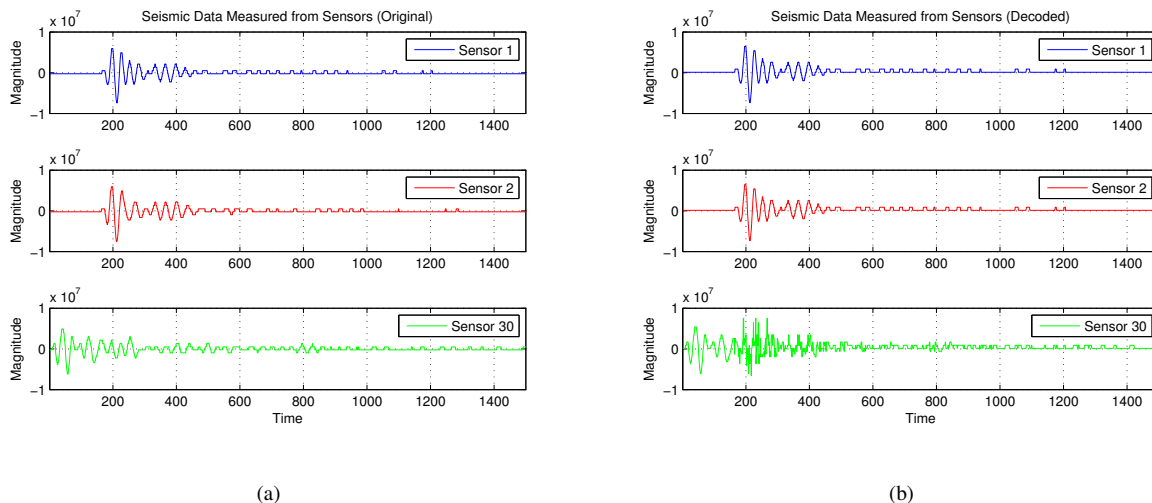


Fig. 7. Measured original seismic data (a) and decoded seismic data based on approximate decoding (b).

## 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 might however be lost or delayed in the delivery of image sequence streams, which prevents a perfect reconstruction of the images. However, the correlation between successive frames can be used for improved decoding performance.

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.,  $X_n = [X_{n,1}, \dots, X_{n,N}]$ . A patch  $X_{n,p}$  contains  $L \times L$  blocks of pixels  $x_{n,p}^b$ ,  $1 \leq b \leq L^2$ , i.e.,  $X_{n,p} = [x_{n,p}^1, \dots, x_{n,p}^{L \times L}]$ . Such a representation is illustrated on Fig. 8. The system implements RLNC and combines patches at similar positions in different frames to produce encoded symbols. In other words, it produces a series of symbols  $Y_p(k) = \sum_{n=1}^N c_{n,p}(k) X_{n,p}$  for a location of patch  $p$ . The coding coefficients  $c_{n,p}(k)$  are randomly chosen in  $\text{GF}(2^r)$ . We assume that the original data (i.e., pixels) has 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. This constraint typically depends 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

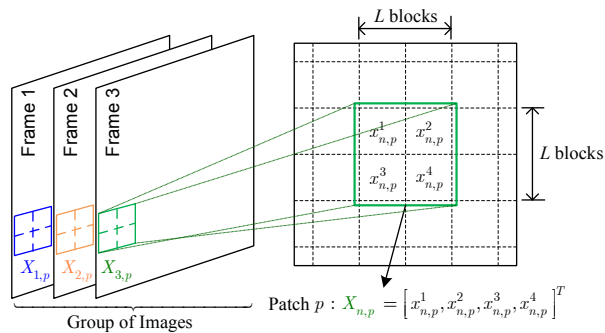


Fig. 8. Illustrative examples of patches in a group of images ( $L = 2$ ).

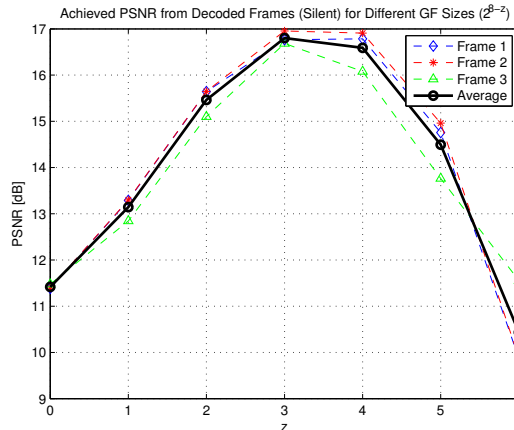


Fig. 9. Achieved PSNR for different GF sizes (i.e.,  $\text{GF}(2^{8-z})$ ) in the approximate decoding of the *Silent* sequence.

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 typically depends on the motion in the image sequence. When motion information permits to add enough additional constraints to the decoding system, 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 corresponds to patches.

### B. Experiment Results

In our experiments, we consider three consecutive frames extracted from *Silent* standard MPEG sequence with QCIF format ( $174 \times 144$ ). The patches are constructed with four blocks of  $8 \times 8$  pixels. We assume that only  $2/3$  of the linear equations required for perfect decoding are received. The decoder implements approximate decoding by assuming that the correlation information is known at the decoder. Additional constraints are added to the decoding system based on the best matched pairs of blocks in consecutive frames, in the sense of the smallest sum of absolute differences of 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  $\text{GF}(2^8)$  to  $\text{GF}(2^{8-z})$ . We reduce the size of the field by discarding the  $z$  least significant bits for each pixel. Fig. 9 shows the PSNR (Peak Signal to Noise Ratio) quality achieved from the decoded frames for different number of discarded bits  $z$ . As discussed in Theorem 3, the expected decoding error can be minimized if  $z^* = \lceil (r-1)/2 \rceil$  or  $z^* = \lfloor (r-1)/2 \rfloor$ , 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 are presented in Fig. 10 for two different sizes of the coding field. We can observe in the decoded images that several patches that are completely black or white. This is because the Gaussian elimination fails during the decoding process, which is the consequence of singular coding coefficient submatrix.

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 imposes that the consecutive frames are identical. Fig. 11 shows the frames decoded with no motion over  $\text{GF}(32)$ . 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 more

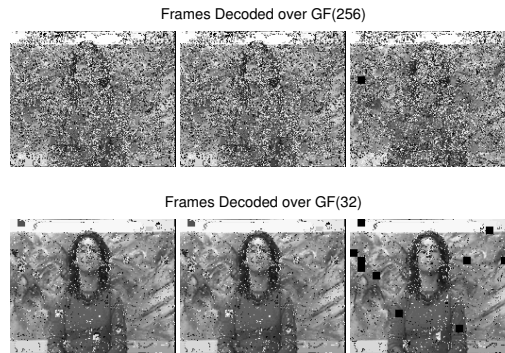


Fig. 10. Decoded frames for the *Silent* sequence for 2 different sizes of the coding field.

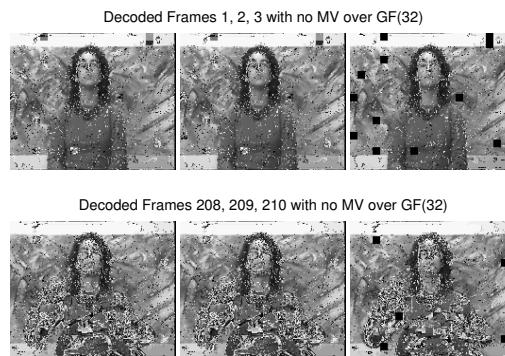


Fig. 11. Decoded frames with no information about motion estimation.

motion is included, we clearly observe significant performance degradation, especially in the positions where high motion exists.

We finally study 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 [14]. Thus, if the group of images (i.e., window size) becomes larger, the coding coefficient matrix becomes large and the probability that Gaussian elimination fails correspondingly becomes 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 contain 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. 12. 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. 12 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 is the

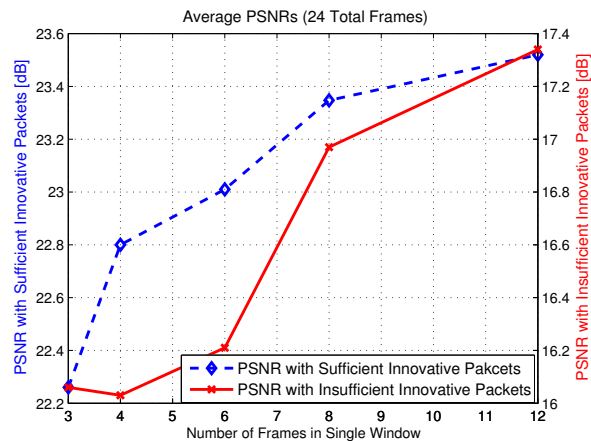


Fig. 12. Decoding PSNR for different window sizes in the encoding of the sequence *Silent*.

lowest when window size is 4. 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.

## VII. RELATED WORK

We describe in this section the work related to network coding and to the delivery of correlated sources. Network coding has recently been proposed as an efficient way to exploit network diversity for data transmission. There have been many works on the different aspects on network coding and we provide here only a short overview of some of the most relevant papers.

Several research works have first focused on the choice of the coding coefficients in network coding systems. The coefficients might be determined in a centralized manner [3], but at the price of large communication overheads between the central node and the users in the system. To reduce the communication overheads, random linear network coding (RLNC) is introduced in [4]. Although the RLNC may incur performance degradation compared to linear network coding schemes such as [3], this degradation often stays negligible [15]. It certainly positions RLNC as one of the most interesting coding solutions for the deployment of practical applications based on network coding.

The network coding schemes have been used in various applications where network diversity can provide important benefits. For example, in [16], [17], a new protocol is devised for a large scale content distribution based on RLNC. It proposes a simple system that enables users to locate neighboring users with novel information that they have not already received. This protocol outperforms BitTorrent-like protocols [18] that are the most popular for current peer-to-peer (P2P) content distribution. Network coding is also deployed for efficient and robust data dissemination. For example, in [15], the RLNC is used with gossiping protocols, where the protocols outperform significantly the systems that store and forward data. Moreover, several approaches for distributed storage are developed based on network coding [19], [20], where the network coding enables the applications to reliably store data across a set of unreliable nodes in a distributed way. It is interesting to note that most of these works have focused on theoretical aspects of network coding such as achievable capacity and coding gain, as well as its practical aspects such as robustness *when* the number of innovative packets is *sufficient* for perfect decoding. Hence, these solutions only provide limited performance if receivers have only small number of innovative packets.

Network coding of correlated sources in the context of Slepian-Wolf problem [6] has been studied in [7], [8] where the sources are distributed encoding and jointly decoded. Slepian-wolf coding by means of joint source and channel coding of correlated sources is shown that it can reduce the decoding error probability [9], [10]. The above distributed coding schemes [7]–[10] assume that the sources are encoded by systematic channel encoders which is quite restricting assumption, especially in sensor networks as it may reduce the possible gains from network diversity. These schemes target to data transmission with errors, while we consider packet erasures. It worth noting that differently from [7]–[10], we focus on cases, where the received data is not sufficient to permit errorless decoding of the original information.

In practice, it is actually not guaranteed that each node receives sufficient innovative packets for successful data recovery. This becomes even more critical if applications are delay-sensitive, as they have stringent time constraints for data delivery. Thus, it is essential to have a methodology that provides a good approximation of the original data, when the number of innovative packets is not sufficient for perfect decoding. A simple but widely used technique is to find the approximation that minimizes the norm of the error in decoding the system with the help of the pseudo-inverse of the encoding matrix. While this regularization technique provides a closed form solution and can be deployed in a general case, it may result in significantly unreasonable approximation [12]. 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. An alternative approach that exploits the correlation for data recovery is studied in compressive sensing, where the

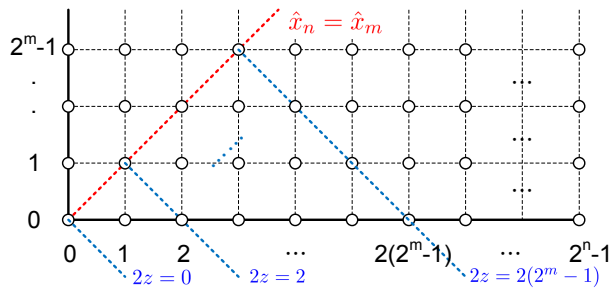


Fig. 13.

original data can be approximately recovered from a small set of equations, under sparsity assumption [21]. However, such approaches are not applicable to the problem considered in this paper, where coding operations are performed in finite fields on data that are not necessarily sparse.

### VIII. CONCLUSIONS

In this paper, we have described a framework for the delivery of correlated information sources with help of network coding and approximate decoding based on correlation information. 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 coding applications, where the experimental results confirm the results of our analytical study, as well as the benefits of approximate decoding solutions.

### IX. 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 show that  $\hat{P} \geq \frac{1}{2}$ , defined in Eq. (14). Using Bayes' rule,

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

Referring to Fig. 13, we have

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

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

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

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

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

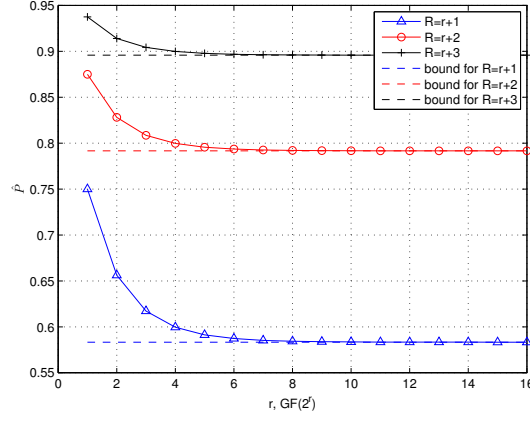


Fig. 14.  $\hat{P}$  for different  $r$  and  $R$ .

Since

$$\lim_{r \rightarrow \infty} \hat{P} = 1 - \frac{5}{6} \cdot \frac{1}{2^\alpha} > \frac{1}{2} \quad \text{for all } \alpha \in \mathbb{N}$$

and  $\hat{P}$  is a non-increasing function of  $r$ ,  $\hat{P} > \frac{1}{2}$  for all  $r, R$ . Illustrative examples for  $R = r + 1$ ,  $R = r + 2$  and  $R = r + 3$  are shown in Fig. 14.

#### APPENDIX II

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 III

In this appendix, we show that

$$h(z) = \frac{1}{3}b(z)(b(z) - 1)(b(z) - 2) - \frac{1}{3}a(z)(a(z)^2 - 1) \quad (22)$$

is an increasing function for  $z \in \mathbb{Z}$  where  $r/2 < z \leq r-1$ . Note that Eq. (22) 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,

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

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  $r/2 < z \leq r - 1$ ,

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

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] 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.
- [5] P. A. Chou, Y. Wu, and K. Jain, "Practical network coding," in *Proc. Allerton Conf. Commun., Control and Comput.*, Monticello, IL, USA, Oct. 2003.
- [6] T. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] A. Neumaier, "Solving ill-conditioned and singular linear systems: A tutorial on regularization," *SIAM Review*, vol. 40, no. 3, pp. 636–666, Sep. 1998.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] B. Cohen, "Incentives build robustness in BitTorrent," in *Proc. P2P Economics Workshop*, Berkeley, CA, 2003.
- [19] 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.
- [20] 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.
- [21] 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.