

# MAP Detection and Robust Lossy Coding Over Soft-Decision Correlated Fading Channels

Shervin Shahidi, *Student Member, IEEE*, Fady Alajaji, *Senior Member, IEEE*, and Tamás Linder, *Fellow, IEEE*

**Abstract**—The joint source–channel coding (JSCC) problem for soft-decision-demodulated time-correlated fading channels is investigated without the use of channel coding and interleaving. For the purpose of system design, the recently introduced nonbinary noise discrete channel with queue-based (QB) noise (NBND-CB) is adopted. This analytically tractable Markovian model has been shown to represent effectively correlated fading channels that are hard to handle analytically. Optimal sequence maximum *a posteriori* (MAP) detection of a discrete Markov source sent over the NBND-CB is first studied. When the Markov source is binary and symmetric, a necessary and sufficient condition under which the MAP decoder is reduced to a simple instantaneous symbol-by-symbol decoder is established. Two robust lossy source coding schemes with a low encoding delay are next proposed for the NBND-CB. The first scheme consists of a scalar quantizer (SQ), a proper index assignment, and a sequence MAP decoder designed to harness the redundancy left in the indexes of the quantizer and the soft-decision output and the noise correlation of the channel. The second scheme is a classical noise-resilient vector quantizer known as the channel-optimized vector quantizer (COVQ). It is demonstrated that both systems can successfully exploit the memory and soft-decision information of the channel. Signal-to-distortion-ratio (SDR) gains of more than 1.7 dB are obtained over hard-decision demodulation by using only two bits for soft decision. Furthermore, gains as high as 4.4 dB can be achieved for a strongly correlated channel, in comparison with systems designed for the ideally interleaved (memoryless) channel. Finally, it is numerically observed that, for low coding rates, the NBND-CB model can accurately approximate discrete fading channels (DFCs) in terms of SDR performance.

**Index Terms**—Channel modeling, channel-optimized vector quantization (COVQ), correlated fading channels, joint source–channel coding (JSCC), maximum *a posteriori* (MAP) decoding, Markov channels, scalar quantization (SQ), soft-decision demodulation.

Manuscript received July 20, 2012; revised November 18, 2012; accepted February 22, 2013. Date of publication March 11, 2013; date of current version September 11, 2013. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada. This paper was presented in part at the Canadian Workshop on Information Theory, Kelowna, BC, Canada, May 17–20, 2011 [1] and at the IEEE International Conference on Communications, Ottawa, ON, Canada, June 10–15, 2012 [2]. The review of this paper was coordinated by Prof. G. Bauch.

S. Shahidi was with Queen's University, Kingston, ON K7L 3N6, Canada. He is currently with the Department of Electrical and Computer Engineering, University of Toronto, ON M5S 1A1, Canada (e-mail: sshahidi@comm.utoronto.ca).

F. Alajaji is with the Department of Mathematics and Statistics and the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: fady@mast.queensu.ca).

T. Linder is with the Department of Mathematics and Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: linder@mast.queensu.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TVT.2013.2252213

## I. INTRODUCTION

IT IS well known that the separate treatment of source and channel coding, which is justified by Shannon's source–channel coding separation theorem [3], is not optimal in the presence of complexity and delay constraints. For example, in a recent information theoretic study [4], it has been shown that the error exponent for joint source–channel coding (JSCC) can be twice as large as the exponent for separate source and channel coding. Hence, for an identical overall probability of error, JSCC would need half the (encoding and decoding) delay of separate coding; this translates into a 2-dB power saving for a wide class of source–channel pairs [4]. Such substantial potential benefits provide an incentive for adopting and integrating JSCC in today's resource-strapped wireless communication systems. Some other advantages of JSCC over separate source and channel coding were quantitatively characterized in [5]. For lossy coding, a variety of different JSCC schemes has been proposed (such as [6]–[14] and many others).<sup>1</sup> It is also known that, if a channel is well behaved (ergodic) and has memory, then its capacity is strictly greater than the capacity of its memoryless counterpart (a channel with an identical 1-D transition distribution) realized via ideal (infinite depth) block interleaving [17], [18]. Consequently, a communication system can be designed to take advantage of the memory of a channel and perform better than a system that discards such memory via interleaving. Furthermore, effective use of the soft-decision information of a channel can improve capacity and system performance over hard-decision-decoded schemes (e.g., see [19]–[22]).

In this paper, we investigate the JSCC problem for soft-decision-demodulated time-correlated fading channels. Our objective is the design of effective schemes having low encoding delay and complexity that aptly exploit the source statistics and both the soft-decision information and statistical memory of a channel without the use of channel error-correcting codes and channel interleaving. Such schemes have pertinent applications in wireless communications, including mobile radio and sensor networks, where the mobile or sensor has stringent processing and encoding delay constraints, whereas the fusion center or the base station has sufficient decoding resources. For this purpose, we use the recently introduced nonbinary noise discrete channel with queue-based (QB) noise (NBND-CB), which is a binary input  $2^q$ -ary output channel (where  $q \geq 1$  is an integer)

<sup>1</sup>In this paper, we focus on lossy source codes that are resilient against channel noise. There are other JSCC approaches; see for example [4], [15], [16] and the references therein.

with  $2^q$ -ary stationary ergodic  $M$ th-order Markov noise in [23]–[25]. This model features closed-form expressions for its transition probabilities, noise entropy rate, and autocorrelation function, making it amenable for tractable analytical performance analysis and code design. It is also shown in [24] and [25] that the NBNDC-QB can accurately represent (in terms of channel capacity and noise autocorrelation function) the correlated Rayleigh discrete fading channel (DFC) used with antipodal signaling and  $q$ -bit soft-decision (nonbinary) output quantization.

We first study the optimal sequence maximum *a posteriori* (MAP) detection problem when a discrete Markov source sent over the NBNDC-QB. We translate the MAP detection problem into simple Viterbi decoding with an appropriately modified decoding metric. If the Markov source is binary and symmetric, we prove a necessary and sufficient condition under which the sequence MAP detector reduces to an instantaneous symbol-by-symbol mapping (having no decoding delay). We also show and verify this condition numerically.

To exploit the memory and soft decision information of NBNDC-QB, we next design two JSCC schemes with low encoding delay and complexity: a scheme that uses a basic scalar quantizer (SQ) and sequence MAP decoding (the SQ-MAP scheme) and a scheme that consists of a channel-optimized vector quantizer (COVQ). Both schemes are source-centric JSCC systems in the sense that they do not use explicit algebraic channel coding and are thus less complex than the conventional separate source–channel coding systems. Furthermore, both schemes avoid the use of channel interleaving, which can introduce considerable additional delay when the underlying physical channel experiences slow fading. Instead, our schemes are tailored to exploit judiciously the statistical time-correlation structure of the channel to better combat channel impairments. Both systems are designed and evaluated for the NBNDC-QB and then tested (with a mismatched encoder/decoder) over the equivalent correlated Rayleigh DFC used with soft-decision demodulation to simulate the performance of the systems in a wireless setup.

This paper builds upon and significantly extends [11] and [26], where only binary (input, noise, and output) channels with Markovian additive noise were considered. We note that the channel considered in [11] and [26] is a special case of the NBNDC-QB model used here and can be obtained by setting the NBNDC-QB parameter  $q = 1$ , which translates into using hard-decision demodulation in the underlying fading channel. Note that, in contrast to this paper, in [11] and [26], correlated fading channels were not examined, and realistic channel modeling and validation (under mismatched coding) were not studied.

The organization of this paper is as follows. We introduce two channel models, i.e., the NBNDC-QB and the Rayleigh DFC, in Section II. In Section III, we study the MAP sequence detection of Markov sources over the NBNDC-QB and obtain the required relationships to implement the MAP decoder via the Viterbi algorithm. We further study the specific case of binary Markov sources sent over the NBNDC-QB with memory order  $M = 1$  and prove a necessary and sufficient condition under which the sequence MAP detector reduces to an instan-

taneous symbol-by-symbol mapping. A sufficient condition is also provided for the case of  $M \geq 1$ . In Section IV, we present and evaluate the two JSCC schemes for the NBNDC-QB (the SQ-MAP and COVQ schemes). In SQ-MAP, we apply the MAP decoding system to a SQ analog-valued Markov source and assess system performance in terms of signal-to-distortion ratio (SDR). This extends [11] where only binary output channels with Markov noise were considered. A COVQ system is then proposed for the NBNDC-QB. Similar to the SQ-MAP, we numerically show that the COVQ can successfully exploit the memory and soft-decision information of the channel and achieve considerably better SDR performance than systems with hard output quantization and systems that disregard the memory of the channel via the use of interleaving. In Section V, we numerically validate the NBNDC-QB model for both the SQ-MAP and COVQ systems as an effective approximation of the Rayleigh DFC. We conclude this paper in Section VI.

## II. NONBINARY NOISE DISCRETE CHANNEL WITH QUEUE-BASED NOISE AND DISCRETE FADING CHANNEL MODELS

Here, we review the two channel models studied in [23] and [25]: the NBNDC-QB and the Rayleigh DFC. We also observe that the DFC is a special instance of the NBNDC.

### A. NBNDC-QB

The NBNDC [23], [25] is a binary-input and  $2^q$ -ary-output channel model, where  $q \geq 1$ . The input data bits  $X_j \in \{0, 1\}$  are affected by noise  $Z_j$  via the following relation:

$$Y_j = (2^q - 1)X_j + (-1)^{X_j}Z_j, \quad j = 1, 2, \dots \quad (1)$$

where  $Y_j$  and  $Z_j$  take values in  $\mathcal{Y} = \{0, 1, \dots, 2^q - 1\}$ , with  $\{Y_j\}$  denoting the channel output process, and where noise process  $\{Z_j\}$  is assumed to be independent of  $\{X_j\}$ . According to (1),  $Z_j$  can also be written in terms of input and output symbols as follows:

$$Z_j = \frac{Y_j - (2^q - 1)X_j}{(-1)^{X_j}}, \quad j = 1, 2, \dots \quad (2)$$

The noise process  $\{Z_j\}$  can be in general any stochastic process. Following [23] and [25], we choose  $\{Z_j\}$  to be a nonbinary generalization of the QB noise [18]. We refer to the ensuing channel model by NBNDC-QB. The model, which is based on a ball sampling mechanism involving an urn and a queue of  $M$  cells, is intuitively described as follows. At time instance  $j > M$ , with probability  $1 - \epsilon$ , noise symbol  $Z_j$  is independent of past noise symbols and is picked according to  $\Pr\{Z_j = i\} = \rho_i$ ,  $i = 0, 1, \dots, 2^q - 1$ . With probability  $\epsilon$ ,  $Z_j$  is chosen to be one of  $M$  preceding noise symbols. (The probability that  $Z_j$  takes a particular value depends on a bias parameter<sup>2</sup>  $\alpha$  and increases in proportion to the number of times that value occurred in the  $M$  past noise symbols). The detailed description of the noise model is given in [18] and [25]. The resulting QB noise  $\{Z_j\}$  is a stationary and ergodic

<sup>2</sup>When  $M = 1$ , the queue has only one cell; in this case, the cell bias parameter is set to  $\alpha = 1$  [18], [25].

(irreducible)  $M$ th-order Markov process described by only  $2^q + 2$  independent parameters: memory order  $M$ , marginal probability distribution  $(\rho_0, \rho_1, \dots, \rho_{2^q-1})$ , and correlation parameters  $0 \leq \varepsilon < 1$  and  $\alpha \geq 0$ .

The state process  $\{\mathbf{S}_j\}$  of the QB noise, which is defined by  $\mathbf{S}_j \triangleq (Z_j, Z_{j-1}, \dots, Z_{j-M+1})$  for  $j \geq M$ , is a homogeneous first-order Markov process taking values in  $\{0, 1, \dots, 2^q - 1\}^M$ . It is shown in [23] and [25] that, for  $j \geq M + 1$ , the noise state transition probabilities  $Q(\mathbf{s}_j | \mathbf{s}_{j-1}) \triangleq \Pr\{\mathbf{S}_j = \mathbf{s}_j | \mathbf{S}_{j-1} = \mathbf{s}_{j-1}\}$ , with  $\mathbf{s}_j = (z_j, z_{j-1}, \dots, z_{j-M+1})$  and  $\mathbf{s}_{j-1} = (z'_j, z'_{j-1}, \dots, z'_{j-M+1})$ , are given by

$$Q(\mathbf{s}_j | \mathbf{s}_{j-1}) = \left( \sum_{\ell=1}^{M-1} \delta_{z_j, z_{j-\ell}} + \alpha \delta_{z_j, z_{j-M}} \right) \frac{\varepsilon}{M-1+\alpha} + (1-\varepsilon)\rho_{z_j} \quad (3)$$

if  $z_{k-1} = z'_k$  for  $k = j, \dots, j - M + 2$  and  $Q(\mathbf{s}_j | \mathbf{s}_{j-1}) = 0$  if otherwise, where  $\delta_{i,i'} = 1$  if  $i = i'$  and  $\delta_{i,i'} = 0$  if  $i \neq i'$ , and  $\sum_{l=1}^0 \triangleq 0$ .

Since the noise process is independent of the input, we have

$$\Pr\{Y^m = y^m | X^m = x^m\} = \Pr\{Z^m = z^m\} \quad (4)$$

where  $y^m = (y_1, y_2, \dots, y_m)$ ,  $x^m = (x_1, x_2, \dots, x_m)$ , and  $z^m = (z_1, z_2, \dots, z_m)$ , noting that  $x_i$  and  $y_i$  determine  $z_i$  through (2). The  $m$ -fold channel transition probability  $\Pr\{Z^m = z^m\} \triangleq P_{\text{NBNDQ-B}}^{(m)}(z^m)$  is given in [25, eqs. (20) and (21)] (with  $n$  corresponding to  $m$  herein); in particular, for  $m = 1$ ,  $P_{\text{NBNDQ-B}}^{(1)}(z_1) = \rho_{z_1}$  for all  $z_1 \in \mathcal{Y}$ . The channel noise correlation coefficient is given by [25]

$$\text{Cor} = \frac{\mathbf{E}[Z_k Z_{k+1}] \mathbf{E}[Z_k]^2}{\text{Var}(Z_k)} = \frac{\frac{\varepsilon}{M-1+\alpha}}{1 - (M-2+\alpha) \frac{\varepsilon}{M-1+\alpha}}.$$

### B. Rayleigh DFC

The Rayleigh DFC that we consider consists of a binary phase-shift keying (BPSK) modulator, a time-correlated flat Rayleigh fading channel with additive white Gaussian noise (AWGN), and a  $q$ -bit soft-quantized coherent demodulator. The input and output alphabets are  $\mathcal{X} = \{0, 1\}$  and  $\mathcal{Y} = \{0, 1, \dots, 2^q - 1\}$ , respectively. Denoting the DFC binary input process by  $\{X_k\}$ , the received channel symbols are given by  $R_k = \sqrt{E_s} A_k S_k + N_k$ ,  $k = 1, 2, \dots$ , where  $E_s$  is the energy of signal sent over the channel,  $S_k = 2X_k - 1$  is the  $\{-1, 1\}$ -valued BPSK modulated signal, and  $\{N_k\}$  is a sequence of independent identically distributed (i.i.d.) Gaussian random variables of variance  $N_0/2$ . Here,  $\{A_k\}$  is the Rayleigh fading process of the channel (which is independent from  $\{N_k\}$  and the input process) with  $A_k = |G_k|$ , where  $\{G_k\}$  is a time-correlated complex wide-sense stationary Gaussian process with Clarke's autocorrelation function given as a Bessel function of the normalized maximum Doppler frequency  $f_D T$  [25], [28]. The SNR of the DFC is given by  $\text{SNR} = E_s/N_0$ .

In the DFC model, a soft-decision demodulator consisting of a  $q$ -bit uniform quantizer takes the output  $R_k$  to produce the

discrete channel output:  $Y_k = j$  if  $R_k \in (T'_{j-1}, T'_j]$ , where the  $T'_j$  are uniformly spaced thresholds with step size  $\Delta$ , given by  $T'_{-1} = -\infty$ ,  $T'_j = (j+1 - 2^{q-1})\Delta$  for  $j = 0, 1, \dots, 2^q - 2$ , and  $T'_{2^q-1} = \infty$ . Letting  $\delta \triangleq \Delta/\sqrt{E_s}$  and  $T_j \triangleq T'_j/\sqrt{E_s}$ , the  $m$ -fold transition probability for the DFC can be calculated via [23], [25]

$$P_{\text{DFC}}^{(m)}(y^m | x^m) \triangleq \Pr\{Y^m = y^m | X^m = x^m\} = \mathbf{E}_{A_1 \dots A_m} \left[ \prod_{k=1}^m q_{x_k, y_k}(A_k) \right] \quad (5)$$

where

$$q_{i,j}(a_k) \triangleq \Pr\{Y_k = j | X_k = i, A_k = a_k\} = Q_G\left(\sqrt{2\text{SNR}}(T_{j-1} - (2i-1)a_k)\right) - Q_G\left(\sqrt{2\text{SNR}}(T_j - (2i-1)a_k)\right)$$

with  $Q_G(\cdot)$  denoting the Gaussian  $Q$ -function, and  $\mathbf{E}_X[\cdot]$  denoting expectation with respect to the random variable  $X$ . For  $m = 1$ , there is a closed-form expression for  $P_{\text{DFC}}^{(1)}(y|x) = P_{\text{DFC}}^{(1)}(j)$ , given by

$$P_{\text{DFC}}^{(1)}(j) = n(-T_{j-1}) - n(-T_j) \quad (6)$$

where  $j = y - (2^q - 1)x/(-1)^x \in \mathcal{Y}$ , and

$$n(T_j) = 1 - Q_G\left(T_j \sqrt{2\text{SNR}}\right) - \frac{\left[1 - Q_G\left(\frac{T_j \sqrt{2}}{\sqrt{\frac{1}{\text{SNR}} + 1}}\right)\right] e^{-\frac{T_j^2}{\left(\frac{1}{\text{SNR}} + 1\right)}}}{\sqrt{\frac{1}{\text{SNR}} + 1}}.$$

For  $m \leq 3$ ,  $P_{\text{DFC}}^{(m)}(y^m | x^m)$  can be calculated in closed form. For  $m > 3$ , (5) can only be determined numerically. Finally, we point out that the DFC is actually an NBNDQ, as given by (1), whose noise process has an  $n$ -fold distribution given by (5) [23], [25].

### C. Fitting the NBNDQ-QB Model to the Rayleigh DFC Model

We fit the NBNDQ-QB model to a given Rayleigh DFC (with fixed SNR and  $f_D T$  and  $q$ ) via the following steps [24], [25]:

- Match the noise 1-D probability distributions by setting  $\rho_j = P_{\text{DFC}}^{(1)}(j)$  for  $j \in \mathcal{Y}$ , where  $P_{\text{DFC}}^{(1)}(j)$  is given by (6), in terms of  $\delta$ ,  $q$ , and SNR. The values of  $\rho_j$  are given in Table I.
- Match the noise correlation coefficients (so that the QB parameter  $\alpha$  is given in terms of  $M$  and  $\varepsilon$ ). The DFC noise correlation coefficient is calculated using  $P_{\text{DFC}}^{(m)}(y^m | x^m)$  in (5) for  $m = 2$  [25].
- Estimate the remaining QB parameters  $(M, \varepsilon)$  by minimizing the Kullback–Leibler divergence rate between the two  $(2^q)$ -ary noise processes.

TABLE I  
 $\rho$  AND  $\delta$  VALUES FOR THE NBNDC-QB FOR SOFT-DECISION  
 RESOLUTIONS  $q = 1, 2, 3$

SNR (dB)	$q$	$\delta$	$\rho_0$	$\rho_1$	$\rho_2$	$\rho_3$	$\rho_4$	$\rho_5$	$\rho_6$	$\rho_7$
15	1	-	0.992	0.008	-	-	-	-	-	-
	2	0.12	0.972	0.020	0.006	0.001	-	-	-	-
	3	0.60	0.955	0.018	0.012	0.008	0.004	0.002	0.001	0.000
10	1	-	0.977	0.023	-	-	-	-	-	-
	2	0.20	0.924	0.053	0.019	0.005	-	-	-	-
	3	0.11	0.865	0.051	0.037	0.024	0.013	0.006	0.003	0.001
5	1	-	0.936	0.064	-	-	-	-	-	-
	2	0.40	0.782	0.154	0.054	0.010	-	-	-	-
	3	0.18	0.703	0.100	0.078	0.055	0.034	0.018	0.008	0.005
2	1	-	0.892	0.108	-	-	-	-	-	-
	2	0.50	0.695	0.196	0.085	0.024	-	-	-	-
	3	0.25	0.563	0.132	0.112	0.084	0.054	0.031	0.015	0.009

It is important to note that, in general, the two channel models cannot be matched to have an identical statistical behavior. However, for the memoryless case (with  $\text{Cor} = 0$ ), the NBNDC-QB is statistically identical to the ideally interleaved DFC. The values of the fitting NBNDC-QB parameters, which are obtained as outlined earlier, are given in [24, Table II] for different DFCs.

### III. MAP DETECTION OF MARKOV SOURCES OVER THE NONBINARY NOISE DISCRETE CHANNEL WITH QUEUE-BASED NOISE

#### A. MAP Decoder Design

Consider a stationary and ergodic discrete source  $\{\mathbf{X}_i\}$  with finite alphabet  $\{0, 1\}^n$ . Here,  $\{\mathbf{X}_i\}$  can represent the output of a source encoder, such as the output of a rate- $n$  SQ (as will be seen in Section IV). Its redundancy (which is due to the statistical memory of the source and the nonuniformity of its marginal distribution [11], [26]) encapsulates the dispensable amount of information of the source that can be eliminated via optimal variable-length lossless or fixed-rate nearly lossless data compression.

Suppose that a sequence of  $N$  source symbols  $\mathbf{x}^N = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \in \{0, 1\}^{nN}$  is sent over the NBNDC-QB in  $nN$  channel uses. The channel contaminates this source sequence via a sequence of  $2^q$ -ary noise symbols  $z^{nN} = (z_1, z_2, \dots, z_{nN}) \in \{0, 1, \dots, 2^q - 1\}^{nN}$  and outputs sequence  $\mathbf{y}^N = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N) \in \{0, 1, \dots, 2^q - 1\}^{nN}$ . Since the transmission over the channel is done bit by bit (and not  $n$ -tuple by  $n$ -tuple), we represent the noise sequence using a bit-by-bit notation, so that the noise symbols  $(z_{ni+1}, z_{ni+2}, \dots, z_{n(i+1)})$ ,  $i = 0, 1, \dots, N - 1$ , correspond to the input  $n$ -tuple  $\mathbf{x}_{i+1}$  and output  $n$ -tuple  $\mathbf{y}_{i+1}$ . The channel output  $\mathbf{y}^N$  is fed to a sequence MAP decoder, which estimates  $\mathbf{x}^N$  by  $\hat{\mathbf{x}}^N$  as

$$\hat{\mathbf{x}}^N = \arg \max_{\mathbf{x}^N} \Pr \{ \mathbf{X}^N = \mathbf{x}^N | \mathbf{Y}^N = \mathbf{y}^N \}. \quad (7)$$

To simplify the derivation of the MAP decoding metric, we first assume that the source  $\{\mathbf{X}_i\}$  is i.i.d. A minor modification of the decoding metric is needed to accommodate (Markov) sources with memory.

For  $i, j, k \in \{1, 2, \dots, nN - 1\}$ , where  $i + j \leq nN$  and  $i - k \geq 1$ , define

$$Q \left( z_{i+1}^{i+j} | z_{i-k}^i \right) \triangleq \Pr \left\{ Z_{i+1}^{i+j} = z_{i+1}^{i+j} | Z_{i-k}^i = z_{i-k}^i \right\}.$$

Since the NBNDC-QB channel noise is Markov of memory order  $M$ , for  $nN \geq M$  (which typically holds as  $N$  is assumed to be large in practice), it can be shown (see [2], [34]) that (7) is equivalent to

$$\hat{\mathbf{x}}^N = \arg \max_{\mathbf{x}^N} \left\{ \log \left[ P_{\text{NBNDC-QB}}^{(n)}(z_1^n) P(\mathbf{x}_1) \right] + \sum_{i=1}^{N-1} \log \left[ Q \left( z_{in+1}^{(i+1)n} | z_{in-(M-1)}^{in} \right) P(\mathbf{x}_{i+1}) \right] \right\} \quad (8)$$

where  $P(\mathbf{x}_i) \triangleq \Pr \{ \mathbf{X}_i = \mathbf{x}_i \}$  is the probability distribution of random vector  $\mathbf{X}_i$  of size  $n$ , and from (3), we have

$$Q \left( z_{j+1}^{j+n} | z_{j-(M-1)}^j \right) = \prod_{i=j+1}^{j+n} \left[ \left( \sum_{\ell=i-(M-1)}^{i-1} \delta_{z_i, z_\ell} + \alpha \delta_{z_i, z_{i-M}} \right) \times \frac{\varepsilon}{M-1+\alpha} + (1-\varepsilon) \rho_{z_i} \right] \quad (9)$$

with  $z_i \triangleq 0$  if  $i < 1$ ,  $z_i^j = (z_i, z_{i+1}, \dots, z_j)$ ,  $P_{\text{NBNDC-QB}}^{(n)}(z_1^n) = \Pr \{ Z_1^n = z_1^n \}$  given in [25, eqs. (20) and (21)], and  $z_i$  being related to its corresponding symbols  $x_i$  and  $y_i$  via (2). In light of (8) and (9), the MAP detection can be implemented using a modified version of the Viterbi algorithm [2], [34].

For the case of sources with memory, we assume that the source forms a discrete first-order Markov chain with state transition probability matrix  $P(\mathbf{x}_{i+1} | \mathbf{x}_i)$ ; in this case, the path metric can be directly obtained from (8) by replacing  $P(\mathbf{x}_{i+1})$  to  $P(\mathbf{x}_{i+1} | \mathbf{x}_i)$ .

#### B. Case Study: Map Detection of Binary Markov Sources

It is useful to know when the MAP detector can be replaced with an instantaneous (symbol-by-symbol) decoding rule without loss of optimality in terms of minimal sequence probability of error. The answer to this question is partly given in [26, Th. 1], which gives necessary and sufficient conditions for the MAP decoder to be *useless* for a *binary* Markov channel and a binary Markov source. In this case, a MAP decoder is called useless if it decodes what it sees (i.e.,  $\hat{X}^N = Y^N$ ) and thus does not improve the channel's sequence error rate (this is also known as *say-what-you-see* or *singlet decoding* [26], [31]). Note that skipping the decoder and accepting the output sequence without further processing can only be achieved for  $q = 1$  since, in that case, the output sequence is also binary.

We examine this problem for  $q \geq 2$ , where the received sequence is not binary, and we provide a nontrivial extension of [26, Th. 1]. Specifically, we apply mapping  $\theta$  to convert each  $2^q$ -ary received symbol  $Y_k$  into binary symbol  $\hat{Y}_k$ . We find optimal mapping  $\theta^*$  in the sense of minimizing the symbol error probability. For a symmetric binary Markov source and an

NBNDQ-B with  $M = 1$  and  $q \geq 2$ , our main result here establishes a necessary and sufficient condition for this mapping  $\theta^*$  to be an optimal sequence detection rule.

Consider binary source  $n = 1$  and the NBNDQ-B channel with  $M = 1$ . In this case, the channel noise  $\{Z_n\}$  is a first-order Markov process with (cf. Section II)

$$\begin{aligned} Q(z_i) &\triangleq \Pr\{Z_n = z_i\} = \rho_{z_i} \\ Q(z_i|z_j) &\triangleq \Pr\{Z_n = z_i|z_{n-1} = z_j\} \\ &= [\varepsilon\delta_{z_i,z_j} + (1 - \varepsilon)\rho_{z_i}], z_i, z_j \in \mathcal{Y}. \end{aligned} \quad (10)$$

Consider mapping  $\theta : \mathcal{Y} \rightarrow \{0, 1\}$ . To replace the MAP detector with mapping  $\theta$ , we simply set  $\hat{x}_n = \tilde{y}_n = \theta(y_n)$ ,  $n = 1, 2, \dots, N$ . The following lemma is proven in Appendix A.

*Lemma 1:* For the NBNDQ-B with parameters satisfying the condition

$$\rho_0 \geq \rho_1 \geq \rho_2 \geq \dots \geq \rho_{2^q-1} \quad (11)$$

among all mappings  $\theta : \mathcal{Y} \rightarrow \{0, 1\}$ , the following mapping  $\theta^*$  yields the lowest symbol probability of error:

$$\theta^*(y_n) = \tilde{y}_n = \begin{cases} 0, & \text{if } y_n < k^* \\ 1, & \text{otherwise} \end{cases} \quad (12)$$

where  $k^* \in \{0, 1, \dots, 2^q\}$  is the smallest value satisfying

$$\frac{\rho_{k^*}}{\rho_{2^q-k^*-1}} \leq \frac{P(1)}{P(0)} \quad (13)$$

where  $\rho_{-1} \triangleq \infty$ ,  $\rho_{2^q} \triangleq 0$ , and  $P(x) \triangleq \Pr\{X = x\}$ .

As in Section III-A, sequence  $\{\mathbf{X}_i\}$  (which is the same as  $\{X_i\}$  since  $n = 1$ ) is assumed to be a first-order stationary Markov chain. Define

$$\begin{aligned} p_{x_n x_{n-1}} &\triangleq P(x_n|x_{n-1}) \triangleq \Pr\{X_n = x_n|X_{n-1} = x_{n-1}\} \\ p_{x_n} &\triangleq P(x_n) \triangleq \Pr\{X_n = x_n\}. \end{aligned} \quad (14)$$

For the special case of a symmetric binary Markov source, i.e.,  $P(0) = P(1)$ , it can be seen from (13) that  $k^* = 2^{q-1}$  since, in this case, we have

$$\begin{aligned} 1) \quad &\frac{\rho_{k^*}}{\rho_{2^q-k^*-1}} = \frac{\rho_{2^{q-1}}}{\rho_{2^{q-1}-1}} \leq 1 \\ 2) \quad &\frac{\rho_{k^*-1}}{\rho_{2^q-k^*}} = \frac{\rho_{2^{q-1}-1}}{\rho_{2^{q-1}}} \geq 1. \end{aligned}$$

Define the *auxiliary* binary noise symbol  $\tilde{z}_n$ , which is related to its corresponding noise symbol  $z_n$  via

$$\tilde{z}_n = \begin{cases} 0, & \text{if } z_n < 2^{q-1} \\ 1, & \text{otherwise.} \end{cases} \quad (15)$$

The sequence of auxiliary binary noise symbols form auxiliary noise process  $\{\tilde{Z}_n\}$ . Since noise process  $\{Z_n\}$  is independent of input process  $\{X_n\}$  and the auxiliary binary noise variable  $\tilde{Z}_n$  is only a function of  $Z_n$ , the auxiliary noise process  $\{\tilde{Z}_n\}$  is also independent of input process  $\{X_n\}$ .

The following lemma is proven in Appendix B.

*Lemma 2:* If  $\{Z_n\}$  is a first-order Markov chain, then the auxiliary noise process  $\{\tilde{Z}_n\}$  forms a first-order Markov chain.

Note that the definition of the NBNDQ-B in (1), the definition of the mapping  $\theta^*$  given in (12), and the fact that, for symmetric binary Markov sources,  $k^* = 2^{q-1}$ , imply that  $\tilde{y}_n = x_n$  if and only if  $z_n < 2^{q-1}$  (in which case  $\tilde{z}_n = 0$ ). As a result, the auxiliary binary noise symbol can also be defined in terms of input  $x_n$  and  $\tilde{y}_n$  as follows:

$$\tilde{z}_n \triangleq \begin{cases} 0, & \text{if } \tilde{y}_n = x_n \\ 1, & \text{if } \tilde{y}_n = x_n^c \end{cases} \quad (16)$$

where  $x_n^c$  is the binary complement of  $x_n$  and  $\tilde{y}_n = \theta^*(y_n)$ .

The following theorem gives a necessary and sufficient condition for the mapping  $\theta^*$  to be an optimal sequence detection rule for  $q \geq 2$ . The proof is given in Appendix C.

*Theorem 1:* Consider a symmetric binary Markov source with  $p_{00} = p_{11} \in [1/2, 1]$  and the NBNDQ-B with correlation parameter  $\varepsilon \geq 0$ , memory order  $M = 1$ ,  $q \geq 2$ , and satisfying (11). Let  $x^N$  be a source sequence of length  $N \geq 3$  and  $y^N$  be a channel output sequence, and let  $\tilde{y}^N = \theta^*(y^N)$  be obtained by applying the mapping  $\theta^*$  component-wise to  $y^N$ . If  $x_1 = \tilde{y}_1$  and  $x_N = \tilde{y}_N$ , then  $\hat{x}^N = \tilde{y}^N$  is an optimal sequence MAP detection rule if and only if

$$\frac{\rho_{2^q-1-1}}{\rho_{2^q-1}} \times \left[ \frac{1-p_{00}}{p_{00}} \right]^2 \geq 1 \quad (17)$$

where  $p_{00}$  is defined via (14).

For binary symmetric first-order Markov sources with  $p_{00} = p_{11} \in [0, 1/2)$ , using the same approach, a similar theorem can be proven under the following condition:

$$\frac{\rho_{2^q-1-1}}{\rho_{2^q-1}} \times \left[ \frac{p_{00}}{1-p_{00}} \right]^2 \geq 1.$$

Note that the conditions  $x_1 = \tilde{y}_1$  and  $x_N = \tilde{y}_N$  in the theorem are not stringent since they can be simply satisfied by sending a preassigned value for  $x_1$  and  $x_N$ , where the receiver is also aware of the preassigned values. On the practical side, simulation results for long sequences confirm the result of the theorem without the need to check these conditions.

It is interesting to note that, for  $q \geq 2$ , the optimality condition (17) is independent of the channel noise correlation  $\varepsilon$ , whereas for  $q = 1$ , the condition depends on the noise correlation. In particular, [26, Corollary 3] shows that, for  $q = 1$ , the condition analogous to (17) is

$$\frac{(\varepsilon + (1 - \varepsilon)\rho_0)^2}{(1 - \varepsilon)^2\rho_0\rho_1} \times \left[ \frac{1-p_{00}}{p_{00}} \right]^2 \geq 1.$$

Theorem 1 is shown in Table II for a binary symmetric Markov source with  $p_{00} = 0.6$  and  $0.7$ , where  $C$  is the term on the left-hand side of (17). In the table, the NBNDQ-B's 1-D noise distribution is calculated by matching it to that of the underlying DFC; i.e., by setting  $\rho_j = P_{\text{DFC}}^{(1)}(j)$  as given in (6) in terms of SNR,  $q$ , and  $\delta$ , where the values of  $\delta$  are chosen so that the capacity of the DFC is maximized (see Table I). From Table II, we clearly observe that, when  $C < 1$ , the MAP decoder is performing better than the mapping  $\theta^*$ . For the cases

TABLE II

SYMBOL ERROR RATE (%) FOR MAP DECODING AND INSTANTANEOUS MAPPING  $\theta$  FOR SYMMETRIC BINARY MARKOV SOURCES WITH  $p_{00} = 0.6$  AND  $0.7$ . THE CHANNEL MODEL IS THE NBNDQ-QB, WITH  $M = 1$ ,  $\text{Cor} = 0.0$ , AND  $q = 2, 3$ . THE VALUES  $C$  ARE CALCULATED FROM (17). THE  $\delta$  VALUES FOR SNRS (15, 10, 5, 2) ARE (0.12, 0.20, 0.40, 0.50) FOR  $q = 2$  AND (0.06, 0.11, 0.18, 0.25) FOR  $q = 3$ , RESPECTIVELY

$p_{00}$	$q$	SNR (dB)								
		15		10		5		2		
		MAP	$\theta^*$	MAP	$\theta^*$	MAP	$\theta^*$	MAP	$\theta^*$	
0.6	2	$C = 1.39 > 1$	$C = 1.25 > 1$	$C = 1.27 > 1$	$C = 1.03 > 1$					
		0.76	0.76	2.30	2.30	6.43	6.43	10.85	10.85	
	3	$C = 0.80 < 1$	$C = 0.80 < 1$	$C = 0.73 < 1$	$C = 0.68 < 1$					
		0.73	0.76	2.22	2.30	6.21	6.43	10.51	10.85	
	0.7	2	$C = 0.57 < 1$	$C = 0.52 < 1$	$C = 0.52 < 1$	$C = 0.42 < 1$				
			0.64	0.76	1.91	2.30	5.55	6.43	9.50	10.85
3		$C = 0.33 < 1$	$C = 0.33 < 1$	$C = 0.30 < 1$	$C = 0.28 < 1$					
		0.62	0.76	1.90	2.30	5.37	6.43	9.27	10.85	

with  $C \geq 1$ , the MAP decoder and the instantaneous mapping  $\theta^*$  have the same performance; although in the simulations, the boundary conditions  $x_1 = \tilde{y}_1$  and  $x_N = \tilde{y}_N$  from the theorem have not been checked. Additional results illustrating the independence of the results of the noise correlation can be found in [34].

*Remark:* For  $M \geq 1$  and under the same setting as in Theorem 1, with the boundary assumptions that  $x_1 = \tilde{y}_1$ ,  $x_2 = \tilde{y}_2, \dots, x_M = \tilde{y}_M$  and  $x_N = \tilde{y}_N$ , where  $N > M + 1$ , one can prove that the *sufficiency* of condition (17) for the optimality of the symbol-by-symbol decoding map  $\theta^*$  can be generalized to

$$\frac{\frac{M-2+\alpha}{M-1+\alpha}\epsilon + (1-\epsilon)\rho_{2^{q-1}-1}}{\frac{M-2+\alpha}{M-1+\alpha}\epsilon + (1-\epsilon)\rho_{2^q-1}} \times \left[ \frac{1-p_{00}}{p_{00}} \right]^2 \geq 1. \quad (18)$$

#### IV. ROBUST LOSSY SOURCE CODING OVER THE NONBINARY NOISE DISCRETE CHANNEL WITH QUEUE-BASED NOISE

The NBNDQ captures a large class of channel models. For example, setting  $q = 1$  and letting the noise be i.i.d. reduces the NBNDQ to the familiar memoryless binary symmetric channel. As noted in Section II, the Rayleigh DFC with Clarke’s fading model is also an NBNDQ with a stationary and ergodic noise process [23], [25]. Furthermore, the model has the ability to delineate properly both the statistical memory structure and the soft-decision information of many real-valued output channels with memory used with antipodal signaling and soft-output quantization (including colored additive Gaussian channels). Here, we design two JSCC schemes, i.e., SQ-MAP and COVQ, for the NBNDQ-QB channel.

##### A. SQ-MAP System

1) *System Description:* Consider the system depicted in Fig. 1. Source  $\{V_i\}_{i=1}^\infty$  is assumed to be a real-valued stationary and ergodic process. The SQ encoder is mapping  $\gamma$  from the real domain  $\mathbb{R}$  of source symbols to the index set  $\{0, 1, \dots, 2^n - 1\}$ , such that  $\gamma(v) = i$  if  $v \in \mathbf{R}_i$ , where  $\{\mathbf{R}_i : i = 0, 1, \dots, 2^n - 1\}$  is a partition of  $\mathbb{R}$ . Hence, the SQ rate

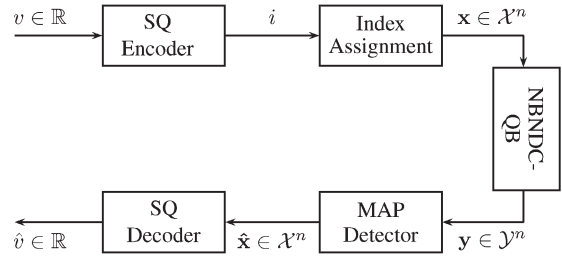


Fig. 1. Block diagram of a JSCC system using SQ-MAP.

is  $R = n$ . The partitions are chosen according to Lloyd–Max formulation [35], with the initial codebook selection obtained via the splitting algorithm [27]. The index assignment module is a one-to-one mapping  $b : \{0, 1, \dots, 2^n - 1\} \rightarrow \{0, 1\}^n$  with  $b(i) = \mathbf{x}$ , which maps each index  $i$  to binary vector  $\mathbf{x} \in \{0, 1\}^n$ . Since  $b$  is one to one, we can denote the quantization regions by  $\mathbf{R}_\mathbf{x}$  instead of  $\mathbf{R}_i$ , where  $b(i) = \mathbf{x}$ . Different index assignment methods such as the natural binary code, the folded binary code (FBC) [11], simulated annealing, and some heuristic assignment methods were tested. The FBC was selected because of its simplicity and good performance.

The  $n$ -tuple codeword  $\mathbf{x}$  is then sent bit by bit over the NBNDQ-QB channel. The channel output  $\mathbf{y} \in \{0, 1, \dots, 2^q - 1\}^n$  is fed to the MAP decoder described earlier, where the source redundancy and channel memory are harnessed for error correction. Finally, the SQ decoder  $\beta$  maps the decoder output  $\hat{\mathbf{x}}$  into the output levels of the quantizer codebook, i.e.,  $\beta(\hat{\mathbf{x}}) = c_{\hat{\mathbf{x}}}$ ,  $c_{\hat{\mathbf{x}}} \in \mathbb{R}$ , and  $\hat{\mathbf{x}} \in \{0, 1\}^n$ .

Using a MAP decoder enables the system to take advantage of interblock memory to minimize the sequence error probability (and, as a result, decrease the distortion).<sup>3</sup> It can be seen that, in this system, most of the complexity load is undertaken by the receiver.

2) *Numerical Results:* We next present numerical results on the performance of the described communication system for the NBNDQ-QB model. The NBNDQ-QB noise 1-D distribution  $\rho_j$  is expressed in terms of the SNR of the underlying Rayleigh DFC (discussed in Section II-B) by setting  $\rho_j = P_{\text{DFC}}^{(1)}(j)$ , where  $P_{\text{DFC}}^{(1)}(j)$  is given in (6) for  $j = 0, 1, \dots, 2^q - 1$ , and  $\delta$  is chosen as in [23], to maximize the DFC capacity as given in Table I.

Several source distributions are tested, including memoryless (i.i.d.) Gaussian and Laplacian sources and correlated Gauss–Markov sources. All sources have zero mean and unit variance. The correlated source is modeled via a first-order Markov process, defined by  $V_i = \phi V_{i-1} + U_i$ , where  $\phi \in (-1, 1)$  is the correlation parameter and  $\{U_i\}$  is a Gaussian i.i.d. process. Note that when  $\{V_i\}$  is i.i.d., the resulting  $2^n$ -ary process  $\{\mathbf{x}_i\}$  is also i.i.d. However, if  $\{V_i\}$  is Markovian,  $\{\mathbf{x}_i\}$  is not necessarily Markovian, but we model it as a Markov process (the distribution  $P(\mathbf{x}_i)$ , and state transition matrix  $[P(\mathbf{x}_i | \mathbf{x}_{i-1})]$  is calculated from a training set of source symbols used for designing the SQ.

<sup>3</sup>Note that the MAP decoder does not directly minimize the end-to-end mean square distortion.

TABLE III  
SQ-MAP TRAINING SDR RESULTS (IN DECIBELS) FOR MEMORYLESS NBND-CQB AND HIGHLY CORRELATED NBND-CQB WITH PARAMETERS  $M = \alpha = 1$ . G: MEMORYLESS GAUSSIAN SOURCE. L: MEMORYLESS LAPLACIAN SOURCE. GM: GAUSS-MARKOV SOURCE WITH  $\phi = 0.9$

Source	$q$	$n$	Fully interleaved (Cor=0)				Cor=0.9			
			SNR (dB)				SNR (dB)			
			15	10	5	2	15	10	5	2
G	1	1	4.17	3.75	2.78	1.94	4.19	3.77	2.85	1.97
		2	8.15	6.49	3.85	2.14	8.37	6.89	4.47	2.84
		3	11.05	7.80	4.02	1.93	11.58	8.43	4.76	2.76
	2	1	4.17	3.75	2.78	1.94	4.19	3.77	2.85	1.97
		2	8.15	6.49	3.85	2.14	8.69	7.68	5.61	4.03
		3	11.10	7.94	4.33	2.53	12.61	10.15	6.64	4.51
	3	1	4.17	3.75	2.78	1.94	4.19	3.77	2.85	1.97
		2	8.17	6.54	4.02	2.41	8.76	7.77	5.91	4.48
		3	11.15	7.98	4.38	2.57	12.86	10.52	7.16	5.12
L	1	1	2.87	2.62	2.00	1.44	2.88	2.63	2.05	1.45
		2	6.65	5.27	2.91	1.30	6.89	5.88	4.28	3.21
		3	9.59	6.49	2.72	0.58	10.14	7.64	4.88	3.34
	2	1	2.87	2.62	2.00	1.44	2.88	2.63	2.05	1.45
		2	6.69	5.42	3.32	2.01	7.26	6.72	5.53	4.47
		3	9.90	7.09	3.81	2.06	11.59	9.86	7.14	5.35
	3	1	2.87	2.62	2.00	1.44	2.88	2.63	2.05	1.45
		2	6.72	5.45	3.27	2.01	7.32	6.91	5.88	4.97
		3	9.96	7.12	3.75	1.99	11.85	10.47	7.94	6.23
GM	1	1	4.21	3.78	3.74	3.22	4.23	3.81	2.89	2.01
		2	8.95	8.29	6.97	6.34	8.94	8.24	6.88	5.71
		3	13.38	11.84	9.46	7.52	13.89	12.69	10.43	8.69
	2	1	4.38	4.23	3.89	3.61	4.35	4.14	3.44	2.65
		2	9.16	8.81	8.00	6.98	9.24	8.97	8.26	7.36
		3	13.98	12.87	10.72	8.91	14.47	14.02	12.88	11.51
	3	1	4.39	4.30	4.01	3.65	4.37	4.20	3.62	2.97
		2	9.18	8.87	8.09	7.24	9.29	9.18	8.64	8.00
		3	14.07	13.06	10.96	9.16	14.57	14.34	13.47	12.58

For each simulation, the SQ training and statistics collection is done over a set of  $10^6$  source symbols. For testing,  $N = 10^5$  independently generated source symbols are transmitted, and the SDR per source symbol is calculated under the mean-square-error distortion measure. We run each simulation ten times and take average for ensuring consistent results. Table III depicts simulation results (in decibels) for different sources over the NBND-CQB model with several values of the parameters, i.e., SNR, SQ codeword length  $n$ , noise correlation Cor, and soft-decision resolution  $q$ .

*Memoryless sources:* Table III indicates that the system exploiting high noise correlation performs significantly better than the system that fully interleaves (Cor = 0) the channel. For example, more than 4.2 dB of SDR gain is obtained for memoryless Laplacian sources at  $q = 3$ ,  $n = 3$ , SNR = 2. In addition, for  $n = 1$ , since the quantized codewords form a symmetric i.i.d. source ( $p_{00} = p_{11} = 0.5$ ), the results illustrate [26, Corollary 3] (when  $q = 1$ ) and Theorem 1 of Section III (when  $q \geq 2$ ). Table IV shows the results of an SQ system, using the instantaneous mapping instead of MAP detection, for various source distributions. A comparison of Tables III and IV for  $n = 1$  and for Gaussian and Laplacian sources reveals that the instantaneous symbol-by-symbol decoder and the MAP decoder are performing for these cases. Considerable gains (up to 2.25 and 3 dB) are also obtained by increasing the quantizer resolution to  $q = 2$  and  $q = 3$ , respectively (for  $n = 3$ , SNR = 5, and Cor = 0.9 for Laplacian sources).

*Gauss-Markov sources:* From Table III, we remark that up to 3.4-dB SDR gains (at  $q = 3$ ,  $n = 3$ , and SNR = 2) can

TABLE IV  
SQ WITH INSTANTANEOUS MAPPING: TRAINING SDR RESULTS (IN DECIBELS) FOR  $n = 1$  AND THE MEMORYLESS NBND-CQB AND THE HIGHLY CORRELATED NBND-CQB WITH PARAMETERS  $M = 1$ ,  $\alpha = 1$ . G: MEMORYLESS GAUSSIAN SOURCE. L: MEMORYLESS LAPLACIAN SOURCE. GM: GAUSS-MARKOV SOURCE WITH  $\phi = 0.9$

Source	$q$	Fully interleaved (Cor=0)				Cor=0.9			
		SNR (dB)				SNR (dB)			
		15	10	5	2	15	10	5	2
G	1, 2, 3	4.17	3.75	2.78	1.94	4.19	3.77	2.85	1.97
L	1, 2, 3	2.87	2.62	2.00	1.44	2.88	2.63	2.05	1.45
GM	1, 2, 3	4.21	3.78	2.82	1.97	4.23	3.81	2.89	2.01

be realized for Gauss-Markov sources by exploiting the noise correlation instead of interleaving the channel. In general, a better performance is observed when the channel is highly correlated.

At low rates, particularly at  $n = 1$ , the SDR performance for the correlated channel is worse than that for the uncorrelated channel. This behavior is expected for  $n = 1$  and  $q = 1$  from [26, Corollary 3]. According to this corollary and the numerical results, for the correlated channel, the source memory has a mismatch with the channel memory. As a result, increasing the channel noise correlation will also increase the mismatch between the source and the channel, making the SQ-MAP perform worse on correlated channels than on uncorrelated channels. However, this mismatch does not occur for higher rates ( $n > 2$ ), and the SDR performance of the system significantly improves with increasing channel noise correlation. For  $n = 1$ , the results of Theorem 1 can also be shown by comparing the Gauss-Markov source results for  $n = 1$  (with  $p_{00} = p_{11} = 0.86$ ) in Tables III and IV.

In addition, using a three-bit soft-decision quantizer in the receiver, gains up to 3.9 dB are observed (at  $n = 3$ , SNR = 2, and Cor = 0.9 for Gauss-Markov sources) over a hard-decision quantizer ( $q = 1$ ). Additional results for other values of Cor and  $M$  are given in [34]. It is seen that, for lower noise correlations, (e.g., Cor = 0.5 and 0.7), the system still has a better SDR performance, although the gain is less than for Cor = 0.9.

## B. COVQ

In contrast to the SQ-MAP, the COVQ system incorporates the intrablock memory of the noise into the quantizer design algorithm. This makes it more robust to channel noise than standard scalar or vector quantizers, such as the Linde-Buzo-Gray vector quantizer [27], without adding extra algebraic (channel coding) redundancy.

The COVQ system of rate  $R = n/k$  is depicted in Fig. 2. We refer the reader to [1] and [34] for a detailed description of the system and its design algorithm based on iteratively applying optimality encoder and decoder conditions [7]–[10]. Instead, we focus on illustrating the performance of the system over the NBND-CQB channel. The same source distributions were used as for the SQ-MAP system, and the algorithm was trained using 500 000 source vectors.

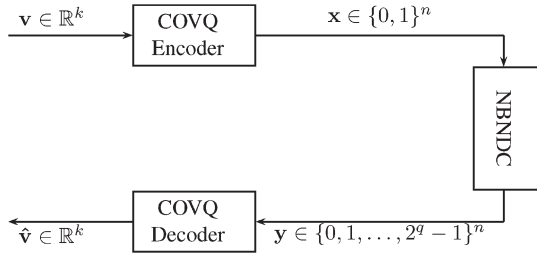


Fig. 2. Block diagram of a COVQ system.

TABLE V  
COVQ TRAINING SDR RESULTS (IN DECIBELS) FOR THE MEMORYLESS NBNDNC-QB AND THE HIGHLY CORRELATED NBNDNC-QB WITH PARAMETERS  $\alpha = 1.0, M = 1, \varepsilon = 0.9$ . MEMORYLESS LAPLACIAN SOURCE

q	R	k	Fully interleaved (Cor=0)				Cor=0.9			
			SNR (dB)				SNR (dB)			
			15	10	5	2	15	10	5	2
1	1	1	2.87	2.63	2.07	1.58	2.87	2.63	2.07	1.58
		2	3.46	3.12	2.42	1.84	3.48	3.16	2.45	1.85
		3	4.06	3.46	2.44	1.90	4.26	3.93	3.38	2.91
	2	1	6.64	5.39	3.48	2.32	6.71	5.64	4.12	4.34
		2	7.67	6.18	4.18	3.12	7.76	7.08	6.44	5.62
		3	8.11	6.61	4.55	3.38	8.40	8.16	7.07	6.38
	3	1	9.70	7.08	4.21	3.54	9.84	9.15	8.01	6.78
		2	10.92	8.52	5.96	4.36	11.27	10.91	9.33	8.17
		3	11.40	9.25	6.41	4.65	12.26	11.97	10.23	9.10
2	1	1	2.90	2.70	2.21	1.77	2.90	2.70	2.21	1.77
		2	3.50	3.22	2.59	2.06	3.52	3.25	2.63	2.08
		3	4.14	3.63	2.66	2.03	4.31	4.04	3.46	3.10
	2	1	6.85	5.79	3.96	2.77	6.95	6.05	4.48	4.74
		2	7.91	6.59	4.88	3.73	8.06	7.01	6.69	5.98
		3	8.37	7.13	5.24	3.94	8.66	7.62	7.25	6.51
	3	1	10.32	7.84	5.97	4.41	10.63	8.46	8.45	7.31
		2	11.45	9.51	7.04	5.30	11.71	10.52	9.02	8.51
		3	12.08	10.13	7.52	5.62	12.53	11.58	10.36	9.65

Table V depicts COVQ training results for the memoryless Laplacian source over NBNDNC-QB with different channel noise correlation coefficients and coding rates. The channel parameters  $\delta$  and  $\rho_j$  are given in Table I. As for the SQ-MAP system, the results show that the COVQ system performs consistently better over the highly correlated channel than over the fully interleaved one: Gains of more than 4.4 dB are achievable (e.g., for  $q = 1, R = 3, k = 3$ , and  $SNR = 2$  in Table V) by exploiting the channel memory. Note that, since the COVQ only makes use of intrablock memory, for rate  $R = 1$  and low dimensions  $k$ , the block length is so small that there is not much channel memory to exploit. As a result, the performance is constant for different channel correlations. However, it is observed that, in some cases, interleaving may give better COVQ performance over channels with lower noise correlations. Since the capacity of the correlated channel is strictly higher than that of the memoryless channel, this degradation may be due to poor selection of the initial codebook for the vector quantizer. In general, the results indicate improvement of COVQ performance when the channel noise correlation is increased. Similar observations can be made for memoryless Gaussian and Gauss–Markov sources and different channel parameters (such as lower correlation values Cor and higher memory orders  $M$ ) [34].

Finally, as in the case of the SQ-MAP system, we remark that the COVQ performs considerably better with soft-decision quantization ( $q \geq 2$ ) as opposed to hard-decision quantization

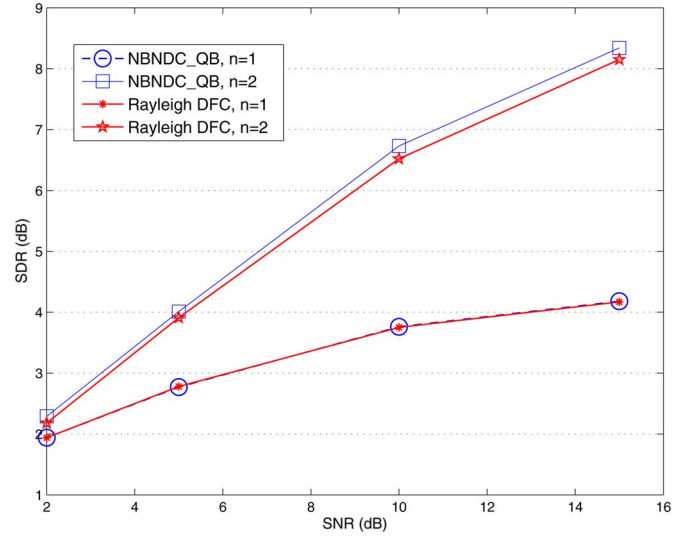


Fig. 3. SQ-MAP simulation SDR results (in decibels) for the DFC-fitted NBNDNC-QB and the DFC. Memoryless Gaussian source:  $q = 2$ ; and  $f_D T = 0.005$ .

( $q = 1$ ). Table V reveals SDR gains as high as 1.7 dB (for  $R = 3, k = 1, Cor = 0$ , and  $SNR = 5$ ) by just using  $q = 2$ . For three-bit quantization, additional (but less pronounced) gains can be realized.

### V. VALIDATING THE NONBINARY NOISE DISCRETE CHANNEL WITH QUEUE-BASED NOISE MODEL IN TERMS OF SIGNAL-TO-DISTORTION RATIO

To illustrate the use of the NBNDNC-QB model in a practical setup, we next assess how well it can represent the correlated Rayleigh DFC in terms of SDR performance when used with the SQ-MAP and COVQ systems. As noted in Section II-B, the  $m$ -fold probability distribution of correlated Rayleigh random variables is not known in closed form for  $m > 3$ , and as a result, the channel transition probabilities  $P_{DFC}^{(m)}(y^m|x^m)$  can only be calculated numerically. It is shown in [24] that the NBNDNC-QB model (for which the channel transition probabilities are known in closed form) can approximate the Rayleigh DFC in terms of channel capacity and noise autocorrelation function.

#### A. SQ-MAP System

To validate the NBNDNC-QB model as a good representation of the Rayleigh DFC for SQ-MAP systems, we design the MAP detector using the path metric obtained for an NBNDNC-QB (whose parameters are obtained using the matching procedure described in Section II-C) and run simulation using both the NBNDNC-QB and the Rayleigh DFC to compare their performance. Hence, the system simulated over the DFC employs a mismatched decoder.

To simulate the Rayleigh DFC, we generate the fading coefficients using the modified Clarke’s method [36]. Simulation results in terms of SDR are shown in Fig. 3 for memoryless Gaussian sources. More results can be found in [2] and [34]. Comparing the performance of the system for the two channels, we observe that for lower rates (codeword lengths  $n = 1$  and 2 for the memoryless Gaussian source), there is good conformity between the results for the two channel models.



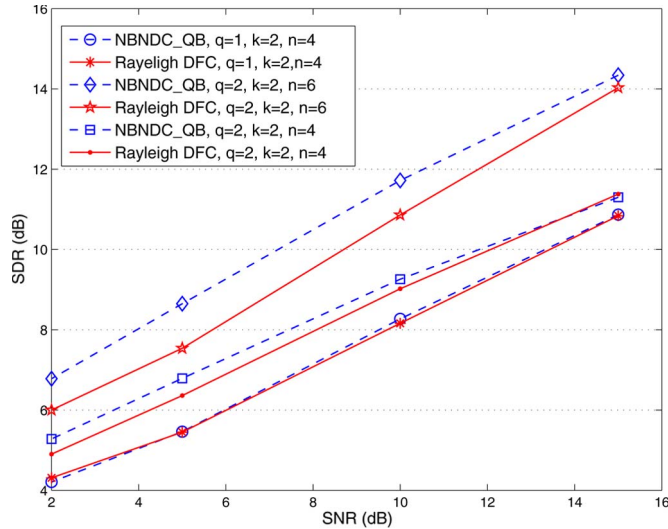


Fig. 4. COVQ simulation SDR results (in decibels) for the DFC-fitted NBNDC-QB and the DFC. Memoryless Gaussian source:  $q = 2$ ; and  $f_{DT} = 0.005$ .

This agreement in SDR performance can be heuristically explained by noting that, for low rates ( $n = 1$  and  $2$ ), the SQ output sent to the channel input is nearly i.i.d. uniform. However, the NBNDC-QB and DFC channels were matched by minimizing the divergence rate between their noise processes. Hence, when both channels are driven by the same capacity-achieving input (which is i.i.d. uniform as both channels are symmetric), they will then have a similar probability of error performance in addition to nearly identical capacities. The same agreement in SDR performance is also observed for memoryless Laplacian and Gauss–Markov sources for  $n = 1$ . We finally note that, for  $n \geq 3$ , some disagreement in SDR performance is observed [34] between the two systems. (In this case, the SQ output is not i.i.d. uniform). Note that the degradation is expected since, for higher rates, the source input distribution becomes less uniform; therefore, the matched NBNDC-QB model becomes less successful in imitating the error performance of the DFC.

### B. COVQ System

We next train a COVQ for an NBNDC-QB whose parameters are obtained by the procedure given in Section II-C to match a given Rayleigh DFC. The resulting performance of the channel-optimized quantizer is then tested over the DFC. (Here again, there is a mismatch in the COVQ system used over the DFC as it is designed for the modeling NBNDC-QB.) Training and simulation results in terms of the SDR are shown in Fig. 4 for memoryless Gaussian sources: We observe that there is a good conformity between the results for the two channel models, i.e., in the case that the NBNDC-QB is used for training and the DFC is used for testing. However, for higher rates, some degradation between the simulation and training results is observed. The same observation applies for memoryless Laplacian and correlated Gauss–Markov sources [1], [34].

## VI. CONCLUSION

Robust source coding for a new channel model called NBNDC-QB was studied. This channel model is analytically

tractable (its transition probabilities, noise entropy rate, and autocorrelation function are known in closed form) and can serve as a good approximation to a discrete Rayleigh fading channel. First, the MAP decoding of a discrete source was considered and implemented using the Viterbi algorithm. For binary symmetric sources, a necessary and sufficient condition was derived for the MAP decoder to be reducible to an instantaneous symbol-by-symbol decoder without loss of optimality. This condition was also numerically illustrated.

Two lossy JSCC schemes with low delay and complexity were implemented and tested for the NBNDC-QB model. In the first system, the MAP decoder was matched to a SQ Markov source (SQ-MAP system). Numerical results demonstrated that the proposed system can successfully utilize memory and soft-decision information over the NBNDC-QB channel model. The second JSCC scheme, which consists of a COVQ implemented over the NBNDC-QB, was also shown to exploit successfully the intrablock memory and soft information of the channel in combating channel errors. Both systems outperform their fully interleaved counterpart systems where the channel memory is discarded using ideal (infinite) interleaving.

Finally, the channel model was compared with the correlated Rayleigh DFC in terms of SDR performance. It was shown numerically that, for both the SQ-MAP and COVQ systems, the NBNDC-QB model can effectively approximate the Rayleigh DFCs for low coding rates. Future research directions include the integration of iterative source–channel coding/decoding methods (e.g., see [37]–[42]) into our JSCC systems for correlated fading channels (used without interleaving) to further improve performance under very noisy conditions while only moderately increasing system delay and complexity.

## APPENDIX A PROOF OF LEMMA 1

We will show that any mapping  $\theta$  can be modified via a sequence of simple changes, each of which can only decrease the error probability, such that, after a finite number of these changes, the modified mapping will be equal to  $\theta^*$ .

To this end, we consider mapping  $\theta : \mathcal{Y} \rightarrow \{0, 1\}$  as a classification rule that classifies  $2^q$  different output symbols from  $\{0, 1, \dots, 2^q - 1\}$  into two classes  $\mathcal{Y}_0$  and  $\mathcal{Y}_1$ . Thus,  $\theta$  is defined by

$$\theta(y) \triangleq \tilde{y} = \begin{cases} 0, & \text{if } y \in \mathcal{Y}_0 \\ 1, & \text{if } y \in \mathcal{Y}_1 \end{cases} \quad (19)$$

where  $\mathcal{Y}_1 \subset \mathcal{Y}$  and  $\mathcal{Y}_0 = \mathcal{Y} \setminus \mathcal{Y}_1$ .

According to (12), for  $\theta^*$ , we have  $\mathcal{Y}_0^* = \{0, 1, \dots, k^* - 1\}$  and  $\mathcal{Y}_1^* = \{k^*, k^* + 1, \dots, 2^q - 1\}$ . Let  $P_e$  denote the symbol error probability under mapping  $\theta$ , i.e.,  $P_e \triangleq \Pr\{\theta(Y) \neq X\}$ , where  $(X, Y)$  has the common joint distribution of the pairs  $(X_i, Y_i)$ . If  $\theta \neq \theta^*$ , at least one of the two following cases hold.

- 1) There exists an element  $a \in \mathcal{Y}_1$ , such that  $a < k^*$ . Removing  $a$  from  $\mathcal{Y}_1$  and then adding it to  $\mathcal{Y}_0$  yield a

mapping with error probability  $\tilde{P}_e$ , such that

$$\begin{aligned}\tilde{P}_e - P_e &= \Pr\{Y = a|X = 1\}Pr\{X = 1\} \\ &\quad - \Pr\{Y = a|X = 0\}Pr\{X = 0\} \\ &= Q(2^q - 1 - a)P(1) - Q(a)P(0) \\ &= P(1)\rho_{2^q-1-a} - P(0)\rho_a.\end{aligned}\quad (20)$$

According to (11),  $\rho_a \geq \rho_{k^*-1}$  and  $\rho_{2^q-k^*} \geq \rho_{2^q-1-a}$ . Hence, by (13), we have

$$\frac{P(1)}{P(0)} \leq \frac{\rho_{k^*-1}}{\rho_{2^q-k^*}} \leq \frac{\rho_a}{\rho_{2^q-1-a}}$$

and therefore,  $\tilde{P}_e - P_e \leq 0$ . Thus, removing  $a$  from  $\mathcal{Y}_1$  and adding it to  $\mathcal{Y}_0$  do not increase the error probability.

- 2) There exists element  $b \in \mathcal{Y}_0$ , such that  $b \geq k^*$ . Similar to (21), it can be shown that

$$\tilde{P}_e - P_e = -P(1)\rho_{2^q-1-b} + P(0)\rho_b.$$

According to (11),  $\rho_b \leq \rho_{k^*}$ , and  $\rho_{2^q-k^*-1} \leq \rho_{2^q-1-b}$ . Hence, by (13), we have

$$\frac{P(1)}{P(0)} \geq \frac{\rho_{k^*}}{\rho_{2^q-k^*-1}} \geq \frac{\rho_b}{\rho_{2^q-1-b}}$$

and thus,  $\tilde{P}_e - P_e \leq 0$ . Hence, removing  $b$  from  $\mathcal{Y}_0$  and adding it to  $\mathcal{Y}_1$  do not increase error probability.

It follows that, if we start from an arbitrary mapping  $\theta$ , after at most  $2^q - 1$  applications of the given replacement steps, we obtain  $\theta^*$ . Since each step can only reduce the error probability, among all mappings  $\theta^*$  must minimize the error probability. ■

## APPENDIX B PROOF OF LEMMA 2

Let  $\{Z_n\}$  be a Markov process with a finite state space  $\mathcal{Y}$ , and let the binary process  $\{\tilde{Z}_n\}$  be defined by  $\tilde{Z}_n = f(Z_n)$  for function  $f: \mathcal{Y} \rightarrow \{0, 1\}$ . We use a result given in [32, pp. 325], [33] which states that a sufficient condition for  $\{\tilde{Z}_n\}$  to be a Markov process is that

$$\begin{aligned}\Pr\{\tilde{Z}_{n+1} = \tilde{z}_{n+1}|Z_n = z_n\} \\ = \Pr\{\tilde{Z}_{n+1} = \tilde{z}_{n+1}|\tilde{Z}_n = f(z_n)\}\end{aligned}\quad (21)$$

for all  $\tilde{z}_{n+1}$  and  $z_n$ . Let  $f$  be given by (15). Then, using (10), we have

$$\begin{aligned}\Pr\{\tilde{Z}_{n+1} = 0|Z_n = z_n\} \\ = \Pr\{Z_{n+1} \in \{0, 1, \dots, 2^{q-1} - 1\}|Z_n = z_n\} \\ = \sum_{i=0}^{2^q-1-1} \Pr\{Z_{n+1} = i|Z_n = z_n\} \\ = \sum_{i=0}^{2^q-1-1} Q(i|z_n) \\ = \begin{cases} \varepsilon + \sum_{i=0}^{2^q-1-1} (1-\varepsilon)\rho_i, & \text{if } z_n < 2^{q-1} \\ \sum_{i=0}^{2^q-1-1} (1-\varepsilon)\rho_i, & \text{if } z_n \geq 2^{q-1}. \end{cases}\end{aligned}\quad (22)$$

In addition

$$\begin{aligned}\Pr\{\tilde{Z}_{n+1} = 0|\tilde{Z}_n = 0\} \\ = \sum_{i=0}^{2^q-1-1} \Pr\{\tilde{Z}_{n+1} = 0|Z_n = i\} \frac{\Pr\{Z_n = i\}}{\Pr\{\tilde{Z}_n = 0\}}.\end{aligned}\quad (23)$$

Note that, according to (22), for all  $j = 0, \dots, 2^q-1$ , we have

$$\Pr\{\tilde{Z}_{n+1} = 0|Z_n = j\} = \varepsilon + \sum_{i=0}^{2^q-1-1} (1-\varepsilon)\rho_i.$$

Hence, (23) is equal to

$$\begin{aligned}\left(\varepsilon + \sum_{i=0}^{2^q-1-1} (1-\varepsilon)\rho_i\right) \frac{\sum_{i=0}^{2^q-1-1} \Pr\{Z_n = i\}}{\Pr\{\tilde{Z}_n = 0\}} \\ = \left(\varepsilon + \sum_{i=0}^{2^q-1-1} (1-\varepsilon)\rho_i\right) \frac{\Pr\{\tilde{Z}_n = 0\}}{\Pr\{\tilde{Z}_n = 0\}} \\ = \varepsilon + \sum_{i=0}^{2^q-1-1} (1-\varepsilon)\rho_i.\end{aligned}$$

Using the same reasoning, it can be shown that

$$\Pr\{\tilde{Z}_{n+1} = 0|\tilde{Z}_n = 1\} = \sum_{i=0}^{2^q-1-1} (1-\varepsilon)\rho_i.\quad (24)$$

Thus, according to (23) and (24)

$$\begin{aligned}\Pr\{\tilde{Z}_{n+1} = 0|\tilde{Z}_n = \tilde{z}_n\} \\ = \begin{cases} \varepsilon + \sum_{i=0}^{2^q-1-1} (1-\varepsilon)\rho_i, & \text{if } \tilde{z}_n = 0 \\ \sum_{i=0}^{2^q-1-1} (1-\varepsilon)\rho_i, & \text{if } \tilde{z}_n = 1. \end{cases}\end{aligned}\quad (25)$$

It can be seen from (22) and (25) that condition (21) is satisfied for the NBNDC-QB with memory order  $M = 1$  and the function  $f$  defined via (15). Consequently,  $\{\tilde{Z}_n\}$  is a first-order Markov chain. ■

## APPENDIX C PROOF OF THEOREM 1

For  $\theta^*$  to be the optimal detection rule, it is necessary and sufficient that, for all  $x^N \in \{0, 1\}^N$  and  $y^N \in \{0, 1, \dots, 2^q - 1\}^N$ , the following holds:

$$\gamma \triangleq \frac{\Pr\{X^N = \tilde{y}^N|Y^N = y^N\}}{\Pr\{X^N = x^N|Y^N = y^N\}} \geq 1.$$

$\gamma$  can be written as

$$\gamma = \frac{\Pr\{Y^N = y^N|X^N = \tilde{y}^N\} \Pr\{X^N = \tilde{y}^N\}}{\Pr\{Y^N = y^N|X^N = x^N\} \Pr\{X^N = x^N\}}.$$

Note that, by (2) and (4), we have  $\Pr\{Y^N = y^N|X^N = x^N\} = \Pr\{Z^N = z^N\}$ , where  $z_i = y_i - (2^q - 1)x_i/(-1)^{x_i} \in \{0, 1, \dots, 2^q - 1\}$ ,  $i = 1, 2, \dots, N$ . In addition, note that, by the

definition of  $\theta^*$  (with  $k^* = 2^{q-1}$ ), if we let  $a_i \triangleq y_i - (2^q - 1)\tilde{y}_i/(-1)^{\tilde{y}_i}$ , with  $\tilde{y}_i = \theta^*(y_i)$ , then  $a_i \in \{0, 1, \dots, 2^{q-1} - 1\}$  for all  $i = 1, \dots, N$ .

Thus we have, with  $a^N = (a_1, \dots, a_N)$ ,  $z^n = (z_1, \dots, z_n)$ , and  $\tilde{y}^n = (\tilde{y}_1, \dots, \tilde{y}_n)$  as above

$$\begin{aligned} \gamma &= \frac{\Pr\{Z^N = a^N\} \Pr\{X^N = \tilde{y}^N\}}{\Pr\{Z^N = z^N\} \Pr\{X^N = x^N\}} \\ &= \frac{\Pr\{Z_1 = a_1\} P(\tilde{y}_1)}{\Pr\{Z_1 = z_1\} P(x_1)} \prod_{k=2}^N \frac{Q(a_k|a_{k-1}) P(\tilde{y}_k|\tilde{y}_{k-1})}{Q(z_k|z_{k-1}) P(x_k|x_{k-1})} \\ &= \prod_{k=2}^N \frac{Q(a_k|a_{k-1}) P(\tilde{y}_k|\tilde{y}_{k-1})}{Q(z_k|z_{k-1}) P(x_k|x_{k-1})} \end{aligned}$$

where the last equality follows from  $\Pr\{Z_1 = a_1\} P(\tilde{y}_1) = \Pr\{Z_1 = z_1\} P(x_1)$  since  $X_1 = \tilde{Y}_1$  according to the hypothesis.

We partition the index set  $\mathcal{K} = \{2, 3, \dots, N\}$  as follows:  $\mathcal{K} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$ , where

$$\mathcal{A}_1 \triangleq \{k \in \mathcal{K} : x_k = \tilde{y}_k, x_{k-1} = \tilde{y}_{k-1}\}$$

$$\mathcal{A}_2 \triangleq \{k \in \mathcal{K} : x_k \neq \tilde{y}_k, x_{k-1} \neq \tilde{y}_{k-1}\}$$

$$\mathcal{A}_3 \triangleq \{k \in \mathcal{K} : x_k \neq \tilde{y}_k, x_{k-1} = \tilde{y}_{k-1}\}$$

$$\mathcal{A}_4 \triangleq \{k \in \mathcal{K} : x_k = \tilde{y}_k, x_{k-1} \neq \tilde{y}_{k-1}\}.$$

Hence

$$\gamma = \prod_{i=1}^4 \prod_{k \in \mathcal{A}_i} \frac{Q(a_k|a_{k-1}) P(\tilde{y}_k|\tilde{y}_{k-1})}{Q(z_k|z_{k-1}) P(x_k|x_{k-1})}.$$

In set  $\mathcal{A}_1$ , since  $x_k = \tilde{y}_k$  and  $x_{k-1} = \tilde{y}_{k-1}$ , we see that  $z_k = a_k$  and  $z_{k-1} = a_{k-1}$ . Thus

$$\begin{aligned} \prod_{k \in \mathcal{A}_1} \frac{Q(a_k|a_{k-1}) P(\tilde{y}_k|\tilde{y}_{k-1})}{Q(z_k|z_{k-1}) P(x_k|x_{k-1})} \\ = \prod_{k \in \mathcal{A}_1} \frac{Q(a_k|a_{k-1}) P(x_k|x_{k-1})}{Q(a_k|a_{k-1}) P(x_k|x_{k-1})} = 1. \end{aligned}$$

In  $\mathcal{A}_2$ ,  $x_k \neq \tilde{y}_k$  and  $x_{k-1} \neq \tilde{y}_{k-1}$  imply that  $x_k = 1 - \tilde{y}_k$  and  $x_{k-1} = 1 - \tilde{y}_{k-1}$ . In addition, if  $x_k \neq \tilde{y}_k$  and  $x_{k-1} \neq \tilde{y}_{k-1}$ , then by the definition of  $\theta^*$  (with  $k^* = 2^q - 1$ ),  $z_k, z_{k-1} \geq 2^{q-1}$ . Now since the Markov source is symmetric ( $p_{00} = p_{11}$ ), we obtain that  $P(\tilde{y}_k|\tilde{y}_{k-1}) = P(x_k|x_{k-1})$ . Noting that  $a_k, a_{k-1} < 2^{q-1}$  and according to (10) and (11), we have

$$\prod_{k \in \mathcal{A}_2} \frac{Q(a_k|a_{k-1}) P(\tilde{y}_k|\tilde{y}_{k-1})}{Q(z_k|z_{k-1}) P(x_k|x_{k-1})} = \prod_{k \in \mathcal{A}_2} \frac{Q(a_k|a_{k-1})}{Q(z_k|z_{k-1})} \geq 1.$$

We next note that, since  $X_1 = \tilde{Y}_1$  and  $X_N = \tilde{Y}_N$ , we clearly must have  $|\mathcal{A}_3| = |\mathcal{A}_4|$ , where  $|\mathcal{B}|$  denotes the number of elements in set  $\mathcal{B}$ . Furthermore, in  $\mathcal{A}_3$ , we have  $x_k \neq \tilde{y}_k$  implying  $x_k = 1 - \tilde{y}_k$ . Therefore, according to the definitions of  $z_k$  and  $a_k$  and using (2), it can be seen that  $z_k = (2^q - 1) - a_k$  and  $z_{k-1} = a_{k-1}$ . Similarly, in set  $\mathcal{A}_4$ , we have  $z_k = a_k$  and  $z_{k-1} =$

$(2^q - 1) - a_{k-1}$ . In addition, due to the source symmetry and noting that  $p_{00} \in [1/2, 1]$ , we have

$$\min_{k \in \mathcal{A}_3} \frac{P(\tilde{y}_k|\tilde{y}_{k-1})}{P(x_k|x_{k-1})} = \min_{j \in \mathcal{A}_4} \frac{P(\tilde{y}_j|\tilde{y}_{j-1})}{P(x_j|x_{j-1})} = \frac{p_{10}}{p_{00}} = \frac{1 - p_{00}}{p_{00}}.$$

Thus, according to (10) and (11), we have

$$\begin{aligned} \prod_{k \in \mathcal{A}_3} \frac{Q(a_k|a_{k-1}) P(\tilde{y}_k|\tilde{y}_{k-1})}{Q(z_k|z_{k-1}) P(x_k|x_{k-1})} &\times \prod_{k \in \mathcal{A}_4} \frac{Q(a_k|a_{k-1}) P(\tilde{y}_k|\tilde{y}_{k-1})}{Q(z_k|z_{k-1}) P(x_k|x_{k-1})} \\ &\geq \prod_{k \in \mathcal{A}_3} \frac{Q(a_k|a_{k-1})(1 - p_{00})}{Q(2^q - 1 - a_k|a_{k-1})p_{00}} \\ &\times \prod_{k \in \mathcal{A}_4} \frac{Q(a_k|a_{k-1})(1 - p_{00})}{Q(a_k|2^q - 1 - a_{k-1})p_{00}} \\ &\geq \prod_{k \in \mathcal{A}_3} \frac{(1 - \varepsilon)\rho_{2^{q-1}-1}(1 - p_{00})}{(1 - \varepsilon)\rho_{2^{q-1}}p_{00}} \times \prod_{k \in \mathcal{A}_4} \frac{(1 - p_{00})}{p_{00}} \\ &= \prod_{i=1}^{|\mathcal{A}_3|} \frac{\rho_{2^{q-1}-1}}{\rho_{2^{q-1}}} \times \left[ \frac{1 - p_{00}}{p_{00}} \right]^2 \end{aligned} \quad (26)$$

where the second inequality follows by taking minimum over the first product and noting that  $(Q(a_k|a_{k-1})/Q(a_k|2^q - 1 - a_{k-1})) \geq 1$  for all  $k$ . Clearly, if (17) holds, then

$$\prod_{i=1}^{|\mathcal{A}_3|} \frac{\rho_{2^{q-1}-1}}{\rho_{2^{q-1}}} \times \left[ \frac{1 - p_{00}}{p_{00}} \right]^2 \geq 1.$$

Thus,  $\gamma \geq 1$ , and mapping  $\theta^*$  is an optimal MAP decoding rule.

To prove the converse, assume that (17) does not hold, i.e.,

$$\frac{\rho_{2^{q-1}-1}}{\rho_{2^{q-1}}} \times \left[ \frac{1 - p_{00}}{p_{00}} \right]^2 < 1.$$

Now, for

$$x^N = (0, 0, \dots, 0)$$

$$y^N = (0, 0, \dots, 0, 2^{q-1}, 0, \dots, 0)$$

where the only nonzero component of  $y^N$  is in an arbitrary position  $i \in \{2, \dots, N - 1\}$ , we have

$$\gamma = 1 \times 1 \times \dots \times 1 \times \left( \frac{\rho_{2^{q-1}-1}}{\rho_{2^{q-1}}} \times \left[ \frac{1 - p_{00}}{p_{00}} \right]^2 \right) \times 1 \times \dots \times 1 < 1.$$

Hence, if (17) does not hold, there exists some  $x^N$  and  $y^N$ , such that the mapping  $\theta^*$  does not decode optimally. ■

## REFERENCES

- [1] S. Shahidi, F. Alajaji, and T. Linder, "Exploiting memory and soft-decision information in channel optimized quantization for correlated fading channels," in *Proc. 12th Can. Workshop Inf. Theory*, Kelowna, BC, Canada, May 2011, pp. 66–69.
- [2] S. Shahidi, F. Alajaji, and T. Linder, "MAP decoding of quantized sources over soft-decision fading channels with memory," in *Proc. IEEE Int. Conf. Commun.*, Ottawa, ON, Canada, Jun. 2012, pp. 2277–2282.
- [3] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Convention Rec.*, vol. 4, pp. 142–163, 1959.

- [4] Y. Zhong, F. Alajaji, and L. L. Campbell, "On the joint source-channel coding error exponent for discrete memoryless systems," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1450–1468, Apr. 2006.
- [5] J. Lim and D. L. Neuhoff, "Joint and tandem source-channel coding with complexity and delay constraints," *IEEE Trans. Commun.*, vol. 51, no. 5, pp. 757–766, May 2003.
- [6] E. Ayanoglu and R. Gray, "The design of joint source and channel trellis waveform coders," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 855–865, Nov. 1987.
- [7] H. Kumazawa, M. Kasahara, and T. Namekawa, "A construction of vector quantizers for noisy channels," *Electron. Commun. Jpn. (Part I: Commun.)*, vol. 67, no. 4, pp. 39–47, 1984.
- [8] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source–Channel coding," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 6, pp. 827–838, Nov. 1987.
- [9] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inf. Theory*, vol. 36, no. 4, pp. 799–809, Jul. 1990.
- [10] N. Farvardin and V. Vaishampayan, "On the performance and complexity of channel-optimized vector quantizers," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 155–160, Jan. 1991.
- [11] N. Phamdo, F. Alajaji, and N. Farvardin, "Quantization of memoryless and Gauss–Markov sources over binary Markov channels," *IEEE Trans. Commun.*, vol. 45, no. 6, pp. 668–675, Jun. 1997.
- [12] D. J. Miller and M. Park, "A sequence-based approximate MMSE decoder for source coding over noisy channels using discrete hidden Markov models," *IEEE Trans. Commun.*, vol. 46, no. 2, pp. 222–231, Feb. 1998.
- [13] D. Miller and K. Rose, "Combined source–channel vector quantization using deterministic annealing," *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 347–356, Feb.–Apr. 1994.
- [14] M. Skoglund, "Soft decoding for vector quantization over noisy channels with memory," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1293–1307, May 1999.
- [15] N. Görtz, *Joint Source–Channel Coding of Discrete-Time Signals With Continuous Amplitudes*. London, U.K.: Imperial College Press, 2007.
- [16] P. Duhamel and M. Kieffer, *Joint Source–Channel Decoding: A Cross-Layer Perspective With Applications in Video Broadcasting over Mobile and Wireless Networks*. New York, NY, USA: Academic, 2009.
- [17] R. L. Dobrushin and M. S. Pinsker, "Memory increases transmission capacity," *Probl. Peredachi Inf.*, vol. 5, no. 1, pp. 94–95, 1969.
- [18] L. Zhong, F. Alajaji, and G. Takahara, "A binary communication channel with memory based on a finite queue," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2815–2840, Aug. 2007.
- [19] N. Phamdo and F. Alajaji, "Soft-decision demodulation design for COVQ over white, colored, and ISI Gaussian channels," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1499–1506, Sep. 2000.
- [20] N. Wernersson and M. Skoglund, "On source decoding based on finite-bandwidth soft information," in *Proc. Int. Symp. Inf. Theory*, Sep. 2005, pp. 87–91.
- [21] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3629–3639, Dec. 2009.
- [22] F. Behnamfar, F. Alajaji, and T. Linder, "Channel-optimized quantization with soft-decision demodulation for space-time orthogonal block-coded channels," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3935–3946, Oct. 2006.
- [23] C. Pimentel and F. Alajaji, "A discrete channel model for capturing memory and soft-decision information: A capacity study," in *Proc. Int. Conf. Commun.*, Jun. 2009, pp. 1–6.
- [24] C. Pimentel, F. Alajaji, and P. Melo, "A discrete queue-based model for soft-decision demodulated correlated fading channels," in *Proc. Veh. Technol. Conf.*, May 2011, pp. 1–5.
- [25] C. Pimentel, F. Alajaji, and P. Melo, "A discrete queue-based model for capturing memory and soft-decision information in correlated fading channels," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1702–1711, Jun. 2012.
- [26] F. Alajaji, N. Phamdo, N. Farvardin, and T. E. Fuja, "Detection of binary Markov sources over channels with additive Markov noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 230–239, Jan. 1996.
- [27] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. COM-28, no. 1, pp. 84–95, Jan. 1980.
- [28] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 957–1000, Jul./Aug. 1968.
- [29] N. Phamdo and N. Farvardin, "Optimal detection of discrete Markov sources over discrete memoryless channels-applications to combined source–channel coding," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 186–193, Jan. 1994.
- [30] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.
- [31] J. Devore, "A note on the observation of a Markov source through a noisy channel," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 6, pp. 762–764, Nov. 1974.
- [32] E. Dynkin, *Markov Processes*. Berlin, Germany: Springer-Verlag, 1965.
- [33] F. Kelly, "Markovian functions of a Markov chain," *Sankhyā: Ind. J. Stat., Ser. A*, vol. 44, no. 3, pp. 372–379, Oct. 1982.
- [34] S. Shahidi, "Robust lossy source coding for correlated fading channels," M.S. thesis, Dept. Elect. Comput. Eng., Queen's Univ., Kingston, ON, Canada, 2011.
- [35] S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.
- [36] Y. Zheng and C. Xiao, "Improved models for the generation of multiple uncorrelated Rayleigh fading waveforms," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 256–258, Jun. 2002.
- [37] N. Görtz, "On the iterative approximation of optimal joint source–channel decoding," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1662–1670, Sep. 2001.
- [38] J. Garcia-Frias, "Decoding of low-density parity-check codes over finite-state binary Markov channels," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1840–1843, Nov. 2004.
- [39] C. Nicola, F. Alajaji, and T. Linder, "Decoding LDPC codes over binary channels with additive Markov noise," in *Proc. CWIT*, Montreal, QC, Canada, Jun. 2005, pp. 187–190.
- [40] A. Ecford, F. Kschischang, and S. Pasupathy, "Analysis of low-density parity-check codes for the Gilbert-Elliott channel," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3872–3889, Nov. 2005.
- [41] J. Kliewer and R. Thobaden, "Iterative source–channel decoding of variable-length codes using residual source redundancy," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 919–929, May 2005.
- [42] L. Schmalen, M. Adrat, T. Clevern, and P. Vary, "EXIT chart based system design for iterative source–channel decoding with fixed-length codes," *IEEE Trans. Commun.*, vol. 59, no. 9, pp. 2406–2413, Sep. 2011.



**Shervin Shahidi** (S'12) was born in Isfahan, Iran, in 1987. He received the B.Sc. degree in electrical engineering from Isfahan University of Technology, Isfahan, Iran, in 2009 and the M.S. degree in electrical engineering from Queen's University, Kingston, ON, Canada, in 2011. He is currently working toward the Ph.D. degree in electrical engineering with the University of Toronto, Toronto, ON.

His research interests are communication systems, wireless communications, coding, probabilistic methods, and algorithms.



**Fady Alajaji** (S'90–M'94–SM'00) received the B.E. degree with distinction from the American University of Beirut, Beirut, Lebanon, in 1988 and the M.Sc. and Ph.D. degrees from the University of Maryland, College Park, MD, USA, in 1990 and 1994, respectively, all in electrical engineering.

In 1994, he held a postdoctoral appointment with the Institute for Systems Research, University of Maryland. Since 1995, he has been with the Department of Mathematics and Statistics, Queen's University, Kingston, ON, Canada, where he is currently

a Professor of mathematics and engineering and, since 1997, has been cross-appointed with the Department of Electrical and Computer Engineering. From 2003 to 2008, he served as the Chair of the Queen's University Mathematics and Engineering Program. His research interests include information theory, digital communications, error control coding, joint source–channel coding, and data compression.

Dr. Alajaji has served as an Organizer and Technical Program Committee member for several international conferences and workshops. He currently serves as the Area Editor and the Editor for Source and Source–Channel Coding for the IEEE TRANSACTIONS ON COMMUNICATIONS. He received the Premier's Research Excellence Award from the Province of Ontario.



**Tamás Linder** (S'92–M'93–SM'00–F'13) was born in Budapest, Hungary, in 1964. He received the M.S. degree in electrical engineering from the Technical University of Budapest in 1988 and the Ph.D. degree in electrical engineering from the Hungarian Academy of Sciences, Budapest, in 1992.

In 1992, he was a Postdoctoral Researcher with the University of Hawaii, Honolulu, HI, USA, and from 1993 to 1994, he was a Fulbright Visiting Scholar with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL, USA.

From 1994 to 1998, he was a faculty member with the Department of Computer Science and Information Theory, Technical University of Budapest. From 1996 to 1998, he was also a Visiting Research Scholar with the Department of Electrical and Computer Engineering, University of California, San Diego, CA, USA. Since 1998, he has been with Queen's University, Kingston, ON, Canada, where he is currently a Professor of mathematics and engineering with the Department of Mathematics and Statistics. His research interests include communications and information theory, source coding and vector quantization, machine learning, and statistical pattern recognition.

Dr. Linder was an Associate Editor for Source Coding for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2003 to 2004. He received the Premier's Research Excellence Award of the Province of Ontario in 2002 and the Chancellor's Research Award of Queen's University in 2003.