This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

Index Mapping for Bit-error Resilient Multiple Description Lattice Vector Quantizer

Sorina Dumitrescu *Senior Member, IEEE*, Yifang Chen, and Jun Chen *Senior Member, IEEE*

***Abstract*—In conventional multiple description coding (MDC) two descriptions of a source are generated and sent over on/off channels. In this work we are interested in exploiting the redundancy built in MDC to additionally confer robustness against other channel errors. In particular, we consider a multiple description lattice vector quantizer (MDLVQ) whose output (a pair of side lattice points) is mapped to a pair of binary indexes and each index is sent over a binary channel. One channel is noiseless, while the other one is noisy. Thus, at the decoder one description is received error-free, while the other one may carry bit errors. Then the decoder uses the error-free description as side information to improve the reconstruction. The effectiveness of the decoder in alleviating the impact of bit errors depends on the mapping γ of side lattice points to binary indexes.**

We propose the design of a structured bit-error resilient mapping γ. For this, the set of side lattice points is first partitioned using Voronoi regions of an appropriate coarse lattice. Next a good linear channel code is selected, each Voronoi region is assigned a coset of this channel code and the side lattice points within each Voronoi region are mapped to binary sequences in the corresponding coset. Additionally, we argue that the performance of γ is improved by assigning cosets close in Hamming distance to neighboring Voronoi regions, and propose a technique to achieve this goal.

We derive a lower bound on the error correction performance of the proposed mapping γ in terms of the performance of the channel code C used in its construction. Interestingly, we prove that, as the rate of the MDLVQ grows to infinity, the mapping γ becomes as good as the code C in correcting bit errors. Simulation results show the significant superiority of the proposed index mapping versus random mappings.

***Index Terms*—Multiple description lattice vector quantizer, bit-error resilience, structured index mapping, error correction performance, asymptotic analysis.**

I. Introduction

Multiple description coding (MDC) generates multiple descriptions of a signal, such that every description can lead to an acceptable reconstruction quality, while the quality gradually increases with the number of descriptions that are successfully decoded. Some tutorial papers on MDC are [1], [2], [3].

Conventionally, the design of MDC has focused mainly on ensuring robustness against description loss. On the other hand, it is natural to expect that the redundancy which is introduced in the system could also be used to mitigate the effect of other channel impairments such as bit errors. Indeed, it was shown in prior work that by using joint source-channel decoding at the central decoder the performance in case of channel errors can be improved [4]-[13]. In particular, the authors of [4] use a two description scalar quantizer (2DSQ) for delay-constrained transmission over a slow fading Rayleigh channel. They show that the 2DSQ significantly outperforms a diversity system consisting of a maximum ratio combiner with the same interleaving delay. In [5], [6] each description generated by a 2DSQ is first applied a convolutional coder and then sent over an AWGN channel using BPSK. The decoder applies an iterative joint source-channel decoding algorithm using the turbo principle, which exploits the correlation between descriptions. Work [6] also develops such an algorithm for the case without the channel coder. Work [7] considers the transmission of a Markov sequence over noisy channels using a 2DSQ with uniform quantizers followed by variable length source codes. They develop a soft decoder based on belief propagation principles. In [9] a joint source-channel coding scheme is proposed, which combines MDC with space-time turbo-coded modulation. The authors show that the system using MDC performs considerably better than a single-description coding scheme for moderate levels of signal-to-noise ratio (SNR). The authors of [8] compare strategies for the transmission of a source signal over two independent, nonergodic channels with random states. They use as a performance measure the so-called distortion exponent, which measures how fast the average distortion decays with the SNR. They compare the use of MDC with channel coding on each individual channel, combined with joint source-channel decoding, against a system sending a single description with optimal channel coding across both channels. They prove that the two strategies have the same performance for continuous state channels, while in the case of on/off channels the former strategy is strictly better.

The above observations raise the question of how to construct systems that are robust against both description loss and bit errors. This problem was addressed in [11], [12], [13]. More specifically, in [11], [12] heuristic algorithms were applied for the design of the index assignment in the case of 2DSQs and multiple description vector quantizers, respectively. On the other hand, in [13] the problem of improving the bit-error resilience of a 2DSQ was addressed by careful design of permutation mappings applied to each description. More specifically, the construction of structured mappings with guaranteed minimum Hamming distance properties, relying on known channel codes, was proposed.

In this work we address the design of a bit-error resilient symmetric MDLVQ, where MDLVQ stands for multiple de-

S. Dumitrescu and J. Chen are with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (emails: sorina/junchen@mail.ece.mcmaster.ca). Y. Chen was with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (email: cheny87@mcmaster.ca).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

scription lattice vector quantizer[1]. An n-dimensional symmetric MDLVQ consists of a so-called *central* lattice $\Lambda_c \subseteq \mathbb{R}^n$, a so-called *side* lattice $\Lambda_s \subseteq \Lambda_c$, and an injective mapping $\alpha : \Lambda_c \to \Lambda_s \times \Lambda_s$, termed the index assignment (IA). The encoder of the MDLVQ first quantizes the input vector to the closest central lattice point $\boldsymbol{\lambda}_c \in \Lambda_c$. This quantizer is also referred to as the *central* quantizer. Next the encoder determines the pair $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \alpha(\boldsymbol{\lambda}_c)$ and outputs $\boldsymbol{\lambda}_i$ as the ith description, $i = 1, 2$. Further, $\boldsymbol{\lambda}_i$ is assigned a binary sequence j_i of some fixed bit-length nR, via a mapping γ_i. The binary index j_i is further sent over a separate binary channel, referred to as channel i.

In the conventional MDLVQ it is assumed that each channel either works correctly or breaks down. Thus, at the receiver, either both j_1 and j_2 are received, in which case the central decoder outputs $\boldsymbol{\lambda}_c$, or only j_i, for some $i = 1, 2$, is received, in which case the side decoder i outputs $\boldsymbol{\lambda}_i$ as the reconstruction. Notice that, for fixed Λ_c and Λ_s, the performance of the MDLVQ is determined by the IA α. The design of good IAs for conventional MDLVQ was addressed in [14]-[20].

In this work we will consider the case when one channel is noiseless, while the other one is noisy. Our work is motivated by the situation where each description, after being applied channel coding followed by modulation, is transmitted over a wireless channel. Specifically, let $j_{i,t}$ denote the binary index output by encoder i at time instance t and let $\boldsymbol{j}_i$ denote the concatenation of T outputs, i.e., $\boldsymbol{j}_i = (j_{i,1}, \cdots, j_{i,T})$. The binary sequence $\boldsymbol{j}_i$ is further applied a CRC (cyclic redundancy check code) for error detection, followed by a systematic[2] binary block channel code. The channel codeword is modulated and transmitted over a wireless physical channel. When the quality of the wireless channel is worse than expected (due to fading, shadowing or interference from other transmitters), the channel decoder fails to correct the errors, situation which is detected by the CRC decoder. In such a case let $\boldsymbol{j}'_i = (j'_{i,1}, \cdots, j'_{i,T})$ denote the systematic part of the received bit sequence, where each $j'_{i,t}$ is a (possibly) bit-error corrupted version of $j_{i,t}$, $1 \leq t \leq T$. Thus, there are four scenarios that can occur: I) the outputs are $\boldsymbol{j}_1$ and $\boldsymbol{j}_2$; II) the outputs are $\boldsymbol{j}_1$ and $\boldsymbol{j}'_2$; III) the outputs are $\boldsymbol{j}'_1$ and $\boldsymbol{j}_2$; IV) the outputs are $\boldsymbol{j}'_1$ and $\boldsymbol{j}'_2$. More importantly, the decoder has knowledge of the scenario that occurs.

From this point on the decoding may proceed in several ways. One possibility is to discard the erroneous sequences $\boldsymbol{j}'_1$, $\boldsymbol{j}'_2$. Such a system can be modeled by the conventional MDC problem since cases II and III reduce to the case when only one description is received, while case IV corresponds to both descriptions being lost. Another possibility is to further apply joint decoding which exploits the correlation between the descriptions, in all cases II, III and IV. As shown in [6] joint decoding can improve the performance. In this work we will consider joint decoding only in cases II and III, while in scenario IV we discard the erroneous sequences. The joint decoding will be performed at each time instance t, i.e., for each pair $(j_{1,t}, j'_{2,t})$ in case II, respectively each pair $(j'_{1,t}, j_{2,t})$ in case III. In the sequel the term "channel i", for $i = 1, 2$, will refer to the channel with input j_i and output j_i or j'_i, incorporating the composite action of the channel encoder, modulator, transmission over the physical wireless channel, followed by demodulator and the channel decoder.

The goal of our work is to design the mapping γ_i to improve the performance at the aforementioned decoder. Notice that since γ_i is a bijection, it does not affect the performance (i.e., it does not change the distortion or rate) when channel i is noiseless. Thus, for the design of γ_i it is enough to focus on the case when channel i is noisy, while the other channel is noiseless. Note that the decoder knows which channel is noisy, but the encoder does not. Fig. 1 depicts the system under consideration in the case when only channel 2 is noisy.

The main contribution of this work is the design of a structured mapping γ_i with good error-correction performance when channel i is noisy, while the other channel is error-free. For this, the set of side lattice points is first partitioned using the Voronoi regions of a properly chosen coarse lattice. Next γ_i is constructed based on two other mappings φ and ψ. The mapping φ maps the side lattice points situated in the Voronoi region of the origin to binary sequences of a good linear channel code C. The mapping ψ maps the Voronoi regions of the coarse lattice to cosets of the chosen channel code. Next we prove that the so-called *local Hamming distance* of ψ, defined as the largest Hamming distance between cosets assigned to neighbouring Voronoi regions, is of crucial importance for the performance of γ_i. Therefore, we propose the design of ψ with small local Hamming distance, based on n one-dimensional Gray mappings. Furthermore, we evaluate the performance of the proposed mapping γ_i in terms of the performance of the channel code C used in its design, and of the local Hamming distance of ψ. We show that the error-correction ability of γ_i reaches the same level as the code C, asymptotically as the rate of the MDLVQ approaches ∞. We also prove that, as the rate of the MDLVQ and the quantizer dimension n go to ∞, the redundancy added in the channel code approaches the value of the redundancy built in the MDLVQ system, which is defined as the difference between the sum-rate and the no excess rate. Thus, by selecting a capacity achieving channel code C for the construction of γ_i, it is guaranteed that the whole redundancy introduced in the MDLVQ system is effectively utilized toward error correction. Finally, experiments are carried out to assess the practical performance of the proposed design. The results show significant performance improvement using the proposed index mapping compared with random mappings. We point out that the conference paper [21] is a shorter version of this work, which omits all the proofs and does not include Proposition 3. Additionally, the present work contains more examples, experimental results and discussions.

It is worth highlighting the differences versus prior work [13]. In [13] the design of bit-error resilient mappings was carried out in the case of 2DSQ for both scenarios where only one or both descriptions may carry bit errors. However, the extension of those designs from the scalar to the vector case is not straightforward. In the current work we consider

[1]The attribute "symmetric" refers to the fact that both descriptions have the same rate and lead to the same distortion when each of them is used alone for the source reconstruction.

[2]A systematic channel code is a channel code where the information sequence is part of the channel codeword.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

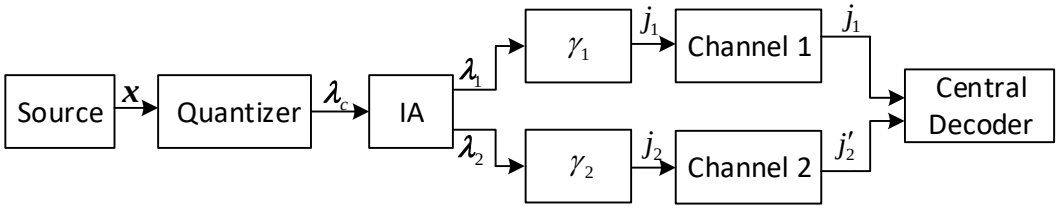

Fig. 1: MDLVQ with bit-errors.

only the scenario when one description is correct, but develop a general framework for the construction of the mapping, applicable to higher dimensions, not just to the scalar case. Additionally, we include the analysis of performance in terms of probability of correct decoding (not just Hamming distance, as in [13]), and the asymptotical analysis of performance as the rate approaches infinity (not present in [13]).

The rest of this paper is organized as follows. Section II presents the main definitions and notations. Section III introduces the structured construction of the mapping γ_i based on a coarse lattice and two other mappings φ and ψ. Additionally, the error correction performance of γ_i is assessed in terms of the channel code used in its construction and of the local Hamming distance of ψ. Section IV presents a technique to construct the mapping ψ. Section V discusses the asymptotical error correction performance of the proposed mapping γ as the rate of the MDLVQ approaches infinity. Finally, Section VI presents experimental results and Section VII concludes the paper.

II. Preliminaries

We will use $\mathbb{F}_2$ to denote the binary field. The lower case letters in bold are used to denote row vectors. The upper case letters in bold will be used to denote matrices. Let $w_H(\boldsymbol{s})$ be the Hamming weight of a binary vector $\boldsymbol{s}$. For any binary vectors $\boldsymbol{s}$ and $\boldsymbol{t}$ with the same dimension, define the Hamming distance $d_H(\boldsymbol{s},\boldsymbol{t})$ as $w_H(\boldsymbol{s}\oplus\boldsymbol{t})$, where "$\oplus$" denotes the component-wise addition in the binary field. Further, define $d_H(A)$ for a set $A \subseteq \mathbb{F}_2^m$, as the minimum Hamming distance between any two distinct bit sequences in A. For $\boldsymbol{s} \in \mathbb{F}_2^m$, denote by $d_H(\boldsymbol{s}, A)$ the smallest Hamming distance between $\boldsymbol{s}$ and any point in A. Additionally, let $\overline{A}$ denote the complement of A in $\mathbb{F}_2^m$.

For each $q, 0 < q < 1$, let $H(q) \triangleq -q\log_2 q - (1-q)\log_2(1-q)$. For any finite set B, let $|B|$ denote its cardinality. For any n-by-n matrix $\boldsymbol{M}$ with elements in $\mathbb{R}$, let $\|\boldsymbol{M}\|$ denote its l_2-norm, defined as $\|\boldsymbol{M}\| = \sup_{\|\boldsymbol{x}\|=1} \|\boldsymbol{x}\boldsymbol{M}\|$, where $\|\boldsymbol{x}\|$ is the Euclidean norm of the vector $\boldsymbol{x} \in \mathbb{R}^n$. Additionally, if the matrix is invertible then $\boldsymbol{M}^{-1}$ denotes its inverse. The volume of a measurable set $\mathcal{S} \subseteq \mathbb{R}^n$ is denoted by $vol(\mathcal{S})$.

A lattice $\Lambda' \subseteq \mathbb{R}^n$ is a discrete set of points $\Lambda' = \{\boldsymbol{\lambda}' = \boldsymbol{z}\boldsymbol{G}', \boldsymbol{z} \in \mathbb{Z}^n\}$, where $\boldsymbol{G}'$ is an n-by-n non-singular matrix whose elements are real numbers. Thus, the lattice consists of all points with integer coordinates with respect to the basis formed by the row vectors of the generator matrix. The *fundamental parallelepiped* of the lattice [25] is defined as the set of all points in $\mathbb{R}^n$ which are linear combinations of the basis vectors, with coefficients between 0 and 1. The *fundamental volume* of the lattice is the volume of the fundamental parallelepiped and is denoted by $v(\Lambda')$. The nearest-neighbor quantizer associated with the lattice Λ', denoted by $Q_{\Lambda'}$, maps each $\boldsymbol{x} \in \mathbb{R}^n$ to its closest lattice point, i.e.,

$$Q_{\Lambda'}(\boldsymbol{x}) \triangleq \arg\min_{\boldsymbol{\lambda}' \in \Lambda'} \|\boldsymbol{x} - \boldsymbol{\lambda}'\|. \tag{1}$$

For each $\boldsymbol{\lambda}' \in \Lambda'$, the *Voronoi region* of $\boldsymbol{\lambda}'$ in Λ', denoted by $V[\boldsymbol{\lambda}' : \Lambda']$, is the set of vectors $\boldsymbol{x} \in \mathbb{R}^n$ mapped to $\boldsymbol{\lambda}'$ by quantizer $Q_{\Lambda'}$. The ties are broken in a systematic manner such that the following relation holds for all $\boldsymbol{\lambda}' \in \Lambda'$,

$$V[\boldsymbol{\lambda}' : \Lambda'] = \boldsymbol{\lambda}' + V[\boldsymbol{0} : \Lambda'],$$

where $\boldsymbol{x} + \mathcal{S} \triangleq \{\boldsymbol{x} + \boldsymbol{y} | \boldsymbol{y} \in \mathcal{S}\}$, for any $\boldsymbol{x} \in \mathbb{R}^n$ and $\mathcal{S} \subset \mathbb{R}^n$. Thus, all Voronoi regions are congruent. Moreover, the volume of a Voronoi region equals $v(\Lambda')$.

For any set $\mathcal{S} \subseteq \mathbb{R}^n$, let $cl(\mathcal{S})$ denote the closure of the set $\mathcal{S}$, i.e., the set consisting of all points of $\mathcal{S}$ together with all limit points of $\mathcal{S}$ [26, Def. 3.63]. Then we have

$$cl(V[\boldsymbol{\lambda}' : \Lambda']) = \{\boldsymbol{x} \in \mathbb{R}^n | \|\boldsymbol{x} - \boldsymbol{\lambda}'\| \leq \|\boldsymbol{x} - \boldsymbol{\lambda}''\| \text{ for any } \boldsymbol{\lambda}'' \in \Lambda'\}.$$

It is worth pointing out that, according to our definition of the Voronoi region, which follows [25], not all the points on the boundary of $V[\boldsymbol{\lambda}' : \Lambda']$ are included in $V[\boldsymbol{\lambda}' : \Lambda']$, therefore $cl(V[\boldsymbol{\lambda}' : \Lambda']) \neq V[\boldsymbol{\lambda}' : \Lambda']$[3]. Two lattice points $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \Lambda'$ are said to be *adjacent* if the closures of their Voronoi regions have points in common. Further, for each $\boldsymbol{x} \in \mathbb{R}^n$ we define

$$\boldsymbol{x} \bmod \Lambda' \triangleq \boldsymbol{x} - Q_{\Lambda'}(\boldsymbol{x}).$$

In this work, we assume a fixed-rate symmetric MDLVQ. Consider a vector source $\boldsymbol{X}$ with elements in $\mathbb{R}^n$, with probability density function (pdf) $\boldsymbol{f}(\boldsymbol{x})$. We will assume that the support of the pdf is included in some bounded measurable set $\mathcal{A}$[4]. Recall that the MDLVQ comprises two lattices, Λ_c and $\Lambda_s \subset \Lambda_c$, and an IA mapping $\alpha : \Lambda_c \to \Lambda_s \times \Lambda_s$. Let $\mathcal{B}_c \triangleq Q_{\Lambda_c}(\mathcal{A})$. Let N denote the index of Λ_s with respect to Λ_c, i.e., $N \triangleq v(\Lambda_s)/v(\Lambda_c)$. For $i = 1, 2$, let $\mathcal{B}_i$ denote the codebook of description i, i.e., $\mathcal{B}_1 \triangleq \{\boldsymbol{\lambda}_1 \in \Lambda_s | \alpha(\boldsymbol{\lambda}_c) = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2),$ for some $\boldsymbol{\lambda}_c \in \mathcal{B}_c, \boldsymbol{\lambda}_2 \in \Lambda_s\}$, and $\mathcal{B}_2 \triangleq \{\boldsymbol{\lambda}_2 \in \Lambda_s | \alpha(\boldsymbol{\lambda}_c) = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2),$ for some $\boldsymbol{\lambda}_c \in \mathcal{B}_c, \boldsymbol{\lambda}_1 \in \Lambda_s\}$. As in most of the prior work on MDLVQ [14]-[18], [20] we will assume that the IA is shift invariant, i.e., for any $\boldsymbol{\lambda}_c \in \Lambda_c$ and $\boldsymbol{\lambda}_s \in \Lambda_s$, the following holds: $\alpha(\boldsymbol{\lambda}_c + \boldsymbol{\lambda}_s) = (\boldsymbol{\lambda}_s, \boldsymbol{\lambda}_s) + \alpha(\boldsymbol{\lambda}_c)$.

[3]Note that the boundary of $V[\boldsymbol{\lambda}' : \Lambda']$ consists of all points $\boldsymbol{x}$ in $cl(V[\boldsymbol{\lambda}' : \Lambda'])$ such that $\|\boldsymbol{x} - \boldsymbol{\lambda}'\| = \|\boldsymbol{x} - \boldsymbol{\lambda}''\|$ for some $\boldsymbol{\lambda}'' \in \Lambda' \setminus \{\boldsymbol{\lambda}'\}$.

[4]In practice, for general distributions we can choose a bounded set $\mathcal{A}$ such that $\mathbb{P}[\boldsymbol{X} \notin \mathcal{A}]$ is very small and the following development can be adapted with no essential modifications.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

Further, for each $i = 1, 2$, let $\gamma_i : \mathcal{B}_i \longrightarrow \{0,1\}^{nR}$ be an injective mapping, where R must satisfy the condition $nR \geq \max_{i=1,2} \lceil \log_2 |\mathcal{B}_i| \rceil$. Note that R is the rate of each description. The MDLVQ encoder operates as follows. For $\boldsymbol{x} \in \mathcal{A}$, the central quantizer determines $\boldsymbol{\lambda}_c = Q_{\Lambda_c}(\boldsymbol{x})$. Then the binary index $j_i = \gamma_i(\boldsymbol{\lambda}_i)$ is sent over the ith channel for $i = 1, 2$, where $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \alpha(\boldsymbol{\lambda}_c)$.

As pointed out in the introductory section we are interested in the scenario when only one of the channels is noisy. The decoder knows which channel is error free, but the encoder does not. Since the mapping γ_i does not affect the performance when channel i is noiseless, the mappings γ_1 and γ_2 can be designed separately. The design of γ_1 and γ_2 is similar, therefore, we will focus on the design of γ_2. For this it is sufficient to consider the case when only the second channel is noisy (Fig. 1). More specifically, we will assume that channel 2 is the nRth extension[5] of a binary symmetric channel (BSC) with crossover probability $p, 0 < p < 1/2$, denoted by BSC(p). For simplicity, we will use the notation γ instead of γ_2. Thus, the central decoder receives indexes j_1 and j_2' over the first, respectively second channel. The decoder recovers $\boldsymbol{\lambda}_1$ as $\gamma_1^{-1}(j_1)$. Then it uses a maximum likelihood decoder, which exploits the knowledge of $\boldsymbol{\lambda}_1$ as side information, to find an estimate $\hat{j}_2$ of the index transmitted over the second channel. For each $\boldsymbol{\lambda}_1 \in \Lambda_s$ denote by $\mathcal{W}(\boldsymbol{\lambda}_1)$ the set of side lattice points $\boldsymbol{\lambda}_2$ such that $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \alpha(\boldsymbol{\lambda}_c)$ for some $\boldsymbol{\lambda}_c \in \Lambda_c$. Then the decoder searches for the index $\hat{j}_2 \in \gamma\left(\mathcal{W}(\boldsymbol{\lambda}_1) \cap \mathcal{B}_2\right)$, which is closest in Hamming distance to j_2'. Then it determines $\hat{\boldsymbol{\lambda}}_2 = \gamma^{-1}(\hat{j}_2)$ and finally outputs the reconstruction $\hat{\boldsymbol{\lambda}}_c = \alpha^{-1}(\boldsymbol{\lambda}_1, \hat{\boldsymbol{\lambda}}_2)$.

We will evaluate the performance of the mapping γ using the maximal conditional probability of error, denoted by $P_e(\gamma)$, and defined as follows,

$$P_e(\gamma) \triangleq \max_{(j_1, j_2)} \mathbb{P}[\hat{J}_2 \neq J_2 | (J_1, J_2) = (j_1, j_2)],$$

where J_1, J_2 and $\hat{J}_2$ denote the random variables associated to indexes j_1, j_2 and $\hat{j}_2$, respectively. As explained previously, the central decoder needs to look only in a subset of $\mathbb{F}_2^{nR}$, namely in $\gamma(\mathcal{W}(\boldsymbol{\lambda}_1) \cap \mathcal{B}_2)$, in order to estimate $\hat{j}_2$. Thus, the set $\gamma(\mathcal{W}(\boldsymbol{\lambda}_1) \cap \mathcal{B}_2)$ can be regarded as the channel code used for the transmission of $\boldsymbol{\lambda}_2$ when the first description lattice point is $\boldsymbol{\lambda}_1$. Then the performance of the mapping γ is determined by the individual performance of each code $\gamma(\mathcal{W}(\boldsymbol{\lambda}_1) \cap \mathcal{B}_2)$, where $\boldsymbol{\lambda}_1 \in \mathcal{B}_1$. Since another way of assessing the error-correction performance of a channel code is in terms of its minimum Hamming distance, we will also consider such a measure. Therefore, let us define the *minimum window Hamming distance* of the mapping γ as

$$d_{H,W}(\gamma) = \min_{\boldsymbol{\lambda}_1 \in \mathcal{B}_1} d_H(\gamma(\mathcal{W}(\boldsymbol{\lambda}_1) \cap \mathcal{B}_2)).$$

Note that the decoder is able to correct all $\lfloor (d_{H,W}(\gamma) - 1)/2 \rfloor$ bit errors, therefore, the value of $d_{H,W}(\gamma)$ can also be used to measure the performance of the mapping γ.

[5]The nRth extension of a BSC is the channel obtained by applying a BSC to each bit of a binary sequence of length nR.

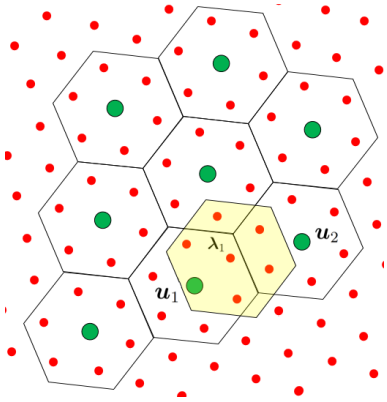

Fig. 2: Partitioning using Voronoi regions of the coarse lattice. The small red-filled circles are points of Λ_s, while big green-filled circles represent points of the coarse lattice $\Lambda \subset \Lambda_s$. The hexagons centered at the coarse lattice points are the Voronoi regons of the coarse lattice. The set $\lambda_1 + \mathcal{W}$ (highlighted) contains λ_1 and the six side lattice points closest to λ_1 and it intersects three Voronoi regions of the coarse lattice.

III. Design of Mapping γ

The proposed design of the mapping γ relies on a key property, which follows from the shift invariance of the IA, namely

$$\mathcal{W}(\boldsymbol{\lambda}_1) = \boldsymbol{\lambda}_1 + \mathcal{W}, \tag{2}$$

for any $\boldsymbol{\lambda}_1 \in \Lambda_s$, where $\mathcal{W} = \mathcal{W}(\mathbf{0})$. In order to prove it take an arbitrary $\boldsymbol{\lambda}_2 \in \mathcal{W}(\boldsymbol{\lambda}_1)$. Then there exists $\boldsymbol{\lambda}_c \in \Lambda_c$ such that $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \alpha(\boldsymbol{\lambda}_c)$. The shift invariance of the IA implies that $\alpha(\boldsymbol{\lambda}_c) = (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_1) + \alpha(\boldsymbol{\lambda}_c - \boldsymbol{\lambda}_1)$, leading further to $\alpha(\boldsymbol{\lambda}_c - \boldsymbol{\lambda}_1) = (\mathbf{0}, \boldsymbol{\lambda}_2 - \boldsymbol{\lambda}_1)$, which implies that $\boldsymbol{\lambda}_2 \in \boldsymbol{\lambda}_1 + \mathcal{W}$. The proof that $\boldsymbol{\lambda}_1 + \mathcal{W} \subseteq \mathcal{W}(\boldsymbol{\lambda}_1)$ follows similarly.

According to the discussion at the end of the previous section, for a good performance of the mapping γ we would like each subset of $\mathcal{B}_2$ of the form $\boldsymbol{\lambda}_1 + \mathcal{W}$, to be mapped to a good channel code for a BSC. To this end we will use cosets of a linear channel code. It is known that any coset of a linear channel code $C \subset \mathbb{F}_2^{nR}$, has the same performance as C, over a BSC with maximum likelihood decoding. Additionally, all cosets of C have the same size as C and form a partition of $\mathbb{F}_2^{nR}$. Then we will partition the set of lattice points $\mathcal{B}_2$ using Voronoi cells of a coarser lattice $\Lambda \subset \Lambda_s$ such that each such Voronoi cell includes a subset $(\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2$ for some $\boldsymbol{\lambda}_1$. Further, we map the lattice points in each Voronoi cell to the binary sequences in a coset of C. The mapping of the Voronoi cells of the coarse lattice to cosets of C can be represented as a mapping ψ from the set of coarse lattice points to the set C' of coset representatives. Finally, the task of ensuring that the sets of the form $(\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2$, which intersect with several Voronoi cells of the coarse lattice (Fig. 2), are also mapped to good channel codes for a BSC, falls on the design of the mapping ψ.

Therefore, consider a coarse lattice $\Lambda \subset \Lambda_s$ such that

$$V[\mathbf{0} : \Lambda] \supseteq \mathcal{W}. \tag{3}$$

Note that since $\mathcal{W}$ is finite, a sublattice Λ of Λ_s satisfying the above condition is guaranteed to exist. For instance, it could be

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

constructed by scaling Λ_s with an integer scaling factor until the Voronoi region $V[\mathbf{0}:\Lambda]$ becomes big enough to satisfy (3). Further, denote by $\mathcal{W}'$ the set of side lattice points which are in the Voronoi region $V[\mathbf{0}:\Lambda]$, i.e., $\mathcal{W}' \triangleq V[\mathbf{0}:\Lambda] \cap \Lambda_s$. Then relation (3) implies that $\mathcal{W} \subseteq \mathcal{W}'$. Note that Λ should be chosen such that (3) holds and $|\mathcal{W}' \setminus \mathcal{W}|$ is as small as possible. Now let us denote by $\mathcal{B}$ the set of points $Q_\Lambda(\boldsymbol{\lambda}_s)$ for $\boldsymbol{\lambda}_s \in \mathcal{B}_2$. This implies that $\mathcal{B}$ is the smallest subset of Λ such that $\bigcup_{\boldsymbol{u}\in\mathcal{B}} V[\boldsymbol{u}:\Lambda] \supseteq \mathcal{B}_2$, and since $V[\boldsymbol{u}:\Lambda] = \boldsymbol{u} + V[\mathbf{0}:\Lambda]$ and $\mathcal{B}_2 \subset \Lambda_s$, it further follows that $\bigcup_{\boldsymbol{u}\in\mathcal{B}} (\boldsymbol{u} + \mathcal{W}') \supseteq \mathcal{B}_2$. Additionally, note that since $\mathcal{B}_2$ is finite, $\mathcal{B}$ must also be finite.

Further, let $k \triangleq \lceil \log_2 |\mathcal{W}'| \rceil$ and let R be selected such that nR is an integer and $nR \geq \lceil \log_2 |\mathcal{B}| \rceil + k$. Next choose two vector subspaces C and C' of the vector space $\mathbb{F}_2^{nR}$, such that C' is a set of coset representatives of C, $|C| = 2^k$ and $|C'| = 2^{nR-k}$. Then construct two injective mappings

$$\varphi : \mathcal{W}' \longrightarrow C, \quad \psi : \mathcal{B} \longrightarrow C'.$$

Finally, we define the mapping $\gamma : \mathcal{B}_2 \to \mathbb{F}_2^{nR}$ as follows. For every $\boldsymbol{\lambda} \in \mathcal{B}_2$

$$\gamma(\boldsymbol{\lambda}) \triangleq \varphi(\boldsymbol{\lambda} \bmod \Lambda) \oplus \psi(Q_\Lambda(\boldsymbol{\lambda})). \tag{4}$$

Now we will prove that the mapping γ is injective. It is known that since C' is a set of coset representatives of C, then for any $\boldsymbol{b} \in \mathbb{F}_2^{nR}$ there is a unique pair $\boldsymbol{c} \in C$, $\boldsymbol{c}' \in C'$ such that $\boldsymbol{b} = \boldsymbol{c} \oplus \boldsymbol{c}'$ [27]. Then the equality $\gamma(\boldsymbol{\lambda}_1) = \gamma(\boldsymbol{\lambda}_2)$ implies that $\varphi(\boldsymbol{\lambda}_1 \bmod \Lambda) = \varphi(\boldsymbol{\lambda}_2 \bmod \Lambda)$ and $\psi(Q_\Lambda(\boldsymbol{\lambda}_1)) = \psi(Q_\Lambda(\boldsymbol{\lambda}_2))$. Since the mappings φ and ψ are injective it further follows that $\boldsymbol{\lambda}_1 \bmod \Lambda = \boldsymbol{\lambda}_2 \bmod \Lambda$ and $Q_\Lambda(\boldsymbol{\lambda}_1) = Q_\Lambda(\boldsymbol{\lambda}_2)$. Using further the fact that $\boldsymbol{x} = \boldsymbol{x} \bmod \Lambda + Q_\Lambda(\boldsymbol{x})$, for any $\boldsymbol{x} \in \mathbb{R}^n$, leads to $\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2$.

Further, notice that the above construction ensures that for each $\boldsymbol{\lambda}_1 = \boldsymbol{u} \in \mathcal{B}$, the points in $\boldsymbol{\lambda}_1 + \mathcal{W}$ are mapped by γ to a coset of C, which has the same performance as C over BSC(p) with the maximum likelihood decoder. However, for side lattice points $\boldsymbol{\lambda}_1$ which are not in the coarse lattice Λ, the set $\boldsymbol{\lambda}_1 + \mathcal{W}$ may intersect several Voronoi cells of the coarse lattice as shown in Fig. 2, thus the set $\gamma(\boldsymbol{\lambda}_1 + \mathcal{W})$ may not be entirely included in a single coset of C. Next we will address the design of the mapping ψ to ensure that the sets $\gamma(\boldsymbol{\lambda}_1 + \mathcal{W})$ are also good channel codes for the points $\boldsymbol{\lambda}_1$ in the latter category.

Clearly, the Voronoi cells intersected by $\boldsymbol{\lambda}_1 + \mathcal{W}$ are close in Euclidian distance. This observation suggests that the Hamming distance of images through ψ of coarse lattice points which are close in Euclidian distance, is linked to the performance of γ. In order to formally state the result we need a few more notations. Two distinct coarse lattice points $\boldsymbol{u}, \boldsymbol{u}' \in \Lambda$ are said to be *neighbors*, if there are $\boldsymbol{\lambda}, \boldsymbol{\lambda}' \in \Lambda_s$ such that $Q_\Lambda(\boldsymbol{\lambda}) = \boldsymbol{u}$, $Q_\Lambda(\boldsymbol{\lambda}') = \boldsymbol{u}'$ and $\boldsymbol{\lambda} \in \boldsymbol{\lambda}' + \mathcal{W}'$ or $\boldsymbol{\lambda}' \in \boldsymbol{\lambda} + \mathcal{W}'$. Let $\mathcal{N}$ denote the set of pairs $(\boldsymbol{u}, \boldsymbol{u}') \in \mathcal{B}^2$ such that $\boldsymbol{u}$ and $\boldsymbol{u}'$ are neighbors. Further, denote

$$d_{H,\mathcal{N}}(\psi) \triangleq \max_{(\boldsymbol{u},\boldsymbol{u}')\in\mathcal{N}} d_H(\psi(\boldsymbol{u}), \psi(\boldsymbol{u}')).$$

We will use the term *maximum local Hamming distance* of ψ to refer to $d_{H,\mathcal{N}}(\psi)$. We will show next that a small value of $d_{H,\mathcal{N}}(\psi)$ ensures a good performance of γ. To this end we need the following Lemma which provides a crucial insight on the form of each set $\gamma((\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2)$. Its proof is deferred to Appendix A.

Lemma 1. *Let γ be defined as in (4). Then for any $\boldsymbol{\lambda}_1 \in \mathcal{B}_1$, there are a subset C_0 of C, a binary sequence $\boldsymbol{s} \in \mathbb{F}_2^{nR}$ and an injective mapping $\omega : C_0 \to \mathbb{F}_2^{nR}$ such that*

$$\gamma((\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2) = \boldsymbol{s} \oplus \omega(C_0), \tag{5}$$

and, for every $\boldsymbol{c} \in C_0$,

$$d_H(\boldsymbol{c}, \omega(\boldsymbol{c})) \leq d_{H,\mathcal{N}}(\psi). \tag{6}$$

Remark 1. Since the performance of the maximum likelihood decoder over BSC(p) is shift invariant, relation (5) implies that the channel code $\gamma((\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2)$ has the same performance as $\omega(C_0)$. On the other hand, relation (6) implies that when the value of $d_{H,\mathcal{N}}(\psi)$ is small the set $\omega(C_0)$ is "close" in Hamming distance to C_0, and thus, the error correction performance of $\omega(C_0)$ is close to the performance of C_0.

The following result, also proved in Appendix A, derives an upper bound for $P_e(\gamma)$ in terms of $d_{H,\mathcal{N}}(\psi)$ and $P_e(C)$, which is the maximal conditional probability of error of channel code C using the maximum likelihood decoder over BSC(p). Let us introduce first an additional notation. Namely, for any $0 < q < 1/2$, and integers $0 < \kappa < m$, let

$$\epsilon(\kappa, m, q) \triangleq H\left(\frac{\kappa}{m}\right) + \frac{\kappa}{m}(1-q). \tag{7}$$

Theorem 1. *Let γ be defined as in (4) and let $\ell \triangleq d_{H,\mathcal{N}}(\psi)$. Then the following holds*

$$P_e(\gamma) \quad \leq \quad 2^{\epsilon(\ell+1, nR, p)nR} P_e(C). \tag{8}$$

The next result obtains a lower bound on the minimum window Hamming distance of γ in terms of the minimum Hamming distance of C and the maximum local Hamming distance of ψ. Its proof is deferred to Appendix A as well.

Proposition 1. *Let γ be defined as in (4). Then the following relation holds*

$$d_{H,W}(\gamma) \geq d_H(C) - 2d_{H,\mathcal{N}}(\psi).$$

Therefore, according to the above results we would like to design a mapping ψ such that $d_{H,\mathcal{N}}(\psi)$ is small. Such a construction is introduced in the following section based on n one-dimensional Gray mappings.

IV. On the Design of Mapping ψ

Notice that in the scalar case we can use Gray encoding for the mapping ψ, which ensures that the bit sequences assigned to consecutive integers have Hamming distance 1. In this section we will develop a method to construct ψ based on n one-dimensional Gray mappings. For this define, for every $m > 0$, the mapping $b_m : \mathcal{I}_{2^m} \to \mathbb{F}_2^m$ such that $b_m(l) = \beta_m(l)\boldsymbol{D}_m$, for any $l \in \mathcal{I}_{2^m}$, where $\mathcal{I}_t \triangleq \{0, 1, \cdots, t-1\}$, for any $t > 0$, and $\beta_m(l) = (b_{m-1}, \cdots, b_0)$, where $b_i \in \mathbb{F}_2$ and $\sum_{i=0}^{m-1} b_i 2^i = l$. Additionally, $\boldsymbol{D}_m$ is the m-by-m matrix with elements in $\mathbb{F}_2$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

satisfying $\boldsymbol{D}_m(i,j) = 1$ if and only if $i \in \{j-1, j\}$. The following result will be useful in the sequel.

Lemma 2. *For every positive integers m, l and t such that $l, l+t \in \mathcal{I}_{2^m}$, we have $d_H(b_m(l), b_m(l+t)) \leq t$.*

Proof: Using the triangle inequality repeatedly, one obtains that

$$d_H(b_m(l), b_m(l+t)) \leq \sum_{s=1}^{t} d_H(b_m(l+s-1), b_m(l+s)).$$

Using further the fact that $d_H(b_m(\kappa), b_m(\kappa+1)) = 1$ for all κ [13], the conclusion follows. ∎

A. General Method Based on n One-dimensional Gray Mappings

Let us assume now that we have a collection $\tilde{\mathcal{Y}}$ of n sets $\tilde{\mathcal{Y}} = \{\mathcal{Y}_1, \mathcal{Y}_2, \cdots, \mathcal{Y}_n\}$, with $\mathcal{Y}_i \subseteq \Lambda$, for $i = 1, \cdots, n$, such that any coarse lattice point $\boldsymbol{u} \in \mathcal{B}$ can be expressed uniquely as

$$\boldsymbol{u} = \sum_{i=1}^{n} \boldsymbol{u}_i, \qquad \boldsymbol{u}_i \in \mathcal{Y}_i, 1 \leq i \leq n. \tag{9}$$

Then for each $i, 1 \leq i \leq n$, construct a mapping $\theta_i : \mathcal{Y}_i \longrightarrow \{0, 1, \cdots, \kappa_i - 1\}$, where $\kappa_i \triangleq |\mathcal{Y}_i|$. Finally, for any coarse lattice point $\boldsymbol{u} \in \mathcal{B}$ with the decomposition given in (9), define $\psi(\boldsymbol{u})$ as

$$\psi(\boldsymbol{u}) = (\boldsymbol{0}_k, b_{m_1}(\theta_1(\boldsymbol{u}_1)), b_{m_2}(\theta_2(\boldsymbol{u}_2)), \cdots, b_{m_n}(\theta_n(\boldsymbol{u}_n))), \tag{10}$$

where $m_i \triangleq \lceil \log_2 \kappa_i \rceil$, for all $1 \leq i \leq n$. Note that in this case, we have

$$R = \frac{1}{n}\left(k + \sum_{i=1}^{n} m_i\right).$$

Further, notice that if C is a systematic code with the systematic bits appearing in the first k positions, then the image of ψ, denoted by $Im(\psi)$, is a set of coset representatives of C.

Clearly, the performance of the mapping ψ depends on the particular choice of the collection of sets $\tilde{\mathcal{Y}}$. In order to take advantage of Lemma 2 to decrease the value of $d_{H,\mathcal{N}}(\psi)$ we would like each function θ_i to map neighbouring points to consecutive integers. Therefore, it is desirable to choose $\mathcal{Y}_i$ in such a way that such a mapping exists. Additionally, we would like to have the quantity $\sum_{i=1}^{n} m_i - \lceil \log_2 \mathcal{B} \rceil$ as small as possible in order to minimize the extra redundancy introduced into the system by the structure imposed on ψ. In the sequel we present a method for choosing $\tilde{\mathcal{Y}}$, which relies on the decomposition of lattice points as a linear combination of the rows of the generator matrix. The method is termed "GM-decomposition", where "GM" stands for "Generator Matrix".

B. GM-decomposition Method

Notice that each lattice point can be uniquely expressed as an integer linear combination of the row vectors of the generator matrix. Let $\boldsymbol{G}$ be an n-by-n generator matrix of the coarse lattice Λ, and let $\boldsymbol{g}_i$ denote its ith row, for $i = 1, \cdots, n$. For each $i, 1 \leq i \leq n$, let

$$\mathcal{Y}_i \triangleq \{z\boldsymbol{g}_i | z \in \mathbb{Z}, a_{i,0} \leq z \leq a_{i,1}\},$$

for some integers $a_{i,0} \leq a_{i,1}$, such that $\mathcal{B} \subseteq \mathcal{Y}_1 + \mathcal{Y}_2 + \cdots + \mathcal{Y}_n$. Clearly, the decomposition (9) is unique. We will use the term "GM-decomposition" (GMD, for short) to refer to the decomposition using this choice of $\tilde{\mathcal{Y}}$. Further, define $\theta_i(z\boldsymbol{g}_i) \triangleq z - a_{i,0}$, for all $1 \leq i \leq n$, $z\boldsymbol{g}_i \in \mathcal{Y}_i$, $1 \leq i \leq n$. We will use the notation ψ_{GM} for the mapping ψ designed based on GMD. Then for $\boldsymbol{u} = \boldsymbol{z}\boldsymbol{G}$, where $\boldsymbol{z} = (z_1, \cdots, z_n)$, relation (10) becomes

$$\psi_{GM}(\boldsymbol{u}) = (\boldsymbol{0}_k, b_{m_1}(z_1 - a_{1,0}), b_{m_2}(z_2 - a_{2,0}), \cdots \\ \cdots, b_{m_n}(z_n - a_{n,0})). \tag{11}$$

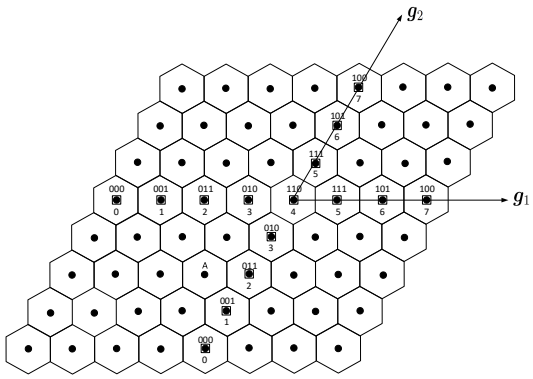

Fig. 3: Illustration of the GMD method.

Example 1. Fig. 3 shows an example for the construction of $\tilde{\mathcal{Y}}$ using the GMD method for the lattice A_2. The points contained in $\mathcal{Y}_1 \cup \mathcal{Y}_2$, are marked with a square. The integer assigned to each point in $\mathcal{Y}_i$ by θ_i is written below the point. The 3-bit binary sequence assigned to each point in $\mathcal{Y}_i$ using the mapping $b_3(\theta_i(\cdot))$ is written above the point. For $\boldsymbol{u} = -\boldsymbol{g}_1 - 2\boldsymbol{g}_2$ (denoted by A in Fig. 3) we have $\psi(\boldsymbol{u}) = (\boldsymbol{0}_k, 010, 011)$.

Next we will derive an upper bound for $d_{H,\mathcal{N}}(\psi_{GM})$. Let ρ denote the covering radius of the coarse lattice, which is defined as the supremum of the Euclidian distance from the origin to any point $\boldsymbol{x} \in V[\boldsymbol{0} : \Lambda]$ [22].

Proposition 2. *Let $\boldsymbol{G}$ be the generator matrix of Λ, used to construct ψ_{GM}, let $\boldsymbol{G}_0 \triangleq \frac{1}{\rho}\boldsymbol{G}$ and let $c = \|\boldsymbol{G}_0^{-1}\| = \rho\|\boldsymbol{G}^{-1}\|$. Then one has $d_{H,\mathcal{N}}(\psi_{GM}) \leq 3c\sqrt{n}$.*

Remark 2. Notice that the constant c defined in the above proposition is invariant to the scaling of the lattice Λ. As we will see in the next section, this result is essential, along with Theorem 1, in establishing that the error correction performance of γ approaches the performance of the code C, as the rate R approaches infinity.

The proof of Proposition 2 relies on the following lemma.

Lemma 3. *Let $\boldsymbol{u}, \boldsymbol{u}' \in \Lambda$ be neighbors. Then $\|\boldsymbol{u} - \boldsymbol{u}'\| \leq 3\rho$.*

Proof: Let $\boldsymbol{\lambda}_1, \boldsymbol{\lambda} \in \Lambda_s$, be such that $\boldsymbol{\lambda} \in \boldsymbol{\lambda}_1 + \mathcal{W}'$, $Q_\Lambda(\boldsymbol{\lambda}) = \boldsymbol{u}$ and $Q_\Lambda(\boldsymbol{\lambda}_1) = \boldsymbol{u}'$. Then

$$\|\boldsymbol{\lambda} - \boldsymbol{u}\| \leq \rho, \quad \|\boldsymbol{\lambda} - \boldsymbol{\lambda}_1\| \leq \rho, \quad \|\boldsymbol{\lambda}_1 - \boldsymbol{u}'\| \leq \rho.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

Using the triangle inequality repeatedly, one obtains

$$\|\boldsymbol{u}-\boldsymbol{u}'\| \leq \|\boldsymbol{u}-\boldsymbol{\lambda}\| + \|\boldsymbol{\lambda}-\boldsymbol{\lambda}_1\| + \|\boldsymbol{\lambda}_1-\boldsymbol{u}'\| \leq 3\rho.$$

Thus, the proof is completed. ∎

Proof of Proposition 2: Let $\boldsymbol{z}, \boldsymbol{z}' \in \mathbb{Z}^n$ such that $\boldsymbol{u}$ and $\boldsymbol{u}'$ are neighbors, where $\boldsymbol{u} = \boldsymbol{z}\boldsymbol{G}$ and $\boldsymbol{u}' = \boldsymbol{z}'\boldsymbol{G}$. According to Lemma 3, one has

$$\|\boldsymbol{u}-\boldsymbol{u}'\| \leq 3\rho. \tag{12}$$

Since $\boldsymbol{G}$ is invertible, one obtains that $\boldsymbol{z} = \boldsymbol{u}\boldsymbol{G}^{-1}$, and $\boldsymbol{z}' = \boldsymbol{u}'\boldsymbol{G}^{-1}$, which leads further to

$$\begin{aligned}\|\boldsymbol{z}-\boldsymbol{z}'\| &= \|(\boldsymbol{u}-\boldsymbol{u}')\boldsymbol{G}^{-1}\| \leq \|\boldsymbol{u}-\boldsymbol{u}'\|\|\boldsymbol{G}^{-1}\| \\ &\leq 3\rho\frac{1}{\rho}\|\boldsymbol{G}_0^{-1}\| = 3c,\end{aligned} \tag{13}$$

where the last inequality follows from (12) and the fact that $\|\boldsymbol{G}^{-1}\| = \frac{1}{\rho}\|\boldsymbol{G}_0^{-1}\|$. By applying the Cauchy-Schwarz inequality and relation (13), one obtains that

$$\sum_{i=1}^{n}|z_i - z_i'| \leq \sqrt{\|\boldsymbol{z}-\boldsymbol{z}'\|^2 n} \leq 3c\sqrt{n}.$$

The above relations in conjunction with (11) and Lemma 2 imply that

$$\begin{aligned}d_H(\psi_{GM}(\boldsymbol{u}), \psi_{GM}(\boldsymbol{u}')) &= \sum_{i=1}^{n} d_H(b_{m_i}(z_i - a_{i,0}), b_{m_i}(z_i' - a_{i,0})) \\ &\leq \sum_{i=1}^{n}|z_i - z_i'| \leq 3c\sqrt{n}.\end{aligned}$$

Now the proof is completed. ∎

V. Asymptotic Analysis

In this section we evaluate the performance of the proposed construction, asymptotically as R goes to ∞, i.e., as $v(\Lambda_s)$ and $v(\Lambda_c)$ go to 0. We would like to point out that the asymptotic analysis of performance under the above high resolution assumption is common practice in the MDLVQ literature [14]-[18], [20]. To this end we will consider a family of MDLVQs as explained next.

First we would like to consider a measure of the rate of the MDLVQ, which does not include the additional redundancy due to the structure imposed on γ. Denote it by R_0 and define it as

$$R_0 \triangleq \frac{1}{n}\log_2\frac{vol(\mathcal{A})}{v(\Lambda_s)}.$$

The motivation for this definition is that the minimum rate needed to encode description i with a fixed-rate coder, is $\frac{1}{n}\lceil\log_2 \mathcal{B}_i\rceil$, value which approaches R_0 as $v(\Lambda_s) \to 0$. Clearly, the interesting case is when $vol(\mathcal{A}) > v(\Lambda_s)$, and thus $R_0 > 0$. Next we introduce the *co-redundancy parameter* a of the the MDLVQ, defined as follows,

$$a \triangleq \frac{\log_2\frac{v(\Lambda_s)}{v(\Lambda_c)}}{\log_2\frac{vol(\mathcal{A})}{v(\Lambda_s)}}. \tag{14}$$

Notice that a can be equivalently expressed as

$$a = \frac{\log_2 N}{nR_0}.$$

Then the minimum rate to encode the points in the central codebook $\mathcal{B}_c$ using a fixed-rate encoder is $\frac{1}{n}\lceil\log_2 \mathcal{B}_c\rceil$, quantity which approaches $R_0(1+a)$ as $v(\Lambda_s) \to 0$. Thus, since $1 - a = \frac{2R_0 - R_0(1+a)}{R_0}$, the quantity $1-a$ can be regarded as a measure of the redundancy between the two descriptions[6]. This observation motivates the use of the term "co-redundancy parameter" for a. The above discussion also implies that $0 < a < 1$. Additionally, it can be easily verified that

$$v(\Lambda_s) = v(\Lambda_c)^{\frac{1}{1+a}} vol(\mathcal{A})^{\frac{a}{1+a}}, \quad N = \left(\frac{vol(\mathcal{A})}{v(\Lambda_c)}\right)^{\frac{a}{1+a}}.$$

For fixed lattices $\Lambda_c, \Lambda_s, \Lambda$ and parameter a satisfying (14), we will consider a family of MDLVQs $\mathcal{FM}(\Lambda_c, \Lambda_s, \Lambda, a) = \{\mathcal{M}(\sigma) | 0 < \sigma \leq 1\}$, where $\mathcal{M}(\sigma)$ is the MDLVQ whose central, side and coarse lattice are, respectively,

$$\Lambda_c(\sigma) \triangleq \sigma\Lambda_c, \Lambda_s(\sigma) \triangleq \sigma^{\frac{1}{1+a}}\Lambda_s, \Lambda(\sigma) \triangleq \mu(\sigma)\Lambda,$$

where $\mu(\sigma)$ will be defined shortly. Then all the MDLVQs in the family have the co-redundancy parameter equal to a. Indeed, one has

$$\begin{aligned}\frac{\log_2\frac{v(\Lambda_s(\sigma))}{v(\Lambda_c(\sigma))}}{\log_2\frac{vol(\mathcal{A})}{v(\Lambda_s(\sigma))}} &= \frac{\log_2\frac{\sigma^{\frac{n}{1+a}}v(\Lambda_s)}{\sigma^n v(\Lambda_c)}}{\log_2\frac{vol(\mathcal{A})}{\sigma^{\frac{n}{1+a}}v(\Lambda_s)}} \\ &= \frac{\log_2\frac{v(\Lambda_s)}{v(\Lambda_c)} - \frac{na}{1+a}\log_2\sigma}{\log_2\frac{vol(\mathcal{A})}{v(\Lambda_s)} - \frac{n}{1+a}\log_2\sigma} \\ &= \frac{a\left(\log_2\frac{vol(\mathcal{A})}{v(\Lambda_s)} - \frac{n}{1+a}\log_2\sigma\right)}{\log_2\frac{vol(\mathcal{A})}{v(\Lambda_s)} - \frac{n}{1+a}\log_2\sigma} \\ &= a,\end{aligned}$$

where the second last equality uses the fact that $\log_2\frac{v(\Lambda_s)}{v(\Lambda_c)} = a\log_2\frac{vol(\mathcal{A})}{v(\Lambda_s)}$. Further, it can be easily seen that as σ approaches 0, the fundamental volume of $\Lambda_c(\sigma)$ and of $\Lambda_s(\sigma)$ approach 0 while $N(\sigma)$, defined as $N(\sigma) \triangleq \frac{v(\Lambda_s(\sigma))}{v(\Lambda_c(\sigma))}$, and $R_0(\sigma)$, defined as $R_0(\sigma) \triangleq \frac{1}{n}\log_2\frac{vol(\mathcal{A})}{v(\Lambda_s(\sigma))}$, both approach ∞. It is important to emphasize that the family $\mathcal{FM}(\Lambda_c, \Lambda_s, \Lambda, a)$ is defined such that the rate R_0 tends to infinity while the co-redundancy parameter a is fixed. The analysis in this section will be performed on this family as $\sigma \to 0$, condition which is equivalent to $R_0 \to \infty$. The same asymptotic regime was considered in [14], where the rate is approaching infinity while keeping fixed some parameter which is similar in spirit to our parameter a.

We extend all the notations related to an MDLVQ, which were introduced in the previous sections, to $\mathcal{M}(\sigma)$ by adding the parameter σ in the notation. Thus, the counterparts of the sets $\mathcal{B}_c$, $\mathcal{B}_i$ for $i = 1, 2$, $\mathcal{B}$, $\mathcal{W}$, $\mathcal{W}'$, C will be denoted, respectively, by $\mathcal{B}_c(\sigma)$, $\mathcal{B}_i(\sigma)$ for $i = 1, 2$, $\mathcal{B}(\sigma)$, $\mathcal{W}(\sigma)$, $\mathcal{W}'(\sigma)$, $C(\sigma)$. The counterparts of the mappings α, γ, φ and ψ will be denoted, respectively, by $\alpha(\sigma)$, $\gamma(\sigma)$, $\varphi(\sigma)$ and $\psi(\sigma)$. Finally, the counterparts of k, R, κ_i, m_i, for $1 \leq i \leq n$,

[6]Note that for fixed dimension n, the value of $1-a$ does not account for the whole redundancy existing in the system since fixed-rate encoding is suboptimal. However, as we will see shortly, as R and n approach ∞, the value of $1-a$ approaches the true redundancy rate, for memoryless vector sources.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

will be denoted, respectively, by $k(\sigma), R(\sigma)$, $\kappa_i(\sigma), m_i(\sigma)$, for $1 \leq i \leq n$. Furthermore, we assume that the IA $\alpha(\sigma)$, is designed as in [14], while the mapping $\gamma(\sigma)$ is constructed as described in Sections III and IV, with $\psi(\sigma)$ based on GMD.

Additionally, we assume that the coarse lattice Λ satisfies the condition $V[\mathbf{0} : \Lambda] \supseteq \mathcal{B}(\mathcal{W})$, where, for any set $\mathcal{S} \in \mathbb{R}^n$, $\mathcal{B}(\mathcal{S})$ denotes the smallest ball centered at the origin, which contains S. Further, we define

$$\mu(\sigma) \triangleq \left(\frac{vol(\mathcal{B}(\mathcal{W}(\sigma)))}{vol(\mathcal{B}(\mathcal{W}))}\right)^{1/n}. \tag{15}$$

Recall that $\Lambda(\sigma) = \mu(\sigma)\Lambda$. It follows that the condition $V[\mathbf{0} : \Lambda(\sigma)] \supseteq \mathcal{B}(\mathcal{W}(\sigma))$ is satisfied, implying that $V[\mathbf{0} : \Lambda(\sigma)] \supseteq W(\sigma)$ holds, too. Additionally, notice that the construction of the IA in [14] ensures that $vol(\mathcal{B}(\mathcal{W}(\sigma)) = N(\sigma)v(\Lambda_s(\sigma))(1 + o(1)) = \sigma^{n\frac{1-a}{1+a}} Nv(\Lambda_s)(1 + o(1))$ as $\sigma \to 0$. This implies that $vol(\mathcal{B}(\mathcal{W}(\sigma)) \to 0$ and, further, that $\mu(\sigma) \to 0$ as $\sigma \to 0$.

The following result, whose proof is given in Appendix B, evaluates the rate loss due to the structure of the mapping γ, asymptotically as $\sigma \to 0$.

Theorem 2. *Consider a family of MDLVQs $\mathcal{FM}(\Lambda_c, \Lambda_s, \Lambda, a) = \{\mathcal{M}(\sigma)|0 < \sigma \leq 1\}$ for some $0 < a < 1$. Then the following relations hold*

$$\frac{1}{n}\log_2\frac{vol(\mathcal{P}_0)}{vol(\mathcal{A})} \leq \lim_{\sigma\to 0}(R(\sigma) - R_0(\sigma)) \leq \frac{1}{n}\log_2\frac{vol(\mathcal{P}_0)}{vol(\mathcal{A})} + 1 + \frac{1}{n}, \tag{16}$$

where $\mathcal{P}_0$ denotes the smallest parallelepiped with edges parallel to the basis vectors of the lattice Λ, which contains the set $\mathcal{A}$.

Next we will evaluate the asymptotic error correction performance of the mapping $\gamma(\sigma)$. For this we need another lemma. Its proof is deferred to Appendix B.

Lemma 4. *Consider a family of MDLVQs $\mathcal{FM}(\Lambda_c, \Lambda_s, \Lambda, a) = \{\mathcal{M}(\sigma)|0 < \sigma \leq 1\}$ for some $0 < a < 1$. Then the following relation is valid*

$$\lim_{\sigma\to 0}\frac{k(\sigma)}{nR(\sigma)} = a.$$

Theorem 3. *Consider a family of MDLVQs $\mathcal{FM}(\Lambda_c, \Lambda_s, \Lambda, a) = \{\mathcal{M}(\sigma)|0 < \sigma \leq 1\}$ for some $0 < a < 1$. Assume that channel 2 is $BSC(p)$.*

1) *If $a < 1 - H(p)$ then*

$$\lim_{\sigma\to 0} P_e(\gamma(\sigma)) = 0, \tag{17}$$

for a suitable choice of the channel code $C(\sigma)$, for $0 < \sigma \leq 1$.

2) *Conversely, if there are channel codes $C(\sigma)$, for $0 < \sigma \leq 1$, such that (17) holds then one must have $a \leq 1 - H(p)$.*

In this work we use a fixed-rate coder for each description, while the prior work uses entropy-coded MDLVQ. Therefore, it is interesting to assess the rate difference between the fixed-rate and entropy-coded scenarios. In the following development we use different vector dimensions n, and will consider a family of MDLVQs $\mathcal{FM}^{(n)}(\Lambda_c^{(n)}, \Lambda_s^{(n)}, \Lambda^{(n)}, a)$ for each n. In order to specify the dimension n, we add the superscript $^{(n)}$ to the related notations. For instance, notations $\mathcal{A}^{(n)}$, $\mathcal{M}^{(n)}(\sigma)$, $R^{(n)}(\sigma)$ will be used, respectively, for $\mathcal{A}$, $\mathcal{M}(\sigma)$, $R(\sigma)$.

Assume that the input random vector $\boldsymbol{X}$ is memoryless with marginal probability density function (pdf) $f(x)$, i.e. $\boldsymbol{f}(\boldsymbol{x}) = f(x_1)\cdots f(x_n)$, where $\boldsymbol{x} = (x_1, \cdots, x_n)$. Also assume that the pdf $f(x)$ has finite mean $\mu \triangleq \int_{\mathbb{R}} xf(x)dx$, finite variance $\sigma^2 \triangleq \int_{\mathbb{R}}(x-\mu)^2 f(x)dx$, and finite differential entropy $h(f) \triangleq -\int_{\mathbb{R}} f(x)\log_2 f(x)dx$. Denote $\boldsymbol{\mu} = (\mu, \cdots, \mu)$.

We will consider the squared error as a distortion measure and denote by $d_i^{(n)}(\sigma)$, for $i = 0, 1, 2$, the distortions (per sample) obtained by applying the fixed-rate MDLVQ $\mathcal{M}^{(n)}(\sigma)$[7]. Specifically, $d_0^{(n)}(\sigma)$ is the distortion when both descriptions are decoded, while for $i = 1, 2$, $d_i^{(n)}(\sigma)$ is the distortion when only description i is decoded. Let $d_{i,e}^{(n)}(\sigma)$, for $i = 0, 1, 2$, denote the corresponding distortions for the entropy-coded MDLVQ. Denote by $R_e^{(n)}(\sigma)$ the rate of a description in the entropy-coded case and let $R_{e,c}^{(n)}(\sigma) = \frac{1}{n}H(Q_{\Lambda_c}(\boldsymbol{X}))$. Further, define the redundancy rate of the MDLVQ $\mathcal{M}^{(n)}(\sigma)$ as

$$\rho^{(n)}(\sigma) \triangleq \frac{2R^{(n)}(\sigma) - R_{e,c}^{(n)}(\sigma)}{R^{(n)}(\sigma)}. \tag{18}$$

The following result is proved in Appendix B.

Proposition 3. *Assume that the random variables $X_1, X_2, \cdots, X_n$ are drawn i.i.d. according to the pdf $f(x)$, i.e., $\boldsymbol{f}(\boldsymbol{x}) = f(x_1)\cdots f(x_n)$. For each $\epsilon > 0$ and $n \geq 1$, denote*

$$\mathcal{A}_\epsilon^{(n)} \triangleq \left\{\boldsymbol{x} \in \mathbb{R}^n \middle| \left|-\frac{1}{n}\log_2 \boldsymbol{f}(\boldsymbol{x}) - h(f)\right| \leq \epsilon\right\}.$$

Consider a sequence[8] ϵ_n of positive values such that $\lim_{n\to\infty}\epsilon_n = 0$ and

$$\lim_{n\to\infty}\mathbb{P}\left[\boldsymbol{X} \in \mathcal{A}_{\epsilon_n}^{(n)}\right] = 1. \tag{19}$$

Fix an arbitrary $a, 0 < a < 1$. For each $n \geq 1$ consider a family of MDLVQs $\mathcal{FM}^{(n)}(\Lambda_c^{(n)}, \Lambda_s^{(n)}, \Lambda^{(n)}, a)$ with $\mathcal{A}^{(n)} = \mathcal{A}_{\epsilon_n}^{(n)}$. Then the following assertions are true.

i) $\lim_{n\to\infty}\lim_{\sigma\to 0}(R_0^{(n)}(\sigma) - R_e^{(n)}(\sigma)) = 0.$

ii) *For every n there is δ_n such that $\lim_{n\to\infty}\delta_n = 0$ and*

$$d_i^{(n)}(\sigma) \leq d_{i,e}^{(n)}(\sigma) + \delta_n, \text{ for all } i = 0, 1, 2, \sigma > 0.$$

iii) $\lim_{n\to\infty}\lim_{\sigma\to 0}(\rho^{(n)}(\sigma) - (1 - a)) = 0.$

Remark 3. Proposition 3 implies that by encoding only the vectors from the set $\mathcal{A}_{\epsilon_n}^{(n)}$ using a fixed-rate MDLVQ, the rates achieved are equal, asymptotically as $\sigma \to 0$ and $n \to \infty$, to those obtained using the entropy-coded MDLVQ, while the distortions for the case of on/off channels are not worse. On the other hand, combining Proposition 3 with Theorem 2 one

[7]When the vector sequence $\boldsymbol{x}$ is not in $\mathcal{A}^{(n)}$, a special index is transmitted, which is always decoded as the mean $\boldsymbol{\mu}$.

[8]The existence of a sequence ϵ_n of positive values such that $\lim_{n\to\infty}\epsilon_n = 0$ and relation (19) holds follows from the Weak Law of Large Numbers.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

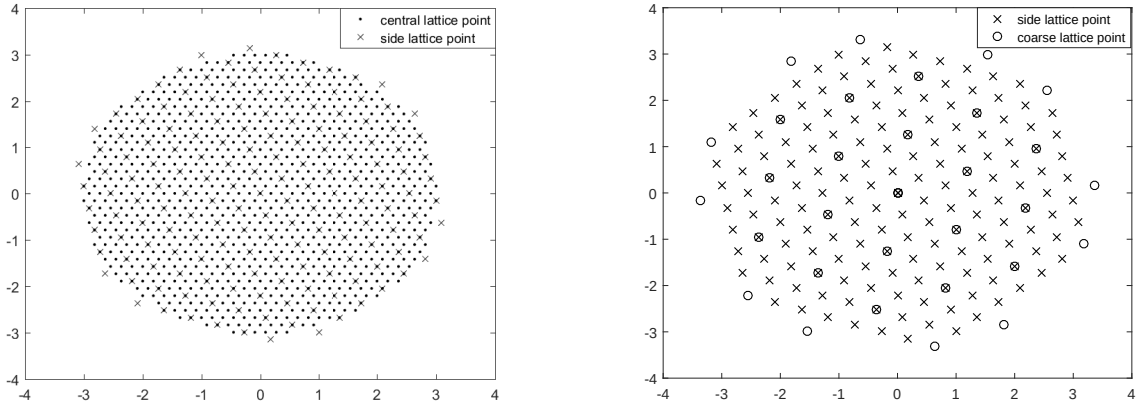

Fig. 4: The sets $\mathcal{B}_c$, $\mathcal{B}_1$, $\mathcal{B}_2$ and $\mathcal{B}$ for Example 2.

obtains that

$$\lim_{n\to\infty}\frac{1}{n}\log_2\frac{vol(\mathcal{P}_0^{(n)})}{vol(\mathcal{A}_{\epsilon_n}^{(n)})} \le \lim_{n\to\infty}\lim_{\sigma\to 0}(R^{(n)}(\sigma)-R_e^{(n)}(\sigma)) \le 1+\lim_{n\to\infty}\frac{1}{n}\log_2\frac{vol(\mathcal{P}_0^{(n)})}{vol(\mathcal{A}_{\epsilon_n}^{(n)})}.$$

Remark 4. Note that the rate of the channel code $C^{(n)}(\sigma)$ is $\frac{k^{(n)}(\sigma)}{nR^{(n)}(\sigma)}$. Lemma 4 implies that the rate of $C^{(n)}(\sigma)$ approaches the value of a as $\sigma\to 0$. If we define the redundancy rate of $C^{(n)}(\sigma)$ as 1 minus its rate, it follows that its redundancy rate approaches $1-a$. On the other hand, according to Proposition 3 the redundancy rate of the MDLVQ can also be approximated by $1-a$ as $\sigma\to 0$ and $n\to\infty$. We may say that a capacity achieving channel code exploits its whole redundancy for error correction. According to Theorem 3, the mapping $\gamma^{(n)}(\sigma)$ has the same error correction performance as the channel code $C^{(n)}(\sigma)$ as $\sigma\to 0$. Therefore, we may say that, by choosing a capacity achieving channel code for $C^{(n)}(\sigma)$, the whole redundancy built in the MDLVQ is used for error correction.

VI. Experimental Results

This section assesses the practical performance of the proposed structured mapping γ in comparison with a random mapping. The dimension of the MDLVQ is 2 and all lattices are scaled versions of A_2. The set $\mathcal{A}$ is the ball of radius 3 centered in the origin. The vector source is a memoryless Gaussian source with zero mean and variance 1, truncated to $\mathcal{A}$. The performance of each mapping is measured using the channel distortion at the central decoder, $D_{0,c}$, i.e., the expected squared distance between the reconstruction at the central decoder when the second channel is $BSC(p)$, and the reconstruction when both descriptions are correct. More specifically, to illustrate the performance we plot the value of $SNR=-10\log_{10}D_{0,c}$ for various values of the channel error probability p. The mapping ψ used in the construction of the structured mapping γ is based on GMD. The mapping used for comparison is selected as the best out of ten randomly generated mappings. We consider two MDLVQ examples highlighted next. Their results are illustrated in Fig. 6.

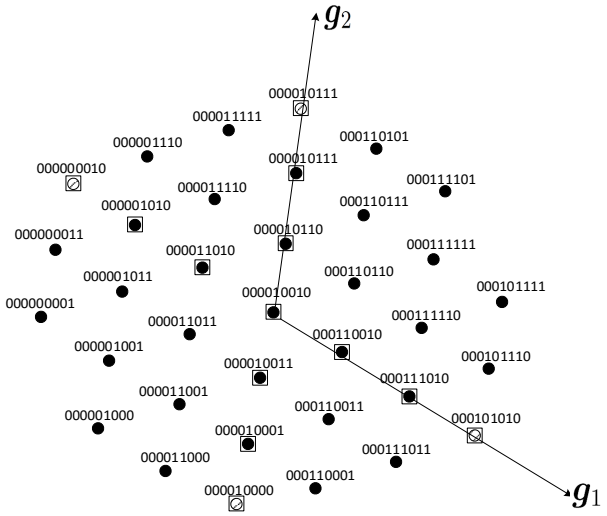

Fig. 5: Mapping ψ_{GM} for Example 2.

Example 2. The settings of this example are $|\mathcal{B}_c|=1015$, $|\mathcal{B}_1|=|\mathcal{B}_2|=163$, $|\mathcal{B}|=31$, $N=|\mathcal{W}'|=7$, $k=3$ and $R=4.5$. The sets $\mathcal{B}_c$, $\mathcal{B}_2$ and $\mathcal{B}$ are shown in Fig. 4. The linear channel code C used in the design of the proposed mapping γ is generated by the following matrix

$$\begin{bmatrix}1&0&0&1&0&1&1&0&0\\0&1&0&1&1&0&0&1&0\\0&0&1&0&1&1&0&0&1\end{bmatrix},$$

thus $d_H(C)=4$. The mapping ψ_{GM} is shown in Fig. 5. It can be easily seen that $d_{H,W}(\gamma)\ge 2$. Note that $a\approx\frac{\log_2 N}{\log_2\mathcal{B}_2}=0.382$, while the rate of the channel code C is $\frac{k}{nR}=0.33$.

Example 3. In this case we have $|\mathcal{B}_c|=6307$, $|\mathcal{B}_1|=|\mathcal{B}_2|=961$, $|\mathcal{B}|=157$, $N=|\mathcal{W}'|=7$, $k=3$ and $R=5.5$. The generator matrix of the linear channel code C is

$$\begin{bmatrix}1&0&0&1&0&1&1&0&1&1&1\\0&1&0&1&1&0&1&1&0&1&1\\0&0&1&0&1&1&0&1&1&1&1\end{bmatrix}.$$

Therefore, we have $d_H(C)=5$, and further $d_{H,W}(\gamma)\ge 3$, which implies that the decoder can correct all one bit error

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

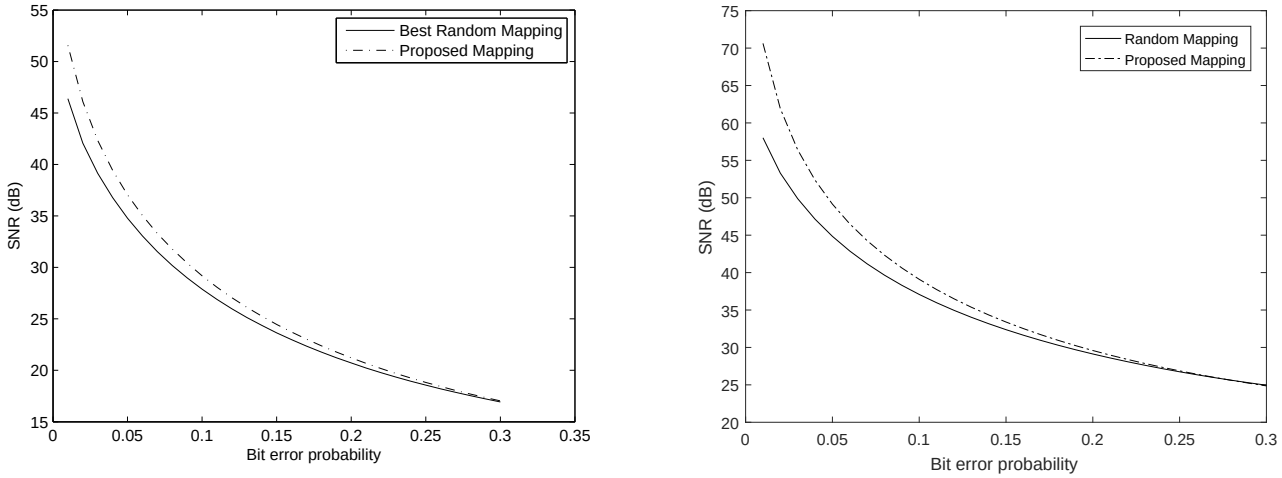

Fig. 6: Performance of the proposed mapping when only one channel is noisy; left: Example 2; right: Example 3.

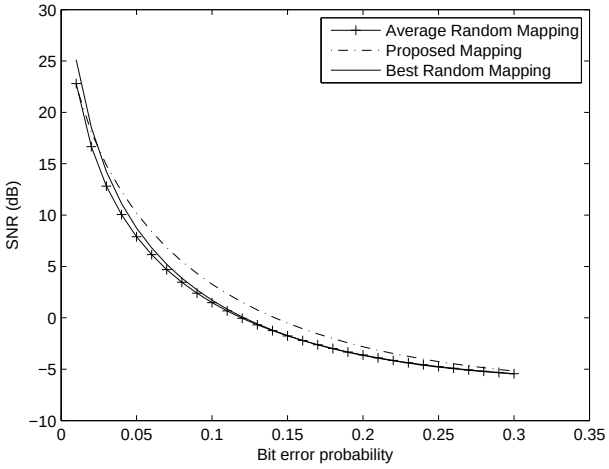

Fig. 7: Performance of the proposed mapping when both channels are noisy.

patterns. The rate of the channel code C is $\frac{k}{nR} = 0.273$, while $a \approx \frac{\log_2 N}{\log_2 \mathcal{B}_2} = 0.283$.

As it can be seen from Fig. 6 for both examples the proposed structured mapping outperforms significantly the best out of ten random mappings.

While the proposed mapping was not designed for the scenario when both channels are noisy, it is interesting to evaluate its performance in such a case. In Fig. 7 we plot the SNR values obtained for Example 2 when both channels are noisy. We consider the same random mappings that were tested in the one noisy channel case. Each time we use the same mapping for both descriptions. We observe that the proposed method is always better than or very close to the average performance of the random mappings. Additionally, we plot the performance of the mapping which was the best in the one noisy channel scenario, which is shown to be inferior to the proposed mapping for p higher than 0.02.

VII. Conclusion

This work is concerned with improving the bit-error resilience of multiple description lattice vector quantizers (MDLVQ) when one of the descriptions is received correctly while the other description may carry bit errors. To this end we propose the design of the mapping γ of side lattice points to binary indexes such that the set of all binary sequences corresponding to possible lattice points of the description carrying errors when the error-free description is fixed, to be a good channel code. Our construction is based on partitioning the set of side lattice points using Voronoi regions of a suitable coarse lattice and mapping the side lattice points in each Voronoi region into a coset of a good linear channel code. Moreover, our design ensures that neighbouring Voronoi regions are mapped to cosets close in Hamming distance. Further, we prove that the error correction performance of the proposed mapping γ approaches the performance of the linear channel code used in its construction, asymptotically as the rate of the MDLVQ goes to ∞. Finally, experimental results show significant improvements in practice by using the proposed mapping compared with a random mapping.

Appendix A
Proof of Results in Section III

Proof of Lemma 1: Fix $\boldsymbol{\lambda}_1 \in \mathcal{B}_1$. Define the mapping $h : (\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2 \to \mathbb{F}_2^{nR}$ such that $h(\boldsymbol{\lambda}) = \varphi(\boldsymbol{\lambda} \bmod \Lambda)$ for every $\boldsymbol{\lambda} \in (\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2$. First we will show that h is injective. Since φ is injective, it is sufficient to prove that the "mod Λ" operation is injective on $\boldsymbol{\lambda}_1 + \mathcal{W}$. For this take two arbitrary distinct points $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ in $\boldsymbol{\lambda}_1 + \mathcal{W}$. It follows that $\boldsymbol{\lambda} - \boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}' - \boldsymbol{\lambda}_1$ are two distinct points in $V[\mathbf{0} : \Lambda]$, and, according to [25, Equation (2.29)], their difference is not in Λ. This implies that $\boldsymbol{\lambda} - \boldsymbol{\lambda}' \notin \Lambda$, hence $\boldsymbol{\lambda} \bmod \Lambda \neq \boldsymbol{\lambda}' \bmod \Lambda$.

Now let C_0 be the image of h. Then clearly, $C_0 \subseteq C$. Since h is a bijection from $(\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2$ to C_0, it follows that it has an inverse $h^{-1} : C_0 \to (\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2$. Then let $\boldsymbol{s} \triangleq \psi(Q_\Lambda(\boldsymbol{\lambda}_1))$ and define ω as follows

$$\omega(\boldsymbol{c}) \triangleq \gamma(h^{-1}(\boldsymbol{c})) \oplus \boldsymbol{s}, \text{ for every } \boldsymbol{c} \in C_0. \tag{20}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

Since γ and h^{-1} are injective, it follows that ω is injective, too. Additionally, relation (20) readily implies (5).

Now we are left to prove relation (6). Let $\boldsymbol{c} \in C_0$ and $\boldsymbol{\lambda} = h^{-1}(\boldsymbol{c})$. Then $\gamma(h^{-1}(\boldsymbol{c})) = \boldsymbol{c} \oplus \psi(Q_\Lambda(\boldsymbol{\lambda}))$. It follows that $d_H(\boldsymbol{c}, \omega(\boldsymbol{c})) = d_H(\psi(Q_\Lambda(\boldsymbol{\lambda})), \psi(Q_\Lambda(\boldsymbol{\lambda}_1)))$. The fact that $\boldsymbol{\lambda} \in \boldsymbol{\lambda}_1 + \mathcal{W}$ leads to $(Q_\Lambda(\boldsymbol{\lambda}), Q_\Lambda(\boldsymbol{\lambda}_1)) \in \mathcal{N}$ and relation (6) follows. ∎

Proof of Theorem 1: Let $m = nR$. For any channel code $A \subseteq \mathbb{F}_2^m$ and each $\boldsymbol{c} \in A$, let us denote by $\mathcal{F}(\boldsymbol{c}, A)$ the set of m-bit sequences $\boldsymbol{t}$ which are mapped to $\boldsymbol{c}$ by the maximum likelihood decoder. Since C is a vector subspace of $\mathbb{F}_2^m$, one has

$$\mathcal{F}(\boldsymbol{c}, C) = \boldsymbol{c} \oplus \mathcal{F}(\boldsymbol{0}, C),$$

for any $\boldsymbol{c} \in C$, which further implies that $P_e(C) = \mathbb{P}[\boldsymbol{e} \in \overline{\mathcal{F}(\boldsymbol{0}, C)}]$, where $\boldsymbol{e}$ denotes the m-bit error sequence. Before proceeding we need one more notation. For any $B \subseteq \mathbb{F}_2^m$ and any integer $\kappa \leq m$, let $\Gamma^\kappa B$ denote the *Hamming κ-neighbourhood* of B defined as follows [28]

$$\Gamma^\kappa B \triangleq \{\boldsymbol{t} \in \mathbb{F}_2^m | d_H(\boldsymbol{t}, B) \leq \kappa\}.$$

Now fix an arbitrary $\boldsymbol{\lambda}_1 \in \mathcal{B}_1$ and let $A = \gamma((\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2)$. We will first show that the following holds for every $\boldsymbol{c}'$ in A:

$$\boldsymbol{e} \notin \Gamma^{\ell+1}\overline{\mathcal{F}(\boldsymbol{0}, C)} \Rightarrow \boldsymbol{c}' \oplus \boldsymbol{e} \in \mathcal{F}(\boldsymbol{c}', A). \tag{21}$$

The above assertion implies that $P_e(\gamma) \leq \mathbb{P}[\boldsymbol{e} \in \Gamma^{\ell+1}\overline{\mathcal{F}(\boldsymbol{0}, C)}]$. Further, according to [28, Proof of Lemma 5.1], one obtains

$$\mathbb{P}[\boldsymbol{e} \in \Gamma^{\ell+1}\overline{\mathcal{F}(\boldsymbol{0}, C)}] \leq 2^{\epsilon(\ell+1, m, p)m} \mathbb{P}[\boldsymbol{e} \in \overline{\mathcal{F}(\boldsymbol{0}, C)}],$$

which further leads to (8).

We are left to prove (21). Note that if $\overline{\Gamma^{\ell+1}\overline{\mathcal{F}(\boldsymbol{0}, C)}}$ is empty then (21) holds trivially. Let us assume that $\overline{\Gamma^{\ell+1}\overline{\mathcal{F}(\boldsymbol{0}, C)}}$ is nonempty and let $\boldsymbol{e} \in \overline{\Gamma^{\ell+1}\overline{\mathcal{F}(\boldsymbol{0}, C)}}$. Then $d_H(\boldsymbol{e}, \overline{\mathcal{F}(\boldsymbol{0}, C)}) > \ell + 1$. Moreover, $\boldsymbol{e} \in \mathcal{F}(\boldsymbol{0}, C)$ also holds. First we will prove that for any $\boldsymbol{c} \in C, \boldsymbol{c} \neq \boldsymbol{0}$, the following relation is true

$$d_H(\boldsymbol{e}, \boldsymbol{0}) < d_H(\boldsymbol{e}, \boldsymbol{c}) - 2\ell. \tag{22}$$

For this, fix some $\boldsymbol{c} \in C, \boldsymbol{c} \neq \boldsymbol{0}$. Let $d \triangleq w_H(\boldsymbol{c})$. We assume without loss of generality (wlg) that all 1's appear in the first d positions in $\boldsymbol{c}$. Let d_1 denote the number of 1's in the first d positions in $\boldsymbol{e}$ and let d_0 denote the number of 1's in the last $m - d$ positions in $\boldsymbol{e}$. Then, one has

$$d_H(\boldsymbol{e}, \boldsymbol{c}) - d_H(\boldsymbol{0}, \boldsymbol{e}) = w_H(\boldsymbol{e} \oplus \boldsymbol{c}) - w_H(\boldsymbol{e}) = (d - d_1 + d_0) - (d_1 + d_0) = d - 2d_1. \tag{23}$$

Let $d_2 \triangleq \min(d - d_1, \ell + 1)$ and let $\boldsymbol{f}$ be the binary sequence obtained from $\boldsymbol{e}$ by changing d_2 0's from the first d positions into 1's. Then $d_H(\boldsymbol{e}, \boldsymbol{f}) = d_2 \leq \ell + 1$. Using further the fact that $d_H(\boldsymbol{e}, \overline{\mathcal{F}(\boldsymbol{0}, C)}) > \ell + 1$, one obtains that $\boldsymbol{f} \in \mathcal{F}(\boldsymbol{0}, C)$, which in turn leads to $d_H(\boldsymbol{f}, \boldsymbol{c}) - d_H(\boldsymbol{0}, \boldsymbol{f}) \geq 0$. Since

$$d_H(\boldsymbol{f}, \boldsymbol{c}) - d_H(\boldsymbol{0}, \boldsymbol{f}) = w_H(\boldsymbol{f} \oplus \boldsymbol{c}) - w_H(\boldsymbol{f}) = (d - d_1 - d_2 + d_0) - (d_1 + d_2 + d_0) = d - 2(d_1 + d_2),$$

it follows that $d - 2(d_1 + d_2) \geq 0$. Using the fact that $d > 0$ one obtains that $d_2 \neq d - d_1$, which leads to $d_2 = \ell + 1$, and further to $d - 2d_1 - 2(\ell + 1) \geq 0$. Using now (23), relation (22) follows.

Let us prove now (21). Assume that $\boldsymbol{e} \in \overline{\Gamma^{\ell+1}\overline{\mathcal{F}(\boldsymbol{0}, C)}}$. Take two arbitrary distinct codewords $\boldsymbol{c}', \boldsymbol{c}'' \in A$. It is sufficient to show that

$$d_H(\boldsymbol{c}', \boldsymbol{c}' \oplus \boldsymbol{e}) < d_H(\boldsymbol{c}'', \boldsymbol{c}' \oplus \boldsymbol{e}). \tag{24}$$

Let $C_0 \subseteq C$, $\boldsymbol{s} \in \mathbb{F}_2^m$ and ω be as in Lemma 1. Then $A \oplus \boldsymbol{s} = \omega(C_0)$. Let $\omega^{-1} : A \oplus \boldsymbol{s} \to C_0$ denote the inverse function of ω and let $\boldsymbol{c} \triangleq \omega^{-1}(\boldsymbol{c}' \oplus \boldsymbol{s}) \oplus \omega^{-1}(\boldsymbol{c}'' \oplus \boldsymbol{s})$ (clearly, $\boldsymbol{c} \in C$) and let $\boldsymbol{t} \triangleq \boldsymbol{c}' \oplus \boldsymbol{c}'' \oplus \boldsymbol{c}$. Then

$$\begin{aligned} w_H(\boldsymbol{t}) &= w_H\left(\left(\boldsymbol{c}' \oplus \boldsymbol{s} \oplus \omega^{-1}(\boldsymbol{c}' \oplus \boldsymbol{s})\right) \oplus \left(\boldsymbol{c}'' \oplus \boldsymbol{s} \oplus \omega^{-1}(\boldsymbol{c}'' \oplus \boldsymbol{s})\right)\right) \\ &\leq w_H\left(\boldsymbol{c}' \oplus \boldsymbol{s} \oplus \omega^{-1}(\boldsymbol{c}' \oplus \boldsymbol{s})\right) + \\ &\qquad w_H\left(\boldsymbol{c}'' \oplus \boldsymbol{s} \oplus \omega^{-1}(\boldsymbol{c}'' \oplus \boldsymbol{s})\right) \\ &\leq \ell + \ell = 2\ell, \end{aligned}$$

where the last inequality follows from (6). Further, one obtains

$$\begin{aligned} d_H(\boldsymbol{c}'', \boldsymbol{c}' \oplus \boldsymbol{e}) &= d_H(\boldsymbol{c}'' \oplus \boldsymbol{c}', \boldsymbol{e}) = d_H(\boldsymbol{c} \oplus \boldsymbol{t}, \boldsymbol{e}) \\ &\geq d_H(\boldsymbol{c}, \boldsymbol{e}) - d_H(\boldsymbol{c} \oplus \boldsymbol{t}, \boldsymbol{c}) \\ &\geq d_H(\boldsymbol{c}, \boldsymbol{e}) - 2\ell, \end{aligned} \tag{25}$$

where the second last relation follows from the triangle inequality, while the last relation is based on $w_H(\boldsymbol{t}) \leq 2\ell$. Note that $\boldsymbol{c} \in C$ and $d_H(\boldsymbol{c}', \boldsymbol{c}' \oplus \boldsymbol{e}) = d_H(\boldsymbol{e}, \boldsymbol{0})$. Then relations (22) and (25) lead to (24), completing the proof. ∎

Proof of Proposition 1: Let us fix $\boldsymbol{\lambda}_1 \in \mathcal{B}_1$ and $\boldsymbol{\lambda}, \boldsymbol{\lambda}' \in (\boldsymbol{\lambda}_1 + \mathcal{W}) \cap \mathcal{B}_2$, $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}'$. Then we need to distinguish between two cases.

Case 1: $Q_\Lambda(\boldsymbol{\lambda}) = Q_\Lambda(\boldsymbol{\lambda}')$. Then $d_H(\gamma(\boldsymbol{\lambda}), \gamma(\boldsymbol{\lambda}')) = d_H(\varphi(\boldsymbol{\lambda} \bmod \Lambda), \varphi(\boldsymbol{\lambda}' \bmod \Lambda)) \geq d_H(C)$, where the last inequality is based on the fact that $\boldsymbol{\lambda} \bmod \Lambda \neq \boldsymbol{\lambda}' \bmod \Lambda$, that φ is injective and that $\varphi(\boldsymbol{\lambda} \bmod \Lambda), \varphi(\boldsymbol{\lambda}' \bmod \Lambda) \in C$.

Case 2: $Q_\Lambda(\boldsymbol{\lambda}) \neq Q_\Lambda(\boldsymbol{\lambda}')$. Let $\boldsymbol{u} = Q_\Lambda(\boldsymbol{\lambda})$ and $\boldsymbol{u}' = Q_\Lambda(\boldsymbol{\lambda}')$. Then relation (4) implies that

$$\begin{aligned} &w_H(\gamma(\boldsymbol{\lambda}) \oplus \gamma(\boldsymbol{\lambda}')) \\ &= w_H(\varphi(\boldsymbol{\lambda} \bmod \Lambda) \oplus \psi(\boldsymbol{u}) \oplus \varphi(\boldsymbol{\lambda}' \bmod \Lambda) \oplus \psi(\boldsymbol{u}')) \\ &\geq w_H(\varphi(\boldsymbol{\lambda} \bmod \Lambda) \oplus \varphi(\boldsymbol{\lambda}' \bmod \Lambda)) - w_H(\psi(\boldsymbol{u}) \oplus \psi(\boldsymbol{u}')). \end{aligned} \tag{26}$$

As in **Case 1**, one obtains that

$$w_H(\varphi(\boldsymbol{\lambda} \bmod \Lambda) \oplus \varphi(\boldsymbol{\lambda}' \bmod \Lambda)) \geq d_H(C). \tag{27}$$

Let $\boldsymbol{u}_1 = Q_\Lambda(\boldsymbol{\lambda}_1)$. Then one has $(\boldsymbol{u}, \boldsymbol{u}_1), (\boldsymbol{u}_1, \boldsymbol{u}') \in \mathcal{N}$, which leads to $w_H(\psi(\boldsymbol{u}) \oplus \psi(\boldsymbol{u}')) \leq w_H(\psi(\boldsymbol{u}) \oplus \psi(\boldsymbol{u}_1)) + w_H(\psi(\boldsymbol{u}_1) \oplus \psi(\boldsymbol{u}')) \leq 2d_{H,\mathcal{N}}(\psi)$. Applying further equations (26) and (27) the claim follows, completing the proof. ∎

APPENDIX B
PROOF OF RESULTS IN SECTION V

Proof of Theorem 2: Recall that

$$2^{nR(\sigma)} = 2^{k(\sigma)} \times 2^{m_1(\sigma) + \cdots + m_n(\sigma)}. \tag{28}$$

Since $k(\sigma) = \lceil \log_2 |\mathcal{W}'(\sigma)| \rceil$, it follows that

$$2^{k(\sigma)} = (1 + \delta_1(\sigma))|\mathcal{W}'(\sigma)|, \tag{29}$$

for some variable $\delta_1(\sigma)$ such that $0 \leq \delta_1(\sigma) < 1$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

Further, the fact that $m_i(\sigma) = \lceil \log_2 \kappa_i(\sigma) \rceil$, $1 \leq i \leq n$, implies that

$$2^{m_1(\sigma)+\cdots+m_n(\sigma)} = (1+\delta_2(\sigma))^n \kappa_1(\sigma) \times \cdots \times \kappa_n(\sigma), \quad (30)$$

for some variable $\delta_2(\sigma)$ satisfying $0 \leq \delta_2(\sigma) < 1$.

Let $\mathcal{P}(\sigma)$ denote the smallest parallelepiped with edges parallel to the basis vectors of the lattice Λ, which contains the set $\mathcal{B}(\sigma)$. Then the definition of $\psi_{GM}(\sigma)$ implies that

$$vol(\mathcal{P}(\sigma)) = \kappa_1(\sigma) \times \cdots \times \kappa_n(\sigma) \times v(\Lambda(\sigma)). \quad (31)$$

Using equations (28)-(31) and the fact that $v(\Lambda(\sigma)) = |\mathcal{W}'(\sigma)| v(\Lambda_s(\sigma))$, it follows that

$$2^{nR(\sigma)} = \frac{vol(\mathcal{P}(\sigma))}{v(\Lambda_s(\sigma))}(1+\delta_1(\sigma))(1+\delta_2(\sigma))^n,$$

and further that,

$$R(\sigma) = \frac{1}{n}\log_2 \frac{vol(\mathcal{P}(\sigma))}{v(\Lambda_s(\sigma))} + \frac{1}{n}\log_2(1+\delta_1(\sigma)) + \log_2(1+\delta_2(\sigma)).$$

Relying further on the facts that $R_0(\sigma) = \frac{1}{n}\log_2 \frac{vol(\mathcal{A})}{v(\Lambda_s(\sigma))}$ and $0 \leq \delta_1(\sigma), \delta_2(\sigma) < 1$, one obtains that

$$\frac{1}{n}\log_2 \frac{vol(\mathcal{P}(\sigma))}{vol(\mathcal{A})} \leq R(\sigma) - R_0(\sigma) < \frac{1}{n}\log_2 \frac{vol(\mathcal{P}(\sigma))}{vol(\mathcal{A})} + 1 + \frac{1}{n}. \quad (32)$$

The design of the IA in [14] and the construction of the coarse lattice guarantee that

$$\mathcal{B}_2(\sigma) \subseteq \cup_{\lambda_s \in Q_{\Lambda_s(\sigma)}(Q_{\Lambda_c(\sigma)}(\mathcal{A}))}(\lambda_s + \mathcal{W}'(\sigma)).$$

Furthermore, recall that $\mathcal{B}(\sigma)$ is the smallest set of lattice points from $\Lambda(\sigma)$ such that $\mathcal{B}_2(\sigma) \subseteq \cup_{\boldsymbol{u} \in \mathcal{B}(\sigma)}(\boldsymbol{u} + \mathcal{W}'(\sigma))$. These observations together with the fact that $v(\Lambda_c(\sigma))$, $v(\Lambda_s(\sigma))$ and $v(\Lambda(\sigma))$ all approach 0 as $\sigma \to 0$, imply that $vol(\mathcal{P}(\sigma)) \to vol(\mathcal{P}_0)$ as $\sigma \to 0$. Then the claim follows by applying the limit as $\sigma \to 0$ in (32). ∎

Proof of Lemma 4: Consider the same notations as in the proof of Theorem 2. We will first show that

$$|\mathcal{W}'(\sigma)| = c_0 N(\sigma)(1+o(1)), \text{ as } \sigma \to 0, \quad (33)$$

where $c_0 \triangleq \frac{v(\Lambda)}{vol(\mathcal{B}(\mathcal{W}))}$. For this notice that

$$\begin{aligned} |\mathcal{W}'(\sigma)| &= \frac{vol(\Lambda(\sigma))}{vol(\Lambda_s(\sigma))} = \frac{(\mu(\sigma))^n v(\Lambda)}{v(\Lambda_s(\sigma))} \\ &= \frac{vol(\mathcal{B}(\mathcal{W}(\sigma)))v(\Lambda)}{vol(\mathcal{B}(\mathcal{W}))v(\Lambda_s(\sigma))} = c_0 \frac{vol(\mathcal{B}(\mathcal{W}(\sigma)))}{v(\Lambda_s(\sigma))}, \end{aligned}$$

where the third equality is based on (15). Using further the fact that $vol(\mathcal{B}(\mathcal{W}(\sigma))) = N(\sigma)(v(\Lambda_s(\sigma))(1+o(1))$ as $\sigma \to 0$, relation (33) follows.

Further, plugging (33) in (29) and applying the logarithm, one obtains that

$$\begin{aligned} k(\sigma) &= \log_2 N(\sigma) + \log_2 (c_0(1+\delta_1(\sigma))(1+o(1))) \\ &= (\log_2 N(\sigma))(1+o(1)), \end{aligned} \quad (34)$$

as $\sigma \to 0$, where the last equality uses the fact that $\lim_{\sigma \to 0} N(\sigma) = \infty$. Notice that relations (16) in conjunction with the fact that $\lim_{\sigma \to 0} R_0(\sigma) = \infty$ imply that

$$R(\sigma) = R_0(\sigma)(1+o(1)). \quad (35)$$

Recall that $a = \frac{\log_2 N(\sigma)}{R_0(\sigma)}$. Then equations (34) and (35) imply that

$$\lim_{\sigma \to 0} \frac{k(\sigma)}{nR(\sigma)} = \lim_{\sigma \to 0} \frac{\log_2 N(\sigma)}{R_0(\sigma)} = a,$$

completing the proof. ∎

Proof of Theorem 3: We introduce a few notations first. For integers $0 < m' < m$, an (m, m') linear channel code is an m'-dimensional vector subspace of $\mathbb{F}_2^m$. For $0 < q < 1/2$ and $0 < r < 1 - H(q)$, let $E(r,q)$ denote the *random coding exponent* defined in [23, Equation (5.6.16)]. It is proved in [23, Theorem 5.6.4] that $E(r,q) > 0$ and that $E(r,q)$ is a convex function of r, for $0 < r < 1 - H(q)$, which implies that $E(r,q)$ is also continuous in r.

Let $r(\sigma) \triangleq \frac{k(\sigma)}{nR(\sigma)}$. Let us prove the first claim. Lemma 4 implies that $\lim_{\sigma \to 0} r(\sigma) = a$. Using further the fact that $a < 1 - H(p)$ it follows that there is some $\sigma_0 > 0$ such that for all $\sigma, 0 < \sigma < \sigma_0$, one has $r(\sigma) < 1 - H(p)$. Then, according to [23, Corollary of Theorem 6.2.1], for each $\sigma, 0 < \sigma < \sigma_0$, there exists an $(nR(\sigma), k(\sigma))$ linear channel code $C(\sigma)$ such that the following holds

$$P_e(C(\sigma)) \leq 2^{-nR(\sigma)E(r(\sigma),p)}.$$

Now let $\ell(\sigma) \triangleq d_{H,\mathcal{N}}(\psi(\sigma))$. Using further Theorem 1, one obtains that

$$P_e(\gamma(\sigma)) \leq 2^{-nR(\sigma)(E(r(\sigma),p) - \epsilon(\ell(\sigma)+1, nR(\sigma), p))}. \quad (36)$$

Let c be the constant defined in Proposition 2. The value of c is invariant to lattice scaling, therefore it does not depend on σ. Proposition 2 implies that $\ell(\sigma) \leq 3c\sqrt{n}$. Additionally, since $\lim_{\sigma \to 0} R(\sigma) = \infty$, it follows that for σ sufficiently small, one has $\frac{3c\sqrt{n}+1}{nR(\sigma)} < \frac{1}{2}$. Using the above observations together with (7) and with the fact that the entropy function $H(\cdot)$ is increasing on $(0, 1/2)$, one obtains that

$$\epsilon(\ell(\sigma)+1, nR(\sigma), p) \leq \epsilon(3c\sqrt{n}+1, nR(\sigma), p)$$

for σ sufficiently small. This relation in conjunction with (36), further leads to

$$P_e(\gamma(\sigma)) \leq 2^{-nR(\sigma)(E(r(\sigma),p) - \epsilon(3c\sqrt{n}+1, nR(\sigma), p))}. \quad (37)$$

The fact that $\lim_{\sigma \to 0} R(\sigma) = \infty$ implies that $\lim_{\sigma \to 0} \epsilon(3c\sqrt{n} + 1, nR(\sigma), p)) = 0$. Using further the fact that $E(r,p)$ is continuous in r, one obtains that

$$\begin{aligned} &\lim_{\sigma \to 0}(E(r(\sigma),p) - \epsilon(3c\sqrt{n}+1, nR(\sigma), p)) \\ &= \lim_{\sigma \to 0} E(r(\sigma),p) - \lim_{\sigma \to 0} \epsilon(3c\sqrt{n}+1, nR(\sigma), p)) \\ &= E(a,p) - 0 > 0. \end{aligned}$$

The above relations, in conjunction with (37) and the fact that $\lim_{\sigma \to 0} R(\sigma) = \infty$, imply that (17) holds.

Let us prove now the second part. Assume that (17) holds for some choice of the channel codes $C(\sigma)$, $0 < \sigma \leq 1$. Fix some $\boldsymbol{\lambda}_1 \in \mathcal{B}_1$ such that $(\boldsymbol{\lambda}_1 + \mathcal{W}(\sigma)) \cap \mathcal{B}_2 = \boldsymbol{\lambda}_1 + \mathcal{W}(\sigma)$. When the first description lattice point is $\boldsymbol{\lambda}_1$, the set of possible indexes j_2 is $S(\sigma) \triangleq \gamma((\boldsymbol{\lambda}_1 + \mathcal{W}(\sigma)))$. If we regard $S(\sigma)$ as a channel code then its rate is $r'(\sigma) = \frac{\log_2 N(\sigma)}{nR(\sigma)}$. Examining

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

the proof of Lemma 4, it follows that $\lim_{\sigma\to 0} r'(\sigma) = a$. Then for each $\epsilon > 0$ there is $\sigma_\epsilon > 0$ such that $r'(\sigma) \geq a - \epsilon$ for all $0 < \sigma < \sigma_\epsilon$. For each ϵ and $0 < \sigma < \sigma_\epsilon$ let $S_\epsilon(\sigma)$ denote the rate-$(a-\epsilon)$ channel code obtained from $S(\sigma)$ by removing $N(\sigma) - 2^{nR(\sigma)(a-\epsilon)}$ codewords. Notice that as $\sigma \to 0$, the blocklength of the channel code $S(\sigma)$ approaches ∞. Additionally, one has $P_e(S_\epsilon(\sigma)) \leq P_e(S(\sigma)) \leq P_e(\gamma(\sigma))$. Using (17) we obtain that $\lim_{\sigma\to 0} P_e(S_\epsilon(\sigma)) = 0$. Then, in virtue of the Channel Coding Theorem [24, Theorem 8.7.1] one has $a - \epsilon \leq 1 - H(p)$. Since the previous inequality holds for all $\epsilon > 0$ it follows that $a \leq 1 - H(p)$, concluding the proof. ■

Proof of Proposition 3: i) In [14] it was shown that, for a smooth pdf $f(x)$, one has

$$\lim_{\sigma\to 0}(R_e^{(n)}(\sigma) + \log_2 v(\Lambda_s^{(n)}(\sigma)) = h(f).$$

Using, additionally, the fact that

$$R_0^{(n)}(\sigma) = \frac{1}{n}\log_2 \frac{vol(\mathcal{A}_{\epsilon_n}^{(n)})}{v(\Lambda_s^{(n)}(\sigma))},$$

leads to

$$\lim_{\sigma\to 0}(R_0^{(n)}(\sigma) - R_e^{(n)}(\sigma)) = \frac{1}{n}\log_2 vol(\mathcal{A}_{\epsilon_n}^{(n)}) - h(f). \quad (38)$$

Using arguments similar to those in the proof of [24, Theorem 8.2.2] one obtains that

$$\mathbb{P}\left[\boldsymbol{X} \in \mathcal{A}_{\epsilon_n}^{(n)}\right] 2^{n(h(f)-\epsilon_n)} \leq vol(\mathcal{A}_{\epsilon_n}^{(n)}) \leq 2^{n(h(f)+\epsilon_n)},$$

which leads to

$$\frac{1}{n}\log_2 \mathbb{P}\left[\boldsymbol{X} \in \mathcal{A}_{\epsilon_n}^{(n)}\right] + h(f) - \epsilon_n \leq \frac{1}{n}\log_2 vol(\mathcal{A}_{\epsilon_n}^{(n)}) \leq h(f) + \epsilon_n.$$

Based on relation (19) and the fact that $\lim_{n\to\infty}\epsilon_n = 0$, it further follows that

$$\lim_{n\to\infty}\frac{1}{n}\log_2 vol(\mathcal{A}_{\epsilon_n}^{(n)}) = h(f). \quad (39)$$

Using further relation (38) claim i) follows.

ii) For each $\boldsymbol{x} \in \mathbb{R}^n$ and $i = 0, 1, 2$, let $\hat{\boldsymbol{x}}_i$, respectively $\hat{\boldsymbol{x}}_{i,e}$, denote the reconstruction at the side decoder i, for $i = 1, 2$, respectively at the central decoder, for $i = 0$, in the fixed-rate MDLVQ, respectively in the entropy-coded MDLVQ. Recall that for $\boldsymbol{x} \in \mathcal{A}^{(n)}$, we have $\hat{\boldsymbol{x}}_i = \hat{\boldsymbol{x}}_{i,e}$, while for $\boldsymbol{x} \in \mathbb{R}^n \setminus \mathcal{A}^{(n)}$ we have $\hat{\boldsymbol{x}}_i = \boldsymbol{\mu}$, for $i = 0, 1, 2$. Let us fix $i \in \{0, 1, 2\}$. It follows that

$$d_{i,e}^{(n)}(\sigma) - d_i^{(n)}(\sigma) = \frac{1}{n}\int_{\mathbb{R}^n\setminus\mathcal{A}^{(n)}} ||\boldsymbol{x} - \hat{\boldsymbol{x}}_{i,e}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} - \frac{1}{n}\int_{\mathbb{R}^n\setminus\mathcal{A}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x},$$

which further implies that

$$d_i^{(n)}(\sigma) \leq d_{i,e}^{(n)}(\sigma) + \frac{1}{n}\int_{\mathbb{R}^n\setminus\mathcal{A}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x}.$$

Thus, to prove the claim it is sufficient to show that

$$\lim_{n\to\infty}\frac{1}{n}\int_{\mathbb{R}^n\setminus\mathcal{A}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} = 0. \quad (40)$$

For this, consider for each positive ζ and $n \geq 1$ the set

$$\mathcal{S}_\zeta^{(n)} \triangleq \left\{\boldsymbol{x} \in \mathbb{R}^n \middle| \left|\frac{1}{n}||\boldsymbol{x} - \boldsymbol{\mu}||^2 - \sigma^2\right| \leq \zeta\right\},$$

Using the Weak Law of Large Numbers one obtains that there is a sequence of positive values ζ_n such that $\lim_{n\to\infty}\zeta_n = 0$ and

$$\lim_{n\to\infty}\mathbb{P}\left[\boldsymbol{X} \in \mathcal{S}_{\zeta_n}^{(n)}\right] = 1. \quad (41)$$

Now notice that

$$\frac{1}{n}\int_{\mathbb{R}^n\setminus\mathcal{A}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} \leq \frac{1}{n}\int_{\mathbb{R}^n\setminus\mathcal{S}_{\zeta_n}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} + \frac{1}{n}\int_{\mathcal{S}_{\zeta_n}^{(n)}\setminus\mathcal{A}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x}. \quad (42)$$

The definition of $\mathcal{S}_\zeta^{(n)}$ implies that

$$\frac{1}{n}\int_{\mathcal{S}_{\zeta_n}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} \geq (\sigma^2 - \zeta_n)\mathbb{P}\left[\boldsymbol{X} \in \mathcal{S}_{\zeta_n}^{(n)}\right].$$

Then

$$\frac{1}{n}\int_{\mathbb{R}^n\setminus\mathcal{S}_{\zeta_n}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} = \frac{1}{n}\int_{\mathbb{R}^n} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} - \frac{1}{n}\int_{\mathcal{S}_{\zeta_n}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} \leq \sigma^2 - (\sigma^2 - \zeta_n)\mathbb{P}\left[\boldsymbol{X} \in \mathcal{S}_{\zeta_n}^{(n)}\right].$$

Relation (41) and the fact that $\lim_{n\to\infty}\zeta_n = 0$ further imply that

$$\lim_{n\to\infty}\frac{1}{n}\int_{\mathbb{R}^n\setminus\mathcal{S}_{\zeta_n}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} = 0. \quad (43)$$

Further, one obtains that

$$\lim_{n\to\infty}\frac{1}{n}\int_{\mathcal{S}_{\zeta_n}^{(n)}\setminus\mathcal{A}^{(n)}} ||\boldsymbol{x} - \boldsymbol{\mu}||^2 \boldsymbol{f}(\boldsymbol{x})d\boldsymbol{x} \leq \lim_{n\to\infty}(\sigma^2 + \zeta_n)\left(1 - \mathbb{P}\left[\boldsymbol{X} \in \mathcal{A}_{\epsilon_n}^{(n)}\right]\right) = 0. \quad (44)$$

Finally, relations (42), (43) and (44) imply that (40) holds, and the proof of claim ii) is completed.

iii) Using (18) one obtains

$$\rho^{(n)}(\sigma) - (1-a) = \frac{(1+a)R^{(n)}(\sigma) - R_{e,c}^{(n)}(\sigma)}{R^{(n)}(\sigma)} = \frac{(1+a)(R^{(n)}(\sigma) - R_0^{(n)}(\sigma))}{R^{(n)}(\sigma)} + \frac{(1+a)R_0^{(n)}(\sigma) - h(f) + \frac{1}{n}\log_2(v(\Lambda_c^{(n)}(\sigma)))}{R^{(n)}(\sigma)} + \frac{h(f) - \frac{1}{n}\log_2(v(\Lambda_c^{(n)}(\sigma))) - R_{e,c}^{(n)}(\sigma)}{R^{(n)}(\sigma)}.$$

The first term in the last expression approaches 0, as $\sigma \to 0$, since $R^{(n)}(\sigma) \to \infty$, while the numerator is bounded in virtue of Theorem 2. The last term also approaches 0, as $\sigma \to 0$, since $R^{(n)}(\sigma) \to \infty$, while the numerator goes to 0 [14]. Therefore, it is sufficient to show that

$$\lim_{n\to\infty}\lim_{\sigma\to 0}\frac{(1+a)R_0^{(n)}(\sigma) - h(f) + \frac{1}{n}\log_2(v(\Lambda_c^{(n)}(\sigma)))}{R^{(n)}(\sigma)} = 0. \quad (45)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCOMM.2018.2816070, IEEE Transactions on Communications

From the definition of a and $R_0^{(n)}(\sigma)$ we obtain that

$$(1+a)R_0^{(n)}(\sigma) = \frac{1}{n}\log_2 \frac{Nvol(\mathcal{A}^{(n)})}{v(\Lambda_s^{(n)}(\sigma))} = \frac{1}{n}\log_2 \frac{vol(\mathcal{A}^{(n)})}{v(\Lambda_c^{(n)}(\sigma))}.$$

Then (45) becomes equivalent to

$$\lim_{n\to\infty}\lim_{\sigma\to 0} \frac{\frac{1}{n}\log_2(vol(\mathcal{A}^{(n)})) - h(f)}{R^{(n)}(\sigma)} = 0,$$

which is true based on (39), the fact that $\mathcal{A}_{\epsilon_n}^{(n)} = \mathcal{A}^{(n)}$ and $\lim_{\sigma\to 0} R^{(n)}(\sigma) = \infty$. ■

REFERENCES

[1] V. K. Goyal, "Multiple description coding: Compression meets the network", *IEEE Signal Process. Magazine*, vol. 18, no. 5, pp. 74–93, Sep. 2001.

[2] Y. Wang, A. R. Reibman, and S. Lin, "Multiple description coding for video delivery", *Proc. of the IEEE*, vol. 93, no. 1, pp. 57–70, Jan. 2005.

[3] N. Gadgil, M. Yang, M. Cormer, and E. Delp, *Multiple description coding*, Academic press Library in Signal Proc.: Image and Video Compression and Multimedia, vol. 5 (2014), 251.

[4] S.-M. Yang and V. A. Vaishampayan, "Low-delay communication for Rayleigh fading channels: an application of the multiple description quantizer", *IEEE Trans. Communic.*, vol. 43, no. 11, pp. 2771–2783, Nov. 1995.

[5] M. Srinivasan, "Iterative decoding of multiple descriptions," in *Proc. IEEE Data Compress. Conf.*, Snowbird, UT, Mar. 1999, pp. 463–472.

[6] J. Barros, J. Hagenauer, and N. Gortz, "Turbo cross decoding of multiple descriptions," in *Proc. IEEE Intern. Conf. Communic.*, New York, NY, USA, Apr.-May 2002, pp. 1398–1402.

[7] T. Guionnet and C. Guillemot, "Soft decoding of multiple descriptions," in *Proc. IEEE Intern. Conf. Multimedia Expo*, Lausanne, Switzerland, Aug. 2002, pp. 601–604.

[8] J. N. Laneman, E. Martinian, G. W. Wornell, and J. G. Apostolopoulos, "Source-channel diversity for parallel channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 10, pp. 3518–3539, Oct. 2005.

[9] I. Bahceci, Y. Altunbasak, and T. M. Duman, "A turbo-coded multiple-description system for multiple antennas," *IEEE Trans. Communic.*, vol. 54, no.2, pp. 187–191, Feb. 2006.

[10] X. Wu, X. Wang, and Z. Wang, "Resource-scalable joint source-channel MAP and MMSE estimation of multiple descriptions," *IEEE Trans. Signal Proc.*, vol. 57, no. 1, pp. 279–288, Jan. 2009.

[11] R. Ma and F. Labeau, "Error-resilient multiple description coding," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3996-4007, Aug. 2008.

[12] Y. Zhou and W. Chan, "Multiple description quantizer design for multiple-antenna systems with MAP detection," *IEEE Trans. Communic.*, vol. 58, no. 1, pp. 136-145, Jan. 2010.

[13] S. Dumitrescu and Y. Wan, "Bit-error resilient index assignment for multiple description scalar quantizers," *IEEE Trans. Inform. Theory*, vol. 61, no. 5, pp. 2748-2763, May 2015.

[14] V. A. Vaishampayan, N. Sloane, and S. Servetto, "Multiple description vector quantization with lattice codebooks: design and analysis," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 1718–1734, Jul. 2001.

[15] S. N. Diggavi, N. Sloane, and V. A. Vaishampayan, "Asymmetric multiple description lattice vector quantizers," *IEEE Trans. Inform. Theory*, vol. 48, no. 1, pp. 174–191, Jan. 2002.

[16] J. Ostergaard, J. Jensen, and R. Heusdens, "n-channel entropy-constrained multiple description lattice vector quantization," *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 1956–1973, May 2006.

[17] X. Huang and X. Wu, "Optimal index assignment for multiple description lattice vector quantization," in *Proc. IEEE Data Compress. Conf.*, Mar. 2006, pp. 272-281.

[18] J. Ostergaard, R. Heusdens, and J. Jensen, "n-channel asymmetric entropy-constrained multiple-description lattice vector quantization," *IEEE Trans. Inform. Theory*, vol. 56, no. 12, pp. 6354–6375, Dec. 2010.

[19] G. Zhang, J. Klejsa, and W. B. Kleijn, "Optimal index assignment for multiple description scalar quantization with translated lattice codebooks," *IEEE Trans. Signal Proc.*, vol. 60, no. 8, pp. 4444–4451, Aug. 2012.

[20] Z. Gao and S. Dumitrescu, "Flexible symmetric multiple description lattice vector quantizer with $L \geq 3$ descriptions, *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4281–4292, Dec. 2014.

[21] S. Dumitrescu, Y. Chen, and J. Chen, "Index mapping for bit-error resilient multiple description lattice vector quantizer", in *Proc. Intern. Symp. Inform. Theory*, Aachen, Germany, June 2017, pp. 2588–2592.

[22] J.H. Conway and N.J.A. Sloane, *Sphere packings, lattices and groups*, Springer, New York, 1993.

[23] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley and Sons, 1968.

[24] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Ed., John Wiley and Sons, 2006.

[25] R. Zamir, *Lattice Coding for Signals and Networks*, Cambridge University Press, 2014.

[26] G. E. Shilov, *Elementary real and complex analysis*, Dover Publications, New York, 1996.

[27] M. Suzuki, *Group Theory. Vol. 1*, Springer-Verlag, 1982.

[28] I. Csiszar and J. Korner, *Information Theory: Coding Theorems for Discrete Memoryless Sources*, 2nd Ed., Cambridge University Press, 2011.

PLACE PHOTO HERE

Sorina Dumitrescu (M'05-SM'13) received the B.Sc. and Ph.D. degrees in mathematics from the University of Bucharest, Romania, in 1990 and 1997, respectively. From 2000 to 2002 she was a Postdoctoral Fellow in the Department of Computer Science at the University of Western Ontario, London, Canada. Since 2002 she has been with the Department of Electrical and Computer Engineering at McMaster University, Hamilton, Canada, where she held Postdoctoral and Research Associate positions, and where she is currently an Associate Professor. Her current research interests include multimedia coding and communications, network-aware data compression, multiple description codes, joint source-channel coding, signal quantization. Her earlier research interests were in formal languages and automata theory. She was a recipient of the NSERC University Faculty Award during 2007-2012.

PLACE PHOTO HERE

Yifang Chen received the B.Eng. degree in communication engineering from Nanjing University of Science and Technology, Nanjing, China, in 2010 and the M.A.Sc. degree in electrical and computer engineering from McMaster University, Hamilton, ON , Canada, in 2015. She is currently with Google, Kitchener, ON. Her research interests include problems in source coding.

PLACE PHOTO HERE

Jun Chen (S'03-M'06-SM'16) received the B.E. degree with honors in communication engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University, Ithaca, NY, in 2004 and 2006, respectively.

He was a Postdoctoral Research Associate in the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign, Urbana, IL, from September 2005 to July 2006, and a Postdoctoral Fellow at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, from July 2006 to August 2007. Since September 2007 he has been with the Department of Electrical and Computer Engineering at McMaster University, Hamilton, ON, Canada, where he is currently an Associate Professor and a Joseph Ip Distinguished Engineering Fellow. His research interests include information theory, machine learning, wireless communications, and signal processing.

He received several awards for his research, including the Josef Raviv Memorial Postdoctoral Fellowship in 2006, the Early Researcher Award from the Province of Ontario in 2010, and the IBM Faculty Award in 2010. He served as an Associate Editor for the IEEE Transactions on Information Theory from 2014 to 2016.