# Multiple Description Quantization Via Gram–Schmidt Orthogonalization

Jun Chen, *Member, IEEE*, Chao Tian, *Member, IEEE*, Toby Berger, *Fellow, IEEE*, and Sheila S. Hemami, *Senior Member, IEEE*

***Abstract*—The multiple description (MD) problem has received considerable attention as a model of information transmission over unreliable channels. A general framework for designing efficient MD quantization schemes is proposed in this paper. We provide a systematic treatment of the El Gamal–Cover (EGC) achievable MD rate–distortion region, and show it can be decomposed into a simplified-EGC (SEGC) region and a superimposed refinement operation. Furthermore, any point in the SEGC region can be achieved via a successive quantization scheme along with quantization splitting. For the quadratic Gaussian case, the proposed scheme has an intrinsic connection with the Gram–Schmidt orthogonalization, which implies that the whole Gaussian MD rate–distortion region is achievable with a sequential dithered lattice-based quantization scheme as the dimension of the (optimal) lattice quantizers becomes large. Moreover, this scheme is shown to be universal for all independent and identically distributed (i.i.d.) smooth sources with performance no worse than that for an i.i.d. Gaussian source with the same variance and asymptotically optimal at high resolution. A class of MD scalar quantizers in the proposed general framework is also constructed and is illustrated geometrically; the performance is analyzed in the high-resolution regime, which exhibits a noticeable improvement over the existing MD scalar quantization schemes.**

***Index Terms*—Gram–Schmidt orthogonalization, lattice quantization, minimum mean-square error (MMSE), multiple descriptions (MDs), quantization splitting.**

## I. Introduction

In the multiple description (MD) problem, the total available bit rate is split between two channels and either channel may be subject to failure. It is desired to allocate rate and coded representations between the two channels, such that if one channel fails, an adequate reconstruction of the source is possible, but if both channels are available, an improved reconstruction over the single-channel reception results. The formal definition of the MD problem is as follows (also see Fig. 1).

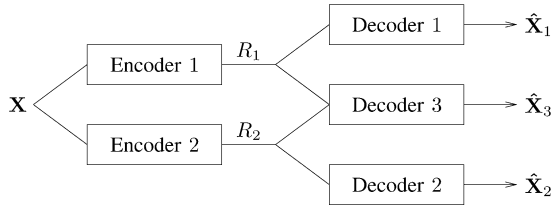


Fig. 1. Encoder and decoder diagram for MDs.

Let $\{X(t)\}_{t=1}^{\infty}$ be an independent and identically distributed (i.i.d.} random process with $X(t) \sim p(x)$ for all $t$. Let $d(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to [0, d_{\max}]$ be a distortion measure.

*Definition 1.1:* The quintuple $(R_1, R_2, D_1, D_2, D_3)$ is called achievable if for all $\varepsilon > 0$, there exist, for $n$ sufficiently large, encoding functions

$$f_i^{(n)} : \mathcal{X}^n \to \mathcal{C}_i^{(n)} \quad \log\left|\mathcal{C}_i^{(n)}\right| \le n(R_i + \varepsilon), \qquad i = 1,2$$

and decoding functions

$$g_i^{(n)} : \mathcal{C}_i^{(n)} \to \mathcal{X}^n, \qquad i = 1,2$$
$$g_3^{(n)} : \mathcal{C}_1^{(n)} \times \mathcal{C}_2^{(n)} \to \mathcal{X}^n$$

such that for $\hat{\boldsymbol{X}}_i = g_i^{(n)}(f_i^{(n)}(\boldsymbol{X})), i = 1,2$, and for $\hat{\boldsymbol{X}}_3 = g_3^{(n)}(f_1^{(n)}(\boldsymbol{X}), f_2^{(n)}(\boldsymbol{X}))$

$$\frac{1}{n}\mathbb{E}\sum_{t=1}^{n} d(X(t), \hat{X}_i(t)) \le D_i + \varepsilon, \qquad i = 1,2,3.$$

The MD rate-distortion region, denoted by $\mathcal{Q}$, is the set of all achievable quintuples.

In this paper, the encoding functions $f_1^{(n)}$ and $f_2^{(n)}$ are referred to as encoder 1 and encoder 2, respectively. Similarly, decoding functions $g_1^{(n)}$, $g_2^{(n)}$ and $g_3^{(n)}$ are referred to as decoder 1, decoder 2, and decoder 3, respectively. It should be emphasized that in a real system, encoders 1 and 2 are just two different encoding functions of a single encoder while decoders 1, 2, and 3 are different decoding functions of a single decoder. Alternatively, in the MD literature decoders 1 and 2 are sometimes referred to as the side decoders because of their positions in Fig. 1, while decoder 3 is referred to as the central decoder.

Early contributions to the MD problem can be found in [1]–[4]. The first general result was El Gamal and Cover's achievable region.

Manuscript received April 1, 2005; revised November 16, 2005. The work of J. Chen and T. Berger was supported in part by the National Science Foundation under Grant CCR-033 0059 and under a grant from the National Academies Keck Futures Initiative (NAKFI). The material in this work was presented in part at the 39th Annual Conference on Information Sciences and Systems, Princeton, NJ, March 2005 and at the IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005.

J. Chen was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853. He is now with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: junchen@us.ibm.com).

C. Tian was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853. He is now with Laboratory of Information and Communication Systems, Swiss Federal Institute of Technology at Lausanne (EPFL), Lausanne, CH-1015 Switzerland (e-mail: chao.tian@epfl.ch).

T. Berger was with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853. He is now with the Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904 USA (e-mail: tb6n@virginia.edu).

S. S. Hemami is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: hemami@ece.cornell.edu).

Communicated by M. Effros, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2006.885498

*Definition 1.2 (El Gamal–Cover (EGC) Region):* For random variables $U_1, U_2$, and $U_3$ jointly distributed with the generic source variable $X$ via conditional distribution $p(u_1, u_2, u_3 \mid x)$, let

$$\mathcal{R}(U_1,U_2,U_3) = \{(R_1,R_2) : R_1 + R_2 \geq I(X;U_1,U_2,U_3) + I(U_1;U_2), R_i \geq I(X;U_i), i = 1,2\}.$$

Let

$$\mathcal{Q}(U_1,U_2,U_3) = \{(R_1,R_2,D_1,D_2,D_3) : (R_1,R_2) \in \mathcal{R}(U_1,U_2,U_3), \exists \hat{X}_i = g_i(U_i) \text{ with } \mathbb{E}d(X,\hat{X}_i) \leq D_i, i = 1,2,3\}.$$

The EGC region[1] is then defined as

$$\mathcal{Q}_{\mathrm{EGC}} = \mathrm{conv}\left( \bigcup_{p(u_1,u_2,u_3 \mid x)} \mathcal{Q}(U_1,U_2,U_3) \right)$$

where $\mathrm{conv}(\mathcal{S})$ denotes the convex hull of $\mathcal{S}$ for any set $\mathcal{S}$ in the Euclidean space.

It was proved in [5] that $\mathcal{Q}_{\mathrm{EGC}} \subseteq \mathcal{Q}$. Ozarow [3] showed that $\mathcal{Q}_{\mathrm{EGC}} = \mathcal{Q}$ for the quadratic Gaussian source. Ahlswede [6] showed that the EGC region is also tight for the "no excess sum-rate" case. Zhang and Berger [7] constructed a counterexample for which $\mathcal{Q}_{\mathrm{EGC}} \subset \mathcal{Q}$. Further results can be found in [8]–[14]. The MD problem has also been generalized to the $n$-channel case [15], [16], but even the quadratic Gaussian case is far from being completely understood. The extension of the MD problem to the distributed source coding scenario has been considered in [17], [18], where the problem is again widely open.

Many practical methods exist for constructing MDs (see [19] for an excellent review). Among them, the MD scalar quantization (MDSQ) proposed by Vaishampayan [20], [21] explicitly takes the quantization approach. The optimization of the key component of this method, namely the index assignment, turns out to be a difficult problem. Several heuristic methods are provided in [20] to construct balanced index assignments, which achieve a central and side distortion product 3.07 dB away from the rate–distortion bound at high resolution [22]; this granular distortion gap can be further reduced [23]. The problem of design good index assignment was also tackled by other researchers [24]–[27]. The MDSQ framework was later extended to MD lattice vector quantization (MDLVQ) for balanced descriptions in [28] and for the asymmetric case in [29]. The design relies heavily on the lattice/sublattice structure to facilitate the construction of index assignments. The analysis on these quantizers shows that the constructions are high-resolution optimal in asymptotically high dimensions.

More relevant to our work is [32], where Frank-Dayan and Zamir proposed a class of MD schemes based on entropy-coded dithered lattice quantizers (ECDQs). The system consists of two independently dithered lattice quantizers as the two side quantizers, with a possible third dithered lattice quantizer to provide refinement information for the central decoder. It was found that even with the quadratic Gaussian source, this system is only optimal in asymptotically high dimensions for the degenerate cases such as successive refinement and the "no excess marginal-rate" case, but not optimal in general. The difficulty lies in generating dependent quantization errors of two side quantizers to simulate the Gaussian MD test channel. Several possible improvements were provided in [32], but the problem remains unsolved.

In this paper we provide a systematic treatment of the EGC achievable MD rate-distortion region and show it can be decomposed into a simplified-EGC (SEGC) region and a superimposed refinement operation. Furthermore, any point in the SEGC region can be achieved via a successive quantization scheme along with quantization splitting. For the quadratic Gaussian case, the MD rate–distortion region is the same as the SEGC region, and the proposed scheme has an intrinsic connection with the Gram–Schmidt orthogonalization method. Thus, we use single-description ECDQs with independent subtractive dithers as building blocks for this MD coding scheme, by which the difficulty of generating dependent quantization errors is circumvented. Analytical expressions for the rate–distortion performance of this system are then derived for general sources, and compared to the optimal rate regions at both high and low lattice dimensions. The proposed scheme is conceptually different from those in [32], and it can achieve the whole Gaussian MD rate–distortion region as the dimension of the (optimal) lattice quantizers becomes large, unlike the method proposed in [32]. The scheme is further illustrated in Section VI using an undithered scalar quantization system, whose high resolution analysis shows promising performance.

The remainder of this paper is divided into six sections. In Section II, ECDQ and the Gram–Schmidt orthogonalization method are briefly reviewed and a connection between the successive quantization scheme and the Gram–Schmidt orthogonalization method is established. In Section III, we present a systematic treatment of the EGC region and show the sufficiency of a successive quantization scheme along with quantization splitting. In Section IV, the quadratic Gaussian case is considered in more depth. In Section V, the proposed scheme based on ECDQ is shown to be universal for all i.i.d. smooth sources with performance no worse than that for an i.i.d. Gaussian source with the same variance and asymptotically optimal at high resolution. A scalar MD quantization scheme in our framework is given in Section VI. Some further extensions are suggested in Section VII, which also serves as the conclusion. Throughout, we use boldfaced letters to indicate ($n$-dimensional) vectors, capital letters for random objects, and small letters for their realizations. For example, we let $\boldsymbol{X} = (X(1), \cdots, X(n))^T$ and $\boldsymbol{x} = (x(1), \cdots, x(n))^T$.

## II. Entropy-Coded Dithered Quantization and Gram–Schmidt Orthogonalization

In this section, we first give a brief review of ECDQ, and then explain the difficulty of applying ECDQ directly to the MD problem. As a method to resolve this difficulty, the Gram–Schmidt orthogonalization is introduced and a connection between the sequential (dithered) quantization and the Gram–Schmidt orthogonalization is established. The purpose of this section is twofold: The first is to review related results on ECDQ and the Gram–Schmidt orthogonalization and show

[1]The form of the EGC region here is slightly different from the one given in [5], but it is straightforward to show that they are equivalent.

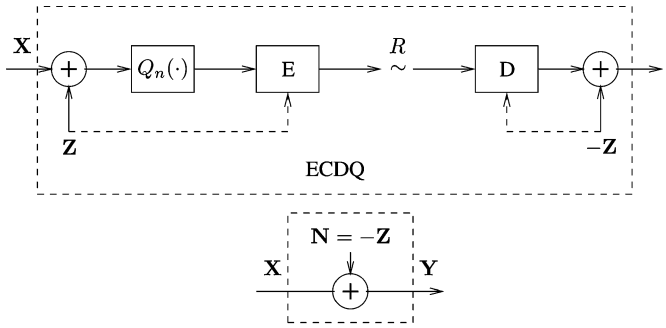


Fig. 2. ECDQ and its equivalent additive-noise channel.

their connection, while the second is to explicate the intuition that motivated this work.

### A. Review of Entropy-Coded Dithered Quantization

Some basic definitions and properties of ECDQ from [32] are quoted below. More detailed discussion and derivation can be found in [37]–[40]. An $n$-dimensional lattice quantizer is formed from a lattice $\boldsymbol{L}_n$. The quantizer $Q_n(\cdot)$ maps each vector $\boldsymbol{x} \in \mathcal{R}^n$ into the lattice point $\boldsymbol{l}_i \in \boldsymbol{L}_n$ that is nearest to $\boldsymbol{x}$. The region of all $n$-vectors mapped into a lattice point $\boldsymbol{l}_i \in \boldsymbol{L}_n$ is the Voronoi region

$$V(\boldsymbol{l}_i) = \{\boldsymbol{x} \in \mathcal{R}^n : \|\boldsymbol{x} - \boldsymbol{l}_i\| \le \|\boldsymbol{x} - \boldsymbol{l}_j\|, \forall j \ne i\}.$$

The dither $\boldsymbol{Z}$ is an $n$-dimensional random vector, independent of the source, and uniformly distributed over the basic cell $V_0$ of the lattice which is the Voronoi region of the lattice point $\boldsymbol{0}$. The dither vector is assumed to be available to both the encoder and the decoder. The normalized second moment $G_n$ of the lattice characterizes the second moment of the dither vector

$$\frac{1}{n}\mathbb{E}\|\boldsymbol{Z}\|^2 = G_n V^{2/n}$$

where $V$ denotes the volume of $V_0$. Both the entropy encoder and the decoder are conditioned on the dither sample $\boldsymbol{Z}$; furthermore, the entropy coder is assumed to be ideal. The lattice quantizer with dither represents the source vector $\boldsymbol{X}$ by the vector $\boldsymbol{W} = Q_n(\boldsymbol{X} + \boldsymbol{Z}) - \boldsymbol{Z}$. The resulting properties of the ECDQ are as follows.

1) The quantization error vector $\boldsymbol{W} - \boldsymbol{X}$ is independent of $\boldsymbol{X}$ and is distributed as $-\boldsymbol{Z}$. In particular, the mean-squared quantization error is given by the second moment of the dither, independently of the source distribution, i.e.,

$$\frac{1}{n}\mathbb{E}\|\boldsymbol{W} - \boldsymbol{X}\|^2 = \frac{1}{n}\mathbb{E}\|\boldsymbol{Z}\|^2 = G_n V^{2/n}.$$

2) The coding rate of the ECDQ is equal to the mutual information between the input and output of an additive noise channel $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}$, where $\boldsymbol{N}$, the channel's noise, has the same probability density function as $-\boldsymbol{Z}$ (see Fig. 2)

$$H(Q_n(\boldsymbol{X} + \boldsymbol{Z}) \,|\, \boldsymbol{Z}) = I(\boldsymbol{X}; \boldsymbol{Y}) = h(\boldsymbol{Y}) - h(\boldsymbol{N}).$$

3) For optimal lattice quantizers, i.e., lattice quantizers with the minimal normalized second moment $G_n$, the autocorrelation of the quantizer noise is "white," i.e., $\mathbb{E}\boldsymbol{Z}\boldsymbol{Z}^T = \sigma^2 I_n$ where $I_n$ is the $n \times n$ identity matrix, $\sigma^2 = G_n^{\text{opt}} V^{2/n}$ is the second moment of the lattice, and

$$G_n^{\text{opt}} = \min_{Q_n(\cdot)} \frac{\int_{V_0} \|\boldsymbol{x}\|^2 \mathrm{d}\boldsymbol{x}}{nV^{1+\frac{2}{n}}}$$

is the minimal normalized second moment of an $n$-dimensional lattice.

Consider the following problem to motivate the general result. Suppose a quantization system is needed with input $X_1$ and outputs $(X_2, \ldots, X_M)$ such that the quantization errors $X_i - X_1, i = 2, \ldots, M$, are correlated with each other in a certain predetermined way, but are uncorrelated with $X_1$. Seemingly, $M-1$ quantizers may be used, each with $X_1$ as the input and $X_i$ as the output for some $i, i = 2, \ldots, M$. By property 1) of ECDQ, if dithers are introduced, the quantization errors are independent of the input of the quantizer. However, it is difficult to make the quantization errors of these $M-1$ quantizers correlated in the desired manner. One may expect it to be possible to correlate the quantization errors by simply correlating the dithers of different quantizers, but this turns out to be not true as pointed out in [32]. Next, we present a solution to this problem by exploiting the relationship between the Gram–Schmidt orthogonalization and sequential (dithered) quantization.

### B. Gram–Schmidt Orthogonalization and Sequential Dithered Quantization

In order to facilitate the treatment, the aforementioned problem is reformulated in an equivalent form: Given $X_1^M$ with an arbitrary covariance matrix, construct a quantization system with $\hat{X}_1$ as the input and $(\hat{X}_2, \ldots, \hat{X}_M)$ as the outputs such that the covariance matrices of $X_1^M$ and $\hat{X}_1^M$ are the same.

Let $\mathcal{H}_s$ denote the set of all finite-variance, zero-mean, real scalar random variables. It is well known [41], [42] that $\mathcal{H}_s$ becomes a Hilbert space under the inner product mapping $\langle X, Y\rangle = \mathbb{E}(XY) : \mathcal{H}_s \times \mathcal{H}_s \to \mathcal{R}$; the norm induced by this inner product is thus $\|X\|^2 = \mathbb{E}X^2$.

For $X_1^M = (X_1, \ldots, X_M)^T$ with $X_i \in \mathcal{H}_s, i = 1, \ldots, M$, the Gram–Schmidt orthogonalization can be used to construct an orthogonal basis $B_1^M = (B_1, \ldots, B_M)^T$ for $X_1^M$. Let $K_{X_1^m}$ denote the covariance matrix of $(X_1, \ldots, X_m)^T$ and let $K_{X_m X_1^{m-1}} = \mathbb{E}[X_m (X_1, \ldots, X_{m-1})^T]$, then

$$B_1 = X_1 \tag{1}$$

$$B_i = X_i - K_{i-1} X_1^{i-1}, \quad i = 2, \ldots, M. \tag{2}$$

Here $K_{i-1} \in \mathcal{R}^{1\times(i-1)}$ is a row vector satisfying $K_{i-1} K_{X_1^{i-1}} = K_{X_i X_1^{i-1}}$. When $K_{X_1^{i-1}}$ is invertible, $K_{i-1}$ is uniquely given by $K_{X_i X_1^{i-1}} K_{X_1^{i-1}}^{-1}$. The product $K_{i-1} X_1^{i-1}$ is the *linear minimum mean-square error (MMSE) estimator* of $X_i$ given $X_1^{i-1}$, and $\mathbb{E}B_i^2$ is its corresponding *linear MMSE estimation error*. $B_1^M$

is sometimes referred to as the *innovation process* [41]. In the special case in which $X_1^M$ are jointly Gaussian

$$K_{i-1}X_1^{i-1}=\mathbb{E}\left(X_i\,\middle|\,X_1^{i-1}\right)=\sum_{j=1}^{i-1}\mathbb{E}(X_i\,|\,B_j),\quad i=2,\ldots,M$$

and $B_1^M$ are zero-mean, independent, and jointly Gaussian. Moreover, since $X_1^i$ is a deterministic function of $B_1^i$, it follows that $B_{i+1}^M$ is independent of $X_1^i$, for $i=1,\ldots,M-1$.

We now show that one can construct a sequential quantization system with $X_1$ as the input to generate a zero-mean random vector $\tilde{X}_1^M=(\tilde{X}_1,\tilde{X}_2,\ldots,\tilde{X}_M)^T$ whose covariance matrix is also $K_{X_1^M}$. Let $X_1^M$ be a zero-mean random vector with covariance matrix $K_{X_1^M}$. By (1) and (2), it is true that

$$X_1=B_1 \tag{3}$$
$$X_i=K_{i-1}X_1^{i-1}+B_i,\qquad i=2,\ldots,M. \tag{4}$$

Assume that $B_i\neq 0$ for $i=2,\ldots,M$. Let $Q_{i,1}(\cdot)$ be a scalar lattice quantizer with step size

$$\Delta_i=\sqrt{12\mathbb{E}B_{i+1}^2},\qquad i=1,2,\ldots,M-1.$$

Let the dither $Z_i\sim\mathcal{U}(-\Delta_i/2,\Delta_i/2)$ be a random variable uniformly distributed over the basic cell of $Q_{i,1}, i=1,2,\ldots,M-1$. Note: the second subscript $n$ of $Q_{i,n}$ denotes the dimension of the lattice quantizer; in this case $n=1$, so it is a scalar quantizer.

Let $(X_1,Z_1,\ldots,Z_{M-1})$ be independent. By the property of ECDQ, we can construct $\tilde{X}_i$ as

$$\tilde{X}_1=X_1 \tag{5}$$
$$\begin{aligned}\tilde{X}_i&=Q_{i-1,1}\left(K_{i-1}\tilde{X}_1^{i-1}+Z_{i-1}\right)-Z_{i-1}\\&=K_{i-1}\tilde{X}_1^{i-1}+N_i,\qquad i=2,\ldots,M\end{aligned} \tag{6}$$

where $N_i\sim\mathcal{U}(\Delta_i/2,\Delta_i/2)$ with $\mathbb{E}N_i^2=\mathbb{E}B_{i+1}^2, i=1,\ldots,M-1$, and $(X_1,N_1,\ldots,N_M)$ are independent. By comparing (3), (4) and (5), (6), it is straightforward to verify that $X_1^M$ and $\tilde{X}_1^M$ have the same covariance matrix. Note that if $B_i=0$ for some $i$, then $\tilde{X}_i=K_{i-1}\tilde{X}_1^{i-1}$ and therefore no quantization operation is needed to generate $\tilde{X}_i$ from $\tilde{X}_1^{i-1}$.

The generalization of the correspondence between the Gram–Schmidt orthogonalization and the sequential (dithered) quantization to the vector case is straightforward; see Appendix I.

## III. Successive Quantization and Quantization Splitting

In the last section, it has been shown that Gram–Schmidt orthogonalization can be used to form any quantization noise correlation structure. However, the rate constraints do not come into the picture, and the quantization order is chosen in an arbitrary fashion. These factors have to be determined in the specific context. In this section, we provide an information-theoretic analysis of the EGC region and show it can be decomposed into a simplified-EGC (SEGC) region and a superimposed refinement operation. Furthermore, any rate pair in the SEGC region can be achieved via a successive quantization scheme along with quantization splitting. In this way, we associate each rate pair in the SEGC region with a natural quantization order. We will see in Section IV that in the Gaussian MD case, the Gram–Schmidt orthogonalization procedure generates sufficient statistics along this quantization order and therefore has the property of rate preservation.

We focus on discrete memoryless source and bounded distortion measure. The results can be generalized to the quadratic Gaussian case, using the technique in [61].

### A. An Information-Theoretic Analysis of the EGC Region

Rewrite $\mathcal{R}(U_1,U_2,U_3)$ in the following form:

$$\begin{aligned}\mathcal{R}(U_1,U_2,U_3)=\{(R_1,R_2):\,&R_1+R_2\geq I(X;U_1,U_2)\\&+I(U_1;U_2)+I(X;U_3\,|\,U_1,U_2),R_i\geq I(X;U_i),i=1,2\}.\end{aligned}$$

Without loss of generality, assume that $X\to U_3\to(U_1,U_2)$ form a Markov chain since otherwise $U_3$ can be replaced by $\tilde{U}_3=(U_1,U_2,U_3)$ without affecting the rate and distortion constraints. Therefore, $U_3$ can be viewed as a fine description of $X$ and $(U_1,U_2)$ as coarse descriptions of $X$. The term $I(X;U_3\,|\,U_1,U_2)$ is the rate used for the superimposed refinement from the pair of coarse descriptions $(U_1,U_2)$ to the fine description $U_3$; in general, this refinement rate is split between the two channels. Since description refinement schemes have been studied extensively in the multiresolution or layered source coding scenario and are well understood, this operation can be separated from other parts of the EGC scheme.

*Definition 3.1 (SEGC Region):* For random variables $U_1$ and $U_2$ jointly distributed with the generic source variable $X$ via conditional distribution $p(u_1,u_2\,|\,x)$, let

$$\begin{aligned}\mathcal{R}(U_1,U_2)=\{(R_1,R_2):\,&R_1+R_2\geq I(X;U_1,U_2)\\&+I(U_1;U_2),R_i\geq I(X;U_i),i=1,2\}.\end{aligned}$$

Let

$$\begin{aligned}&\mathcal{Q}(U_1,U_2)\\&\quad=\{(R_1,R_2,D_1,D_2,D_3):(R_1,R_2)\in\mathcal{R}(U_1,U_2),\exists\hat{X}_1\\&\quad=g_1(U_1),\hat{X}_2=g_2(U_2),\hat{X}_3=g_3(U_1,U_2)\\&\qquad\qquad\text{with }\mathbb{E}d(X,\hat{X}_i)\leq D_i,i=1,2,3\}.\end{aligned}$$

The SEGC region is defined as

$$\mathcal{Q}_{\mathrm{SEGC}}=\mathrm{conv}\left(\bigcup_{p(u_1,u_2\,|\,x)}\mathcal{Q}(U_1,U_2)\right).$$

The SEGC region first appeared in [1] and was attributed to El Gamal and Cover. It was shown in [7] that $\mathcal{Q}_{\mathrm{SEGC}}\subseteq\mathcal{Q}_{\mathrm{EGC}}$.

Using the identity

$$I(A;BC)=I(A;B)+I(A;C)+I(B;C\,|\,A)-I(B;C)$$

$\mathcal{R}(U_1,U_2)$ can be written as

$$\begin{aligned}\mathcal{R}(U_1,U_2)=\{(R_1,R_2):\,&R_1+R_2\geq I(X;U_1)+I(X;U_2)\\&+I(U_1;U_2\,|\,X),R_i\geq I(X;U_i),i=1,2\}.\end{aligned}$$

The typical shape of $\mathcal{R}(U_1,U_2)$ is shown in Fig. 3.

It is noteworthy that $\mathcal{R}(U_1,U_2)$ resembles Marton's achievable region [43] for a two-user broadcast channel. This is not

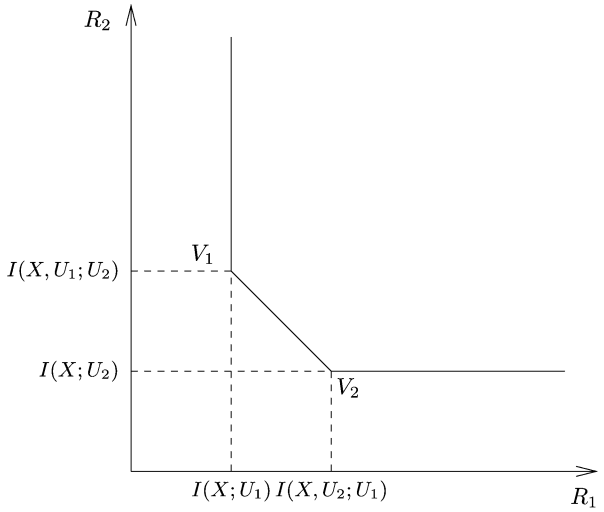


Fig. 3. The shape of $\mathcal{R}(U_1, U_2)$.

surprising since the proof of the EGC theorem relies heavily on the results in [44] which were originally for a simplified proof of Marton's coding theorem for the discrete memoryless broadcast channel.[2] Since the corner points of Marton's region can be achieved via a relatively simple coding scheme due to Gel'fand and Pinsker [46], which for the Gaussian case becomes Costa's dirty paper coding [47], it is natural to conjecture that simple quantization schemes may exist for the corner points of $\mathcal{R}(U_1, U_2)$. This conjecture turns out to be correct as will be shown later.

Since $I(U_1; U_2 \,|\, X) \geq 0$, the sum-rate constraint in $\mathcal{R}(U_1, U_2)$ is always effective. Thus

$$\{(R_1, R_2) : R_1 + R_2 = I(X;U_1) + I(X;U_2) + I(U_1;U_2 \,|\, X), R_i \geq I(X;U_i), i = 1,2\}$$

will be called *the dominant face* of $\mathcal{R}(U_1, U_2)$. Any rate pair inside $\mathcal{R}(U_1, U_2)$ is inferior to some rate pair on the dominant face in terms of compression efficiency. Hence, in searching for the optimal scheme, attention can be restricted to rate pairs on the dominant face without loss of generality. The dominant face of $\mathcal{R}(U_1, U_2)$ has two vertices $V_1$ and $V_2$. Let $(R_1(V_i), R_2(V_i))$ denote the coordinates of vertex $V_i, i = 1, 2$, then

$$V_1 : R_1(V_1) = I(X;U_1), R_2(V_1) = I(X, U_1; U_2)$$
$$V_2 : R_1(V_2) = I(X, U_2; U_1), R_2(V_2) = I(X;U_2).$$

The expressions of these two vertices directly lead to the following successive quantization scheme. By symmetry, we shall only consider $V_1$.

### B. Successive Quantization Scheme

For vertex $V_1$, the successive quantization coding scheme suffices, which can also be understood as a special case of the coding scheme in [5]. It is outlined below because the quantization splitting scheme in the next subsection will utilize this idea.

1) Codebook Generation: Encoder 1 independently generates $2^{n[I(X;U_1)+\epsilon_1]}$ codewords $\boldsymbol{U_1}(j) \sim \prod p(u_1)$. Encoder 2 independently generates $2^{n[I(X,U_1;U_2)+\epsilon_2]}$ codewords $\boldsymbol{U_2}(k) \sim \prod p(u_2)$.
2) Encoding Procedure: Given $\boldsymbol{X}$, encoder 1 finds the codeword $\boldsymbol{U}_1(j^*)$ which is strongly typical with $\boldsymbol{X}$. Then encoder 2 finds the codeword $\boldsymbol{U}_2(k^*)$ which is strongly typical with $\boldsymbol{X}$ and $\boldsymbol{U}_1(j^*)$. Index $j^*$ is transmitted through channel 1 and index $k^*$ is transmitted through channel 2.
3) Reconstruction: Decoder 1 reconstructs $\hat{\boldsymbol{X}}_1$ with $\hat{X}_1(t) = g_1(U_1(j^*, t))$. Decoder 2 reconstructs $\hat{\boldsymbol{X}}_2$ with $\hat{X}_2(t) = g_2(U_2(k^*, t))$. Decoder 3 reconstructs $\hat{\boldsymbol{X}}_3$ with $\hat{X}_3(t) = g_3(U_1(j^*, t), U_2(k^*, t))$. Here, $U_1(j^*, t)$ and $U_2(k^*, t)$ are the $t$th entries of $\boldsymbol{U}_1(j^*)$ and $\boldsymbol{U}_2(k^*)$, respectively, $t = 1, 2, \ldots, n$.

For this scheme, encoder 1 does the encoding first and then encoder 2 follows. The main complexity of this scheme resides in encoder 2, since it needs to construct a codebook that covers the $(\boldsymbol{X}, \boldsymbol{U}_1)$-space instead of just the $\boldsymbol{X}$-space. Observe that, if a function $f(X, U_1) = V$ can be found such that $V$ is a *sufficient statistic* for estimation $U_2$ from $(X, U_1)$, i.e., $(X, U_1) \to V \to U_2$ form a Markov chain,[3] then

$$I(X, U_1; U_2) = I(V; U_2).$$

The importance of this observation is that encoder 2 then only needs to construct a codebook that covers the $\boldsymbol{V}$-space instead of the $(\boldsymbol{X}, \boldsymbol{U}_1)$-space. This is because the Markov lemma [48] implies that if $\boldsymbol{U}_2$ is jointly typical with $\boldsymbol{V}$, then $\boldsymbol{U}_2$ is jointly typical with $(\boldsymbol{X}, \boldsymbol{U}_1)$ with high probability. This observation turns out to be crucial for the quadratic Gaussian case.

We point out that the successive coding structure associated with the corner points of $\mathcal{R}(U_1, U_2)$ is not a special case in network information theory. Besides its resemblance to the successive Gel'fand–Pinsker coding structure associated with the corner points of the Marton's region previously mentioned, other noteworthy examples include the successive decoding structure associated with the corner points of the Slepian–Wolf region [49] (and more generally, the Berger–Tung region [48], [50], [51]) and the corner points of the capacity region of the memoryless multiaccess channel [52], [53].

### C. Successive Quantization Scheme With Quantization Splitting

To achieve an arbitrary rate pair on the dominant face of $\mathcal{R}(U_1, U_2)$, a straightforward method is to timeshare the coding schemes that achieve the two vertices. However, such a scheme generally requires four different quantizers, which is not desirable. Instead, the scheme based on quantization splitting introduced below needs only three quantizers. Before presenting this coding scheme, we shall first prove the following theorem.

[2]The resemblance between $\mathcal{R}(U_1, U_2)$ and Marton's region has a deeper reason. See [45] for a detailed discussion.

[3]Such a function $f(\cdot, \cdot)$ always exists provided $|\mathcal{V}| \geq |\mathcal{X}||\mathcal{U}_1|$.

*Theorem 3.1:* For any rate pair $(R_1, R_2)$ on the dominant face of $\mathcal{R}(U_1, U_2)$, there exists a random variable $U_2'$ with $(X, U_1) \to U_2 \to U_2'$ forming a Markov chain such that

$$R_1 = I(X, U_2'; U_1)$$
$$R_2 = I(X; U_2') + I(X, U_1; U_2 \,|\, U_2').$$

The apparent symmetric claim by switching the role of $U_1$ and $U_2$ also holds.

Before proceeding to prove this theorem, we make the following remarks.

- Since $(X, U_1) \to U_2 \to U_2'$ form a Markov chain, if $U_2'$ is independent of $U_2$, then it must be independent of $(X, U_1, U_2)$ altogether.[4] Then in this case
$$R_1 = I(X; U_1), \quad R_2 = I(X, U_1; U_2)$$
which are the coordinates of $V_1$.
- At the other extreme, letting $U_2'$ be $U_2$ gives
$$R_1 = I(X, U_2; U_1), \quad R_2 = I(X; U_2)$$
which are the coordinates of $V_2$.

*Proof:* Since the desired $U_2'$ has the property that $(X, U_1) \to U_2 \to U_2'$ form a Markov chain, we only need to specify the distribution of $U_2'$ conditioned on $U_2$.

Construct a class of transition probabilities[5] $p_\epsilon(u_2' \,|\, u_2)$ parameterized by $\epsilon$ such that $I(U; U_2')$ varies continuously from 0 to $H(U_2)$ as $\epsilon$ changes from 0 to 1. It remains to show that

$$R_1 + R_2 = I(X; U_1) + I(X; U_2) + I(U_1; U_2 \,|\, X).$$

This is indeed true since

$$\begin{aligned} R_1 + R_2 &= I(X, U_2'; U_1) + I(X; U_2') + I(X, U_1; U_2 \,|\, U_2') \\ &= I(X, U_2'; U_1) + I(X; U_2') + I(X; U_2 \,|\, U_2') \\ &\quad + I(U_1; U_2 \,|\, X, U_2') \\ &= I(X, U_2, U_2'; U_1) + I(X; U_2, U_2') \\ &\overset{(a)}{=} I(X, U_2; U_1) + I(X; U_2) \\ &= I(X; U_1) + I(X; U_2) + I(U_1; U_2 \,|\, X), \end{aligned}$$

where (a) follows from the fact that $(X, U_1) \to U_2 \to U_2'$ form a Markov chain, and this completes the proof. □

The successive quantization scheme with quantization splitting is outlined as follows.

1) Codebook Generation: Encoder 1 independently generates about $2^{n[I(X,U_2';U_1)+\epsilon_1']}$ codewords $\boldsymbol{U_1}(i) \sim \prod p(u_1)$. Encoder 2 independently generates about $2^{n[I(X;U_2')+\epsilon_2']}$ codewords $\bar{\boldsymbol{U}}_2'(j) \sim \prod p(u_2')$. For each codeword $\bar{\boldsymbol{U}}_2'(j)$, encoder 2 independently generates $2^{n[I(X,U_1;U_2 \,|\, U_2')+\epsilon_3']}$ codewords $\boldsymbol{U}_2(j,k) \sim \prod_t p(u_2 \,|\, U_2'(j,t))$. Here $U_2'(j,t)$ is the $t$th entry of $\boldsymbol{U}_2'(j)$
2) Encoding Procedure: Given $\boldsymbol{X}$, encoder 2 finds the codeword $\boldsymbol{U}_2'(j^*)$ such that $\boldsymbol{U}_2'(j^*)$ is strongly typical with $\boldsymbol{X}$. Then encoder 1 finds the codeword $\boldsymbol{U}_1(i^*)$ such that $\boldsymbol{U}_1(i^*)$ is strongly typical with $\boldsymbol{X}$ and $\boldsymbol{U}_2'(j^*)$. Finally, encoder 2 finds the codeword $\bar{\boldsymbol{U}}_2(j^*, k^*)$ such that $\boldsymbol{U}_2(j^*, k^*)$ is strongly typical with $\boldsymbol{X}, \boldsymbol{U}_1(i^*)$ and $\boldsymbol{U}_2'(j^*)$. Index $i^*$ is transmitted through channel 1. Indices $j^*$ and $k^*$ are transmitted through channel 2.
3) Reconstruction: Decoder 1 reconstructs $\hat{\boldsymbol{X}}_1$ with $\hat{X}_1(t) = g_1(U_1(i^*, t))$. Decoder 2 reconstructs $\hat{\boldsymbol{X}}_2$ with $\hat{X}_2(t) = g_2(U_2(j^*, k^*, t))$. Decoder 3 reconstructs $\hat{\boldsymbol{X}}_3$ with $\hat{X}_3(t) = g_3(U_1(i^*, t), U_2(j^*, k^*, t))$. Here $U_1(i^*, t)$ is the $t$th entry of $\boldsymbol{U}_1(i^*)$ and $U_2(j^*, k^*, t)$ is the $t$th entry of $\boldsymbol{U}_2(j^*, k^*), t = 1, 2, \ldots, n$.

This approach is a natural generalization of the successive quantization scheme for the vertices of $\mathcal{R}(U_1, U_2)$. $U_2'$ can be viewed as a coarse description of $X$ and $U_2$ as a fine description of $X$. The idea of introducing an auxiliary coarse description to convert a joint coding scheme to a successive coding scheme has been widely used in the distributed source coding problems [54]–[56]. Similar ideas have also found application in multiaccess communications [57]–[60].

## IV. The Gaussian MD Region

In this section, we apply the general results in the preceding section to the quadratic Gaussian case. The Gaussian MD rate–distortion region is first analyzed to show that $\mathcal{Q}_{\mathrm{EGC}} = \mathcal{Q}_{\mathrm{SEGC}}$ in this case. Then, by incorporating the Gram–Schmidt orthogonalization with successive quantization and quantization splitting, a coding scheme that achieves the whole Gaussian MD region is presented. The rates and distortions are derived explicitly for the Gaussian source, and they will be used to bound the corresponding rates and distortions in the next section when ECDQs are utilized for more general sources. To facilitate reading, the linear estimation coefficients and statistics of the innovation process are collected in Appendix II.

### A. An Analysis of the Gaussian MD Region

Let $\{X^G(t)\}_{t=1}^{\infty}$ be an i.i.d. Gaussian process with $X^G(t) \sim \mathcal{N}(0, \sigma_X^2)$ for all $t$. Let $d(\cdot, \cdot)$ be the squared error distortion measure. For the quadratic Gaussian case, the MD rate–distortion region was characterized in [3], [5], [62]. Namely, $(R_1, R_2, D_1, D_2, D_3) \in \mathcal{Q}$ if and only if

$$R_i \geq \frac{1}{2} \log \frac{\sigma_X^2}{D_i}, \qquad i = 1, 2$$
$$R_1 + R_2 \geq \frac{1}{2} \log \frac{\sigma_X^2}{D_3} + \frac{1}{2} \log \psi(D_1, D_2, D_3)$$

where the expression of $\psi(D_1, D_2, D_3)$ is given at the top of the following page.

The case $D_3 < D_1 + D_2 - \sigma_X^2$ and the case $D_3 > (1/D_1 + 1/D_2 - 1/\sigma_X^2)^{-1}$ are degenerate. It is easy to verify that for any $(R_1, R_2, D_1, D_2, D_3) \in \mathcal{Q}$ with $D_3 < D_1 + D_2 - \sigma_X^2$, there exist $D_1^* \leq D_1, D_2^* \leq D_2$ such that $(R_1, R_2, D_1^*, D_2^*, D_3) \in \mathcal{Q}$ and $D_3 = D_1^* + D_2^* - \sigma_X^2$. Similarly, for any $(R_1, R_2, D_1, D_2, D_3) \in \mathcal{Q}$ with $D_3 > (1/D_1 + 1/D_2 - 1/\sigma_X^2)^{-1}$, there exist $D_3^* = (1/D_1 + 1/D_2 - 1/\sigma_X^2)^{-1} < D_3$ such that $(R_1, R_2, D_1, D_2, D_3^*) \in \mathcal{Q}$. Hence, without loss of optimality, we shall only consider the

[4] This is because $I(X, U_1, U_2; U_2') = I(U_2; U_2') = 0$.

[5] There are many ways to construct such a class of transition probabilities. For example, we can let $p_0(u_2' \,|\, u_2) = p(u_2'), p_1(u_2' \,|\, u_2) = \delta(u_2, u_2')$, and set $p_\epsilon(u_2' \,|\, u_2) = (1 - \epsilon) p_0(u_2' \,|\, u_2) + \epsilon p_1(u_2' \,|\, u_2)$. Here $\delta(u_2, u_2') = 1$ if $u_2 = u_2'$ and $= 0$ otherwise.

$$\psi(D_1,D_2,D_3)=\begin{cases}1, & D_3<D_1+D_2-\sigma_X^2\\ \dfrac{\sigma_X^2 D_3}{D_1D_2}, & D_3>\left(\frac{1}{D_1}+\frac{1}{D_2}-\frac{1}{\sigma_X^2}\right)^{-1}\\ \dfrac{\left(\sigma_X^2-D_3\right)^2}{\left(\sigma_X^2-D_3\right)^2-\left[\sqrt{\left(\sigma_X^2-D_1\right)\left(\sigma_X^2-D_2\right)}-\sqrt{(D_1-D_3)(D_2-D_3)}\right]^2}, & \text{o.w.}\end{cases}$$

case $(1/D_1+1/D_2-1/\sigma_X^2)^{-1}\geq D_3\geq D_1+D_2-\sigma_X^2$, for which $D_1,D_2$, and $D_3$ all are effective.

Following the approach in [5], let

$$U_1=X^G+T_0+T_1 \tag{7}$$
$$U_2=X^G+T_0+T_2 \tag{8}$$

where $(T_1,T_2),T_0,X$ are zero-mean, jointly Gaussian and independent, and $\mathbb{E}(T_1T_2)=-\sigma_{T_1}\sigma_{T_2}$. Let $\hat{X}_i^G=\mathbb{E}(X^G\,|\,U_i)=\alpha_iU_i$ $(i=1,2)$, and $\hat{X}_3^G=\mathbb{E}(X^G\,|\,U_1,U_2)=\beta_1U_1+\beta_2U_2$ (see Appendix II for $\alpha_i,\beta_i,i=1,2$). In this case, the achievable rates and distortions can be computed analytically.

Set $\mathbb{E}(X^G-\hat{X}_i^G)^2=D_i,i=1,2,3$; then

$$\sigma_{T_0}^2=\frac{D_3\sigma_X^2}{\sigma_X^2-D_3} \tag{9}$$
$$\sigma_{T_i}^2=\frac{D_i\sigma_X^2}{\sigma_X^2-D_i}-\frac{D_3\sigma_X^2}{\sigma_X^2-D_3},\qquad i=1,2. \tag{10}$$

After further computation, it can be verified that

$$\begin{aligned}&\mathcal{R}^G(U_1,U_2)\\&\quad\triangleq\{(R_1,R_2):R_1+R_2\geq I(X^G;U_1)+I(X^G;U_2)\\&\qquad+I(U_1;U_2\,|\,X^G),R_i\geq I(X^G;U_i),i=1,2\}\\&\quad=\left\{(R_1,R_2):R_1+R_2\geq\frac{1}{2}\log\frac{\sigma_X^2}{D_3}\right.\\&\qquad\left.+\frac{1}{2}\log\psi(D_1,D_2,D_3),R_i\geq\frac{1}{2}\log\frac{\sigma_X^2}{D_i},i=1,2\right\}.\end{aligned} \tag{11}$$

Hence, for the quadratic Gaussian case

$$\mathcal{Q}=\mathcal{Q}_{\mathrm{EGC}}=\mathcal{Q}_{\mathrm{SEGC}}$$

and there is no need to introduce $U_3$ (more precisely, $U_3$ can be written as a deterministic function of $U_1$ and $U_2$).

We can see that for fixed $(D_1,D_2,D_3)$, $\sigma_{T_i}^2$ $(i=0,1,2)$ are uniquely determined by (9) and (10), and consequently $\mathcal{R}^G(U_1,U_2)$ is given by (11). Since only the optimal MD coding scheme is of interest, the sum-rate $R_1+R_2$ should be minimized with respect to the distortion constraints $(D_1,D_2,D_3)$, i.e., $(R_1,R_2)$ must be on the dominant face of $\mathcal{R}^G(U_1,U_2)$. Thus, for fixed $(D_1,D_2,D_3)$

$$R_1+R_2=\frac{1}{2}\log\frac{\sigma_X^2}{D_3}+\frac{1}{2}\log\psi(D_1,D_2,D_3). \tag{12}$$

### *B. Successive Quantization for Gaussian Source*

If we view $U_1,U_2$ as two different quantizations of $X^G$ and let $U_1-X^G$ and $U_2-X^G$ be their corresponding quantization errors, then (13) follows, as shown at the bottom of the page, where the inequality is strict unless $D_3=(1/D_1+1/D_2-1/\sigma_X^2)^{-1}$. The existence of negative correlation between the quantization errors is the main difficulty in designing optimal MD quantization schemes. To circumvent it, $U_1$ and $U_2$ can be represented in a different form by using the Gram–Schmidt orthogonalization. It yields that

$$\begin{aligned}B_1&=X^G\\B_2&=U_1-E(U_1\,|\,X^G)=U_1-X^G\\B_3&=U_2-E(U_2\,|\,X^G,U_1)=U_2-a_1X^G-a_2U_1\end{aligned}$$

and expressions for $a_1,a_2,\mathbb{E}B_2^2$, and $\mathbb{E}B_3^2$ are given in Appendix II.

Now consider the quantization scheme for vertex $V_1^G$ of $\mathcal{R}^G(U_1,U_2)$. $R_1(V_1^G)$ is given by

$$R_1(V_1^G)=I(X^G;U_1)=I(X^G;X^G+B_2). \tag{14}$$

Since $U_2=\mathbb{E}(U_2\,|\,X^G,U_1)+B_3$, where $B_3$ is independent of $(X^G,U_1)$, it follows that $(X^G,U_1)\to\mathbb{E}(U_2\,|\,X^G,U_1)\to U_2$ form a Markov chain. Clearly, $\mathbb{E}(U_2\,|\,X^G,U_1)\to(X^G,U_1)\to U_2$ also form a Markov chain since $\mathbb{E}(U_2\,|\,X^G,U_1)$ is a deterministic function of $(X^G,U_1)$. These two Markov relationships imply that $I(X^G,U_1;U_2)=I(\mathbb{E}(U_2\,|\,X^G,U_1);U_2)$, and thus

$$\begin{aligned}R_2(V_1^G)&=I(X^G,U_1;U_2)\\&=I(\mathbb{E}(U_2\,|\,X^G,U_1);U_2)\\&=I(a_1X^G+a_2U_1;a_1X^G+a_2U_1+B_3).\end{aligned} \tag{15}$$

$$\begin{aligned}\mathbb{E}[(U_1-X^G)(U_2-X^G)]&=\mathbb{E}[(T_0+T_1)(T_0+T_2)]\\&=\sigma_{T_0}^2-\sigma_{T_1}\sigma_{T_2}\\&=\frac{D_3\sigma_X^2}{\sigma_X^2-D_3}-\sqrt{\left(\frac{D_1\sigma_X^2}{\sigma_X^2-D_1}-\frac{D_3\sigma_X^2}{\sigma_X^2-D_3}\right)\left(\frac{D_2\sigma_X^2}{\sigma_X^2-D_2}-\frac{D_3\sigma_X^2}{\sigma_X^2-D_3}\right)}\\&\leq 0\end{aligned} \tag{13}$$

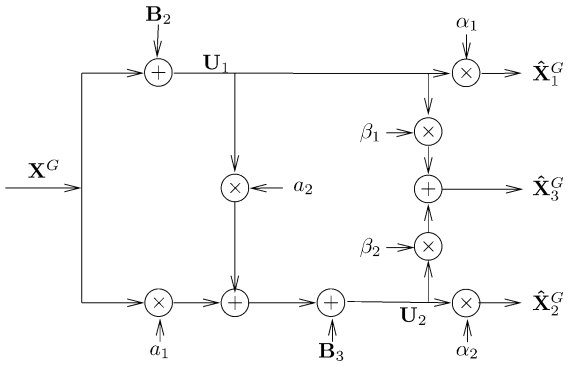


Fig. 4. MD quantization scheme for $V_1^G$.

Although the preceding expressions are all of single-letter type, it does not mean that symbol-by-symbol operations can achieve the optimal bound. Instead, when interpreting these information-theoretic results, one should think of a system that operates on long blocks.

Intuitively speaking (see Fig. 4), (14) and (15) suggest the following quantization scheme. Encoder 1 quantizes at rate $R_1(V_1^G)$ with input $\boldsymbol{X}^G$ and output $\boldsymbol{U}_1$. The quantization error is $\boldsymbol{B}_2 = \boldsymbol{U}_1 - \boldsymbol{X}^G$, which is a zero-mean Gaussian vector with covariance matrix $\mathbb{E}B_2^2 I_n$. Encoder 2 quantizes at rate $R_2(V_1^G)$ with input $a_1\boldsymbol{X}^G + a_2\boldsymbol{U}_1$ and output $\boldsymbol{U}_2$. The quantization error $\boldsymbol{B}_3 = \boldsymbol{U}_2 - a_1\boldsymbol{X}^G - a_2\boldsymbol{U}_1$ is a zero-mean Gaussian vector with covariance matrix $\mathbb{E}B_3^2 I_n$.

*Remarks:*

1) $\boldsymbol{U}_1$ (or $\boldsymbol{U}_2$) is not a deterministic function of $\boldsymbol{X}^G$ (or $a_1\boldsymbol{X}^G + a_2\boldsymbol{U}_1$), and for classical quantizers the quantization noise is generally not Gaussian. Thus, strictly speaking, the "noise-adding" components in Fig. 4 are not quantizers in the traditional sense. We nevertheless refer to them as quantizers[6] in this section for simplicity.
2) From Fig. 4, it is obvious that the MD quantization for $V_1^G$ is essentially the Gram–Schmidt orthogonalization of $(\boldsymbol{X}^G, \boldsymbol{U}_1, \boldsymbol{U}_2)$. As previously shown in Section II, the Gram–Schmidt orthogonalization can be simulated by sequential (dithered) quantization. The formal description and analysis of this quantization scheme in the context of MDs for general sources will be given in Section V.

### C. Successive Quantization With Quantization Splitting for Gaussian Source

Now we study the quantization scheme for an arbitrary rate pair $(R_1^G, R_2^G)$ on the dominant face of $\mathcal{R}^G(U_1, U_2)$. Note that since the rate sum $R_1^G + R_2^G$ is given by (12), $(R_1^G, R_2^G)$ only has one degree of freedom.

Let $U_2' = X^G + T_0 + T_2 + T_3$, where $T_3$ is zero-mean, Gaussian, and independent of $(X^G, T_0, T_1, T_2)$. It is easy to verify that $(X^G, U_1) \to U_2 \to U_2'$ form a Markov chain. Applying the Gram–Schmidt orthogonalization algorithm to $(X^G, U_2', U_1)$, we have (with expressions for $b_1, b_2, \mathbb{E}\tilde{B}_2^2$ and $\mathbb{E}\tilde{B}_3^2$ in Appendix II)

$$\begin{aligned}
\tilde{B}_1 &= X \\
\tilde{B}_2 &= U_2' - \mathbb{E}(U_2' \,|\, X^G) = U_2' - X^G \\
\tilde{B}_3 &= U_1 - \mathbb{E}(U_1 \,|\, X^G, U_2') = U_1 - b_1 X^G - b_2 U_2'.
\end{aligned}$$

Since $U_1 = \mathbb{E}(U_1 \,|\, X^G, U_2') + \tilde{B}_3$, where $\tilde{B}_3$ is independent of $(X^G, U_2')$, it follows that

$$(X^G, U_2') \to \mathbb{E}(U_1 \,|\, X^G, U_2') \to U_1$$

form a Markov chain. Clearly

$$\mathbb{E}(U_1 \,|\, X^G, U_2') \to (X^G, U_2') \to U_1$$

also form a Markov chain because $\mathbb{E}(U_1 \,|\, X^G, U_2')$ is determined by $(X^G, U_2')$. Thus, we have

$$I(X^G, U_2'; U_1) = I(\mathbb{E}(U_1 \,|\, X^G, U_2'); U_1)$$

and this gives

$$R_1^G = \frac{1}{2}\log \frac{\left(\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_1}^2\right)\left(\sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2\right)}{\sigma_{T_0}^2(\sigma_{T_1} + \sigma_{T_2})^2 + \sigma_{T_3}^2\left(\sigma_{T_0}^2 + \sigma_{T_1}^2\right)}.$$

Hence, $\sigma_{T_3}^2$ is uniquely determined by (16) at the bottom of the page. $R_2^G$ can also be readily computed in terms of $\sigma_{T_i}^2, i = 0, 1, 2, 3$ and $\sigma_x^2$, and it can be shown straightforwardly that as $\sigma_{T_3}^2$ varies from 0 to $\infty$, all the rate pairs on the dominant face of $\mathcal{R}^G(U_1, U_2)$ can be achieved.

For a specific rate pair $(R_1^G, R_2^G)$, we have

$$\begin{aligned}
R_1^G &= I(X^G, U_2'; U_1) = I(\mathbb{E}(U_1 \,|\, X^G, U_2'); U_1) \\
&= I(b_1 X^G + b_2 U_2'; b_1 X^G + b_2 U_2' + \tilde{B}_3) \\
R_2^G &= I(X^G; U_2') + I(X^G, U_1; U_2 \,|\, U_2') \\
&= I(X^G; X^G + \tilde{B}_2) + I(X^G, U_1; U_2 \,|\, U_2'). \qquad (17)
\end{aligned}$$

To remove the conditioning term $U_2'$ in $I(X^G, U_1; U_2 \,|\, U_2')$, we apply the Gram–Schmidt procedure to $(U_2', X^G, U_1, U_2)$. It yields (with expressions for $b_i, i = 3, \ldots, 8$ and $\mathbb{E}\bar{B}_j^2, j = 2, 3, 4$ in Appendix II)

$$\begin{aligned}
\bar{B}_1 &= U_2' \\
\bar{B}_2 &= X - \mathbb{E}(X^G \,|\, \bar{B}_1) = X^G - b_3\bar{B}_1 \\
\bar{B}_3 &= U_1 - \mathbb{E}(U_1 \,|\, \bar{B}_1) - \mathbb{E}(U_1 \,|\, \bar{B}_2) = U_1 - b_4\bar{B}_1 - b_5\bar{B}_2 \\
\bar{B}_4 &= U_2 - \sum_{i=1}^{3} \mathbb{E}(U_2 \,|\, \bar{B}_i) = U_2 - b_6\bar{B}_1 - b_7\bar{B}_2 - b_8\bar{B}_3.
\end{aligned}$$

[6]This slight abuse of the word "quantizer" can be justified in the context of ECDQ (as we will show in the next section) since the quantization noise of the optimal lattice quantizer is indeed asymptotically Gaussian; furthermore, the quantization noise is indeed independent of the input for ECDQ [38].

$$\sigma_{T_3}^2 = \frac{\sigma_{T_0}^2(\sigma_{T_1} + \sigma_{T_2})^2 2^{2R_1^G} - \left(\sigma_{T_0}^2 + \sigma_{T_2}^2\right)\left(\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_1}^2\right)}{\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_1}^2 - 2^{2R_1^G}\left(\sigma_{T_0}^2 + \sigma_{T_1}^2\right)}. \qquad (16)$$

Now write

$$\begin{aligned} &I(X^G, U_1; U_2 \,|\, U_2') \\ &\quad = I(b_3\bar{B}_1 + \bar{B}_2, b_4\bar{B}_1 + b_5\bar{B}_2 + \bar{B}_3; b_6\bar{B}_1 \\ &\qquad + b_7\bar{B}_2 + b_8\bar{B}_3 + \bar{B}_4 \,|\, \bar{B}_1) \\ &\quad = I(\bar{B}_2, b_5\bar{B}_2 + \bar{B}_3; b_7\bar{B}_2 + b_8\bar{B}_3 + \bar{B}_4 \,|\, \bar{B}_1) \\ &\quad = I(\bar{B}_2, b_5\bar{B}_2 + \bar{B}_3; b_7\bar{B}_2 + b_8\bar{B}_3 + \bar{B}_4) \end{aligned}$$

where the last step follows from the fact that $\bar{B}_1$ is independent of $(\bar{B}_2, \bar{B}_3, \bar{B}_4)$. The independence of $\bar{B}_4$ and $(\bar{B}_2, \bar{B}_3)$ implies that $(\bar{B}_2, b_5\bar{B}_2 + \bar{B}_3) \to b_7\bar{B}_2 + b_8\bar{B}_3 \to b_7\bar{B}_2 + b_8\bar{B}_3 + \bar{B}_4$ form a Markov chain. This observation, along with the fact that $b_7\bar{B}_2 + b_8\bar{B}_3$ is a deterministic function of $(\bar{B}_2, b_5\bar{B}_2 + \bar{B}_3)$, yields

$$\begin{aligned} &I(\bar{B}_2, b_5\bar{B}_2 + \bar{B}_3; b_7\bar{B}_2 + b_8\bar{B}_3 + \bar{B}_4) \\ &\qquad = I(b_7\bar{B}_2 + b_8\bar{B}_3; b_7\bar{B}_2 + b_8\bar{B}_3 + \bar{B}_4). \end{aligned} \tag{18}$$

Now substitute $\bar{B}_2$ and $\bar{B}_3$ in (18) with corresponding linear combinations of $X^G, U_1$, and $U_2'$, and define $b_1^* = b_1, b_2^* = b_2, b_3^* = b_7 - b_5b_8, b_4^* = b_8, b_5^* = b_3b_5b_8 - b_3b_7 - b_4b_8$, and $b_6^* = b_6$, it follows that

$$\begin{aligned} R_1^G &= I(b_1X^G + b_2U_2'; U_1) \\ &= I(b_1^*X^G + b_2^*U_2'; b_1^*X^G + b_2^*U_2' + \tilde{B}_3), \end{aligned} \tag{19}$$

$$\begin{aligned} R_2^G &= I(X^G; X^G + \tilde{B}_2) + I(b_3^*X^G + b_4^*U_1 + b_5^*U_2'; b_3^*X^G \\ &\quad + b_4^*U_1 + b_5^*U_2' + \bar{B}_4) \\ &= I(X^G; U_2') + I(b_3^*X^G + b_4^*U_1 + b_5^*U_2'; U_2 - b_6^*U_2') \end{aligned} \tag{20}$$

where the last step follows from the observation that $b_7\bar{B}_2 + b_8\bar{B}_3 + \bar{B}_4 = U_2 - b_6U_2'$.

Intuitively speaking, (19) and (20) suggest the following optimal MD quantization system (also see Fig. 5): Encoder 1 quantizes at rate $R_1^G$ with input $b_1^*\boldsymbol{X}^G + b_2^*\boldsymbol{U}_2'$ and output $\boldsymbol{U}_1$. The quantization error $\tilde{\boldsymbol{B}}_3 = \boldsymbol{U}_1 - b_1^*\boldsymbol{X} - b_2^*\boldsymbol{U}_2'$ is Gaussian with covariance matrix $\mathbb{E}\tilde{B}_3^2 I_n$. Encoder 2 consists of two quantizers. The rate of the first quantizer is $R_{2,1}^G$. Its input and output are $\boldsymbol{X}^G$ and $\boldsymbol{U}_2'$, respectively. Its quantization error $\tilde{\boldsymbol{B}}_2 = \boldsymbol{U}_2' - \boldsymbol{X}^G$ is Gaussian with covariance matrix $\mathbb{E}\tilde{B}_2^2 I_n$. The second quantizer is of rate $R_{2,2}^G$. It has input $b_3^*\boldsymbol{X}^G + b_4^*\boldsymbol{U}_1 + b_5^*\boldsymbol{U}_2'$ and output $\boldsymbol{U}_2 - b_6^*\boldsymbol{U}_2'$. Its quantization error $\bar{\boldsymbol{B}}_4$ is Gaussian with covariance matrix $\mathbb{E}\bar{B}_4^2 I_n$. The sum-rate of these two quantizers is the rate of encoder 2, which is $R_2^G$. Here $R_{2,1}^G = I(X^G; U_2')$, and $R_{2,2}^G = R_2^G - R_{2,1}^G$.

*Remarks:*

1) $\boldsymbol{U}_1$ is revealed to decoder 1 and decoder 3. $\boldsymbol{U}_2'$ and $\boldsymbol{U}_2 - b_6^*\boldsymbol{U}_2'$ are revealed to decoder 2 and decoder 3. Decoder 1 constructs $\hat{\boldsymbol{X}}_1^G = \alpha_1\boldsymbol{U}_1$. Decoder 2 first constructs $\boldsymbol{U}_2$ using $\boldsymbol{U}_2'$ and $\boldsymbol{U}_2 - b_6^*\boldsymbol{U}_2'$, and then constructs $\hat{\boldsymbol{X}}_2^G = \alpha_2\boldsymbol{U}_2$. Decoder 3 also first constructs $\boldsymbol{U}_2$, then constructs $\hat{\boldsymbol{X}}_3 = \beta_1\boldsymbol{U}_1 + \beta_2\boldsymbol{U}_2$. It is clear what decoders 2 and 3 want is $\boldsymbol{U}_2$, not $\boldsymbol{U}_2'$ or $\boldsymbol{U}_2 - b_6^*\boldsymbol{U}_2'$. Furthermore, the construction of $\boldsymbol{U}_2$ can be moved to the encoder part. That is, encoder 2 can directly construct $\boldsymbol{U}_2$ with $\boldsymbol{U}_2'$ and $\boldsymbol{U}_2 - b_6^*\boldsymbol{U}_2'$; then, only $\boldsymbol{U}_2$ needs to be revealed to decoder 2 and decoder 3.

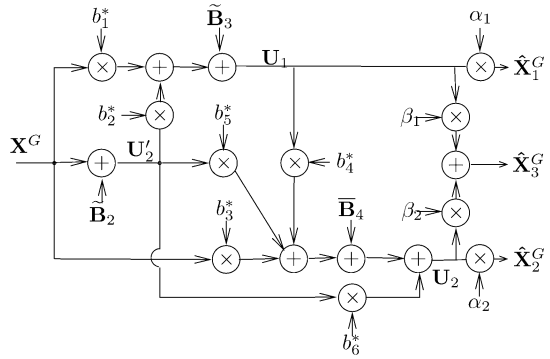


Fig. 5. MD quantization scheme for $(R_1^G, R_2^G)$.

2) That $\bar{\boldsymbol{B}}_4$ is independent of $(\boldsymbol{X}^G, \boldsymbol{U}_1, \boldsymbol{U}_2')$ and $(\tilde{\boldsymbol{B}}_2, \tilde{\boldsymbol{B}}_3)$ is a deterministic function of $(\boldsymbol{X}^G, \boldsymbol{U}_1, \boldsymbol{U}_2')$ implies that $\bar{\boldsymbol{B}}_4$ is independent of $(\tilde{\boldsymbol{B}}_2, \tilde{\boldsymbol{B}}_3)$.
3) The MD quantization scheme for $(R_1^G, R_2^G)$ essentially consists of two Gram–Schmidt procedures, one operating on $(\boldsymbol{X}^G, \boldsymbol{U}_2', \boldsymbol{U}_1)$ and the other on $(\boldsymbol{U}_2', \boldsymbol{X}^G, \boldsymbol{U}_1, \boldsymbol{U}_2)$. The formal description and analysis of this scheme from the perspective of dithered quantization is left to Section V.

### D. Discussion of Special Cases

Next we consider three cases for which the MD quantizers have some special properties.

*1) The case $D_3 = (1/D_1 + 1/D_2 - 1/\sigma_X^2)^{-1}$:* For this case, we have

$$\begin{aligned} \left(R_1\left(V_1^G\right), R_2\left(V_1^G\right)\right) &= \left(R_1\left(V_2^G\right), R_2\left(V_2^G\right)\right) \\ &= \left(\frac{1}{2}\log\frac{\sigma_X^2}{D_1}, \frac{1}{2}\log\frac{\sigma_X^2}{D_2}\right) \end{aligned}$$

which is referred to as the case of no excess marginal rate. Since the dominant face of $\mathcal{R}^G(U_1, U_2)$ degenerates to a single point, the quantization splitting becomes unnecessary. Moreover, we have $\mathbb{E}(U_1 - X^G)(U_2 - X^G) = 0$, i.e., two quantization errors are uncorrelated (and thus independent since $(X^G, U_1, U_2)$ are jointly Gaussian) in this case. This further implies that $U_1 \to X^G \to U_2$ form a Markov chain. Due to this fact, the Gram–Schmidt othogonalization for $(X^G, U_1, U_2)$ becomes particularly simple

$$\begin{aligned} B_1 &= X^G \\ B_2 &= U_1 - \mathbb{E}(U_1 \,|\, X) = U_1 - X^G \\ B_3 &= U_2 - \mathbb{E}(U_2 \,|\, X, U_1) = \mathbb{E}(U_2 \,|\, X^G) = U_2 - X^G. \end{aligned}$$

Thus, the successive quantization scheme degenerates into the conventional separate quantization scheme [32], which in fact suffices for this special case. See Fig. 6.

*2) The case $D_3 = D_1 + D_2 - \sigma_X^2$:* For this case, we have

$$R_1\left(V_1^G\right) + R_2\left(V_1^G\right) = R_1\left(V_2^G\right) + R_2\left(V_2^G\right) = \frac{1}{2}\log\frac{\sigma_X^2}{D_3}$$

which corresponds to the case of no excess sum-rate. Since $D_3 = D_1 + D_2 - \sigma_X^2$ implies $\sigma_X^2 - D_1 = D_2 - D_3$ and

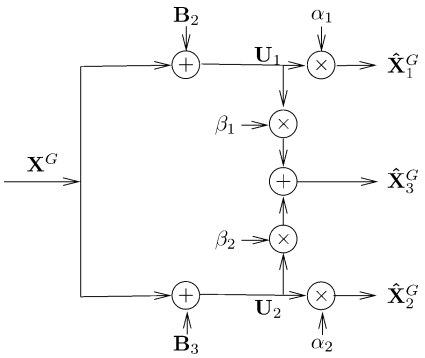


Fig. 6. Special case: $D_3 = (1/D_1 + 1/D_2 - 1/\sigma_X^2)^{-1}$.

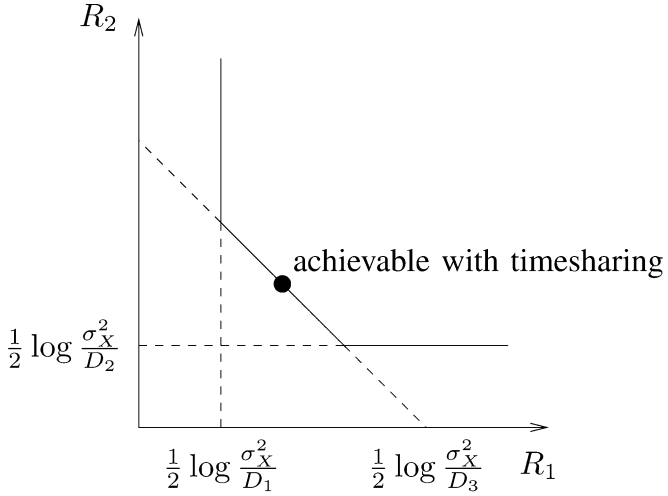


Fig. 7. Special case: $D_3 = D_1 + D_2 - \sigma_X^2$.

$\sigma_X^2 - D_2 = D_1 - D_3$, we get (21) at the bottom of the page. Since $U_1$ and $U_2$ are jointly Gaussian, (21) implies $U_1$ and $U_2$ are independent. This is consistent with the result in [6] although only discrete memoryless sources were addressed there. The interpretation of (21) is that the outputs of the two encoders (quantizers) should be independent. This is intuitively clear because otherwise these two outputs can be further compressed to reduce the sum-rate but still achieve distortion $D_3$ for the joint description. But that would violate the rate distortion theorem, since $\frac{1}{2}\log\frac{\sigma_X^2}{D_3}$ is the minimum $D_3$-admissible rate for the quadratic Gaussian case.

Now consider the following time-sharing scheme: Construct an optimal rate–distortion codebook of rate $\frac{1}{2}\log\frac{\sigma_X^2}{D_3}$ that can achieve distortion $D_3$. Encoder 1 uses this codebook a fraction $\gamma$ $(0 \le \gamma \le 1)$ of the time and encoder 2 uses this codebook the remaining $1-\gamma$ fraction of the time. For this scheme, the resulting rates and distortions are given by $R_1 = \frac{\gamma}{2}\log\frac{\sigma_X^2}{D_3}$, $R_2 = \frac{1-\gamma}{2}\log\frac{\sigma_X^2}{D_3}$, $D_1 = \gamma D_3 + (1-\gamma)\sigma_X^2$, $D_2 = (1-\gamma)D_3 + \gamma\sigma_X^2$, and $D_3$. Conversely, for any fixed $D_1^*$ and $D_2^*$ with $D_1^* + D_2^* = \sigma_X^2 + D_3$, there exists a $\gamma^* \in [0,1]$ such that $D_1^* = \gamma^* D_3 + (1-\gamma^*)\sigma_X^2$, $D_2^* = (1-\gamma^*)D_3 + \gamma^*\sigma_X^2$. The associated rates are $R_1 = \frac{\gamma^*}{2}\log\frac{\sigma_X^2}{D_3}$ and $R_2 = \frac{1-\gamma^*}{2}\log\frac{\sigma_X^2}{D_3}$. So, this specific time-sharing scheme can achieve any point on the dominant face of the rate region for the special case $D_3 = D_1 + D_2 - \sigma_X^2$ (see Fig. 7). Specifically, for the symmetric case where $D_1^* = D_2^* = \frac{1}{2}(D_3 + \sigma_X^2)$, we have $\gamma^* = \frac{1}{2}$ and $R_1 = R_2 = \frac{1}{4}\log\frac{\sigma_X^2}{D_3}$.

*3) The symmetric case $D_1 = D_2 \triangleq D_{12}$:* The symmetric case is of particular practical importance. Moreover, many previously derived expressions take simpler forms if $D_1 = D_2$. Specifically, we have

$$\sigma_{T_1}^2 = \sigma_{T_2}^2 = \frac{D_{12}\sigma_X^2}{\sigma_X^2 - D_{12}} - \frac{D_3\sigma_X^2}{\sigma_X^2 - D_3} \triangleq \sigma_{T_{12}}^2.$$

The coordinates of $V_1^G$ and $V_2^G$ become

$$R_1\left(V_1^G\right) = R_2\left(V_2^G\right) = \frac{1}{2}\log\frac{\sigma_X^2}{D_{12}},$$

$$R_2\left(V_1^G\right) = R_1\left(V_2^G\right) = \frac{1}{2}\log\frac{D_{12}\left(\sigma_X^2 - D_3\right)^2}{4D_3\left(\sigma_X^2 - D_{12}\right)\left(D_{12} - D_3\right)}.$$

The expressions for $R_1^G$ and $R_2^G$ can be simplified to the expression shown at the bottom of the following page. To keep the rates equal, i.e., $R_1^G = R_2^G$, it must be true that

$$\begin{aligned} &4\sigma_{T_0}^2\sigma_{T_{12}}^2 + \sigma_{T_3}^2\left(\sigma_{T_0}^2 + \sigma_{T_{12}}^2\right) \\ &\quad = 2\sigma_{T_0}\sigma_{T_{12}}\left(\sigma_{T_0}^2 + \sigma_{T_{12}}^2 + \sigma_{T_3}^2\right) \\ &\quad \Leftrightarrow \left(\sigma_{T_0} - \sigma_{T_{12}}\right)^2\left(\sigma_{T_3}^2 - 2\sigma_{T_0}\sigma_{T_{12}}\right) = 0. \end{aligned}$$

If $\sigma_{T_0} \neq \sigma_{T_{12}}$, then

$$\sigma_{T_3}^2 = 2\sigma_{T_0}\sigma_{T_{12}}. \tag{22}$$

$$\begin{aligned} \mathbb{E}(U_1U_2) &= \sigma_X^2 + \sigma_{T_0}^2 - \sigma_{T_1}\sigma_{T_2} \\ &= \sigma_X^2 + \frac{D_3\sigma_X^2}{\sigma_X^2 - D_3} - \sqrt{\left(\frac{D_1\sigma_X^2}{\sigma_X^2 - D_1} - \frac{D_3\sigma_X^2}{\sigma_X^2 - D_3}\right)\left(\frac{D_2\sigma_X^2}{\sigma_X^2 - D_2} - \frac{D_3\sigma_X^2}{\sigma_X^2 - D_3}\right)} \\ &= \sigma_X^2 + \frac{D_3\sigma_X^2}{\sigma_X^2 - D_3} - \sqrt{\frac{\sigma_X^8(D_1 - D_3)(D_2 - D_3)}{\left(\sigma_X^2 - D_1\right)\left(\sigma_X^2 - D_2\right)\left(\sigma_X^2 - D_3\right)^2}} \\ &= \sigma_X^2 + \frac{D_3\sigma_X^2}{\sigma_X^2 - D_3} - \frac{\sigma_X^4}{\sigma_X^2 - D_3} \\ &= 0. \end{aligned} \tag{21}$$

If $\sigma_{T_0} = \sigma_{T_{12}}$, then

$$R_1\left(V_1^G\right) = R_2\left(V_1^G\right) = R_1^G = R_2^G$$
$$= R_1\left(V_2^G\right) = R_2\left(V_2^G\right), \qquad \forall \sigma_{T_3}^2 \in \mathcal{R}^+$$

i.e., $(R_1^G, R_2^G)$ is not a function of $\sigma_{T_3}^2$.

## V. Optimal MD Quantization System

In the MD quantization scheme for the quadratic Gaussian case outlined in the preceding section, only the second-order statistics are needed and the resulting quantization system naturally consists mainly of linear operations. In this section, we develop this system in the context of the entropy coded dithered (lattice) quantization (ECDQ) for general sources with the squared error distortion measure. The proposed system may not be optimal for general sources; however, if all the underlying second-order statistics are kept identical with those of the quadratic Gaussian case, then the resulting distortions will also be the same. Furthermore, since among all the i.i.d. sources with the same variance, the Gaussian source has the highest differential entropy, the rates of the quantizers can be upper-bounded by the rates in the quadratic Gaussian case. At high resolution, we prove a stronger result in Section V-C that the proposed MD quantization system is asymptotically optimal for all i.i.d. sources that have finite differential entropy.

In the sequel, we discuss the MD quantization schemes in an order that parallels the development in the preceding section. The source $\{X(t)\}_{t=1}^{\infty}$ is assumed to be an i.i.d. random process (not necessarily Gaussian) with $\mathbb{E}X(t) = 0$ and $\mathbb{E}X^2(t) = \sigma_X^2$ for all $t$.

### A. Successive Quantization Using ECDQ

Consider the MD quantization system depicted in Fig. 8, which corresponds to the Gaussian MD coding scheme for $V_1^G$. Let $Q_{1,n}(\cdot)$ and $Q_{2,n}(\cdot)$ denote optimal $n$-dimensional lattice quantizers. Let $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ be $n$-dimensional random vectors which are statistically independent and each is uniformly distributed over the basic cell of the associated lattice quantizer. The lattices have a "white" quantization noise covariance matrix of the form $\sigma_i^2 I_n = \mathbb{E}\boldsymbol{Z}_i\boldsymbol{Z}_i^T$, where $\sigma_i^2$ is the second moment of the lattice quantizer $Q_{i,n}(\cdot), i = 1, 2$; more specifically, let $\sigma_1^2 = \mathbb{E}B_2^2, \sigma_2^2 = \mathbb{E}B_3^2$, where $\mathbb{E}B_2^2$ and $\mathbb{E}B_3^2$ are given by (38) and (39), respectively. Furthermore, let

$$\boldsymbol{W}_1 = Q_{1,n}(\boldsymbol{X} + \boldsymbol{Z}_1) - \boldsymbol{Z}_1$$
$$\boldsymbol{W}_2 = Q_{2,n}(a_1\boldsymbol{X} + a_2\boldsymbol{W}_1 + \boldsymbol{Z}_2) - \boldsymbol{Z}_2$$

where $a_1$ and $a_2$ are given by (28) and (29), respectively.

*Theorem 5.1:* The first- and second-order statistics of $(\boldsymbol{X}, \boldsymbol{W}_1, \boldsymbol{W}_2)$ are the same as the first- and second-order statistics of $(\boldsymbol{X}^G, \boldsymbol{U}_1, \boldsymbol{U}_2)$ in Section IV.

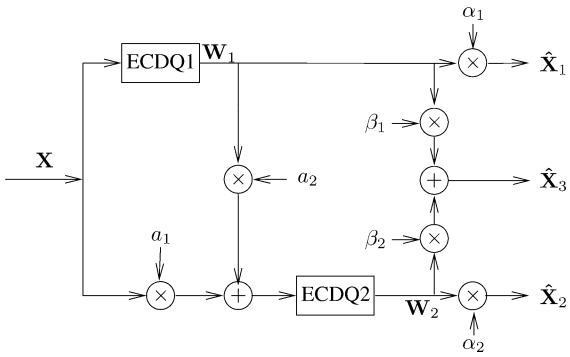


Fig. 8. Successive quantization.

*Remark:* The first-order statistics of $(\boldsymbol{X}, \boldsymbol{W}_1, \boldsymbol{W}_2)$ and $(\boldsymbol{X}^G, \boldsymbol{U}_1, \boldsymbol{U}_2)$ are all zero, so we focus on the second-order statistics. The same is true for the other random variables in this section.

*Proof:* The theorem follows directly from the correspondence between the Gram–Schmidt orthogonalization and the sequential (dithered) quantization established in Section II, and it is straightforward by comparing Figs. 4 and 8. Essentially, $\boldsymbol{X}, \boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ serve as the innovations that generate the first- and second-order statistics of the whole system. □

By Theorem 5.1,

$$\frac{1}{n}\mathbb{E}\|\boldsymbol{X} - \alpha_i\boldsymbol{W}_i\|^2 = \frac{1}{n}\mathbb{E}\|\boldsymbol{X}^G - \alpha_i\boldsymbol{U}_i\|^2 = D_i, \quad i = 1, 2$$
$$\frac{1}{n}\mathbb{E}\|\boldsymbol{X} - \sum_{i=1}^{2}\beta_i\boldsymbol{W}_i\|^2 = \frac{1}{n}\mathbb{E}\|\boldsymbol{X}^G - \sum_{i=1}^{2}\beta_i\boldsymbol{U}_i\|^2 = D_3.$$

Let $\boldsymbol{N}_i$ be an $n$-dimensional random vector distributed as $-\boldsymbol{Z}_i, i = 1, 2, 3$. By property 2) of the ECDQ, the rate $R_1$ of the quantizer $Q_{1,n}(\cdot)$ can be computed and bounded as follows:

$$\begin{aligned} &H(Q_{1,n}(\boldsymbol{X} + \boldsymbol{Z}_1) \,|\, \boldsymbol{Z}_1) \\ &\quad = I(\boldsymbol{X}; \boldsymbol{X} + \boldsymbol{N}_1) \\ &\quad = h(\boldsymbol{X} + \boldsymbol{N}_1) - h(\boldsymbol{N}_1), \\ &\quad = \frac{1}{n}h(\boldsymbol{W}_1) - \frac{1}{n}h(\boldsymbol{N}_1) \\ &\quad \le \frac{1}{n}h(\boldsymbol{U}_1) - \frac{1}{n}h(\boldsymbol{N}_1) \\ &\quad = \frac{1}{2}\log\left[2\pi e\left(\sigma_X^2 + \mathbb{E}B_2^2\right)\right] - \frac{1}{2}\log\frac{\mathbb{E}B_2^2}{G_n^{\text{opt}}} \\ &\quad = R_1\left(V_1^G\right) + \frac{1}{2}\log\left(2\pi e G_n^{\text{opt}}\right) \end{aligned}$$

where the inequality follows from Theorem 5.1 and the fact that for a given covariance matrix, the joint Gaussian distribution

$$R_1^G = \frac{1}{2}\log\frac{\left(\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_{12}}^2\right)\left(\sigma_{T_0}^2 + \sigma_{T_{12}}^2 + \sigma_{T_3}^2\right)}{4\sigma_{T_0}^2\sigma_{T_{12}}^2 + \sigma_{T_3}^2\left(\sigma_{T_0}^2 + \sigma_{T_{12}}^2\right)}$$
$$R_2^G = \frac{1}{2}\log\frac{\left[4\sigma_{T_0}^2\sigma_{T_{12}}^2 + \sigma_{T_3}^2\left(\sigma_{T_0}^2 + \sigma_{T_{12}}^2\right)\right]\left(\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_{12}}^2\right)}{4\sigma_{T_0}^2\sigma_{T_{12}}^2\left(\sigma_{T_0}^2 + \sigma_{T_{12}}^2 + \sigma_{T_3}^2\right)}.$$

maximizes the differential entropy. Similarly, the rate $R_2$ of the quantizer $Q_{2,n}(\cdot)$ can be computed and bounded as follows:

$$\begin{aligned} H(Q_{2,n}(a_1\boldsymbol{X}+a_2\boldsymbol{W}_1+\boldsymbol{Z}_2)\,|\,\boldsymbol{Z}_2) &= I(a_1\boldsymbol{X}+a_2\boldsymbol{W}_1; a_1\boldsymbol{X}+a_2\boldsymbol{W}_1+\boldsymbol{N}_2) \\ &= h(a_1\boldsymbol{X}+a_2\boldsymbol{W}_1+\boldsymbol{N}_2)-h(\boldsymbol{N}_2) \\ &= \frac{1}{n}h(\boldsymbol{W}_2)-\frac{1}{n}h(\boldsymbol{N}_1) \\ &\le \frac{1}{n}h(\boldsymbol{U}_2)-\frac{1}{n}h(\boldsymbol{N}_1) \\ &= \frac{1}{2}\log\left[2\pi e\left(\sigma_X^2+\sigma_{T_0}^2+\sigma_{T_2}^2\right)\right]-\frac{1}{2}\log\frac{\mathbb{E}B_3^2}{G_n^{\text{opt}}} \\ &= R_2\left(V_1^G\right)+\frac{1}{2}\log\left(2\pi e G_n^{\text{opt}}\right). \end{aligned}$$

Since $G_n^{\text{opt}} \to \frac{1}{2\pi e}$ as $n \to \infty$, we have $R_1 \le R_1(V_1^G)$ and $R_2 \le R_2(V_1^G)$ as $n \to \infty$.

*Remark:* As suggested by Zamir [64] and an anonymous reviewer, since $a_1\boldsymbol{X}+a_2\boldsymbol{W}_1 = \boldsymbol{X}+a_2\boldsymbol{N}_1$, we can view the input of $Q_{2,n}(\cdot)$ as a linear combination of source $\boldsymbol{X}$ and the quantization error of ECDQ1. Now it becomes transparent that $\boldsymbol{W}_1-\boldsymbol{X}$ and $\boldsymbol{W}_2-\boldsymbol{X}$ are negatively correlated since

$$\mathbb{E}[(\boldsymbol{W}_1-\boldsymbol{X})(\boldsymbol{W}_2-\boldsymbol{X})^T]=\mathbb{E}[\boldsymbol{N}_1(a_2\boldsymbol{N}_1+\boldsymbol{N}_2)^T]=a_2\mathbb{E}[\boldsymbol{N}_1\boldsymbol{N}_1^T]$$

and $a_2 \le 0$ by (13) and (29).

### *B. Successive Quantization With Quantization Splitting Using ECDQ*

Now we proceed to construct the MD quantization system using ECDQ in a manner which corresponds to that for the Gaussian MD quantization scheme for an arbitrary rate pair $(R_1^G, R_2^G)$.

Let $Q_{1,n}^*(\cdot), Q_{2,n}^*(\cdot)$, and $Q_{3,n}^*(\cdot)$ denote optimal $n$-dimensional lattice quantizers. Let $\boldsymbol{Z}_1^*, \boldsymbol{Z}_2^*$, and $\boldsymbol{Z}_3^*$ be $n$-dimensional random vectors which are statistically independent and each is uniformly distributed over the basic cell of the associated lattice quantizer. The lattices have a "white" quantization noise covariance matrix of the form $\sigma_i^{*2} I_n = \mathbb{E}\boldsymbol{Z}_i^*\boldsymbol{Z}_i^{*T}$, where $\sigma_i^{*2}$ is the second moment of the lattice quantizer $Q_{i,n}^*(\cdot), i=1,2,3$; more specifically, let $\sigma_1^{*2} = \mathbb{E}\tilde{B}_2^2, \sigma_2^{*2} = \mathbb{E}\tilde{B}_3^2$, and $\sigma_3^{*2} = \mathbb{E}\bar{B}_4^2$, where $\mathbb{E}\tilde{B}_2^2, \mathbb{E}\tilde{B}_3^2$ and $\mathbb{E}\bar{B}_4^2$ are given by (40), (41), and (42), respectively. Define

$$\begin{aligned} \tilde{\boldsymbol{W}}_2' &= Q_{1,n}^*(\boldsymbol{X}+\boldsymbol{Z}_1^*)-\boldsymbol{Z}_1^* \\ \tilde{\boldsymbol{W}}_1^n &= Q_{2,n}^*(b_1^*\boldsymbol{X}+b_2^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{Z}_2^*)-\boldsymbol{Z}_2^* \\ \boldsymbol{\Delta} &= Q_{3,n}^*(b_3^*\boldsymbol{X}+b_4^*\tilde{\boldsymbol{W}}_1+b_5^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{Z}_3^*)-\boldsymbol{Z}_3^* \\ \tilde{\boldsymbol{W}}_2 &= \boldsymbol{\Delta}+b_6^*\tilde{\boldsymbol{W}}_2'. \end{aligned}$$

The system diagram is shown in Fig. 9.

*Theorem 5.2:* The first- and second-order statistics of $(\boldsymbol{X}, \tilde{\boldsymbol{W}}_1, \tilde{\boldsymbol{W}}_2, \tilde{\boldsymbol{W}}_2', \boldsymbol{\Delta})$ equal the first- and second-order statistics of $(\boldsymbol{X}^G, \boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{U}_2', \boldsymbol{U}_2 - b_6\boldsymbol{U}_2')$ in Section IV.

*Proof:* By comparing Figs. 5 and 9, it is clear that the theorem follows from the correspondence between the Gram–Schmidt orthogonalization and the sequential (dithered) quantization. The following one-to-one correspondences

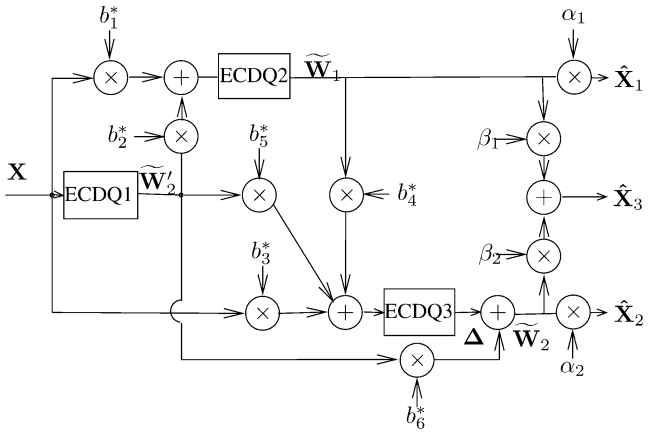


Fig. 9. Successive quantization with quantization splitting.

should be emphasized: $\tilde{\boldsymbol{B}}_2$ and $-\boldsymbol{Z}_1^*, \tilde{\boldsymbol{B}}_3$ and $-\boldsymbol{Z}_2^*, \bar{\boldsymbol{B}}_4$ and $-\boldsymbol{Z}_3^*$. $\boldsymbol{X}, \boldsymbol{Z}_1^*, \boldsymbol{Z}_2^*$, and $\boldsymbol{Z}_3^*$ are the innovations that generate the first- and second-order statistics of the whole system. □

It follows from Theorem 5.2 that

$$\frac{1}{n}\mathbb{E}\|\boldsymbol{X}-\alpha_i\tilde{\boldsymbol{W}}_i\|^2 = \frac{1}{n}\mathbb{E}\|\boldsymbol{X}^G-\alpha_i\boldsymbol{U}_i\|^2 = D_i, \qquad i=1,2$$

$$\frac{1}{n}\mathbb{E}\left\|\boldsymbol{X}-\sum_{i=1}^{2}\beta_i\tilde{\boldsymbol{W}}_i\right\|^2 = \frac{1}{n}\mathbb{E}\left\|\boldsymbol{X}^G-\sum_{i=1}^{2}\beta_i\boldsymbol{U}_i\right\|^2 = D_3.$$

Let $\boldsymbol{N}_i^*$ be an $n$-dimensional random vector distributed as $-\boldsymbol{Z}_i^*, i=1,2,3$. By property 2) of the ECDQ, we bound the rate $R_1$ of $Q_{2,n}(\cdot)$ (conditioned on $\boldsymbol{Z}_2^*$) as follows:

$$\begin{aligned} R_1 &= \frac{1}{n}H(Q_{2,n}^*(b_1^*\boldsymbol{X}+b_2^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{Z}_2^{*n})\,|\,\boldsymbol{Z}_2^*) \\ &= I(b_1^*\boldsymbol{X}+b_2^*\tilde{\boldsymbol{W}}_2'; b_1^*\boldsymbol{X}+b_2^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{N}_2^*) \\ &= \frac{1}{n}h(b_1^*\boldsymbol{X}+b_2^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{N}_2^*)-\frac{1}{n}h(\boldsymbol{N}_2^*) \\ &= \frac{1}{n}h(\tilde{\boldsymbol{W}}_1)-\frac{1}{n}h(\boldsymbol{N}_2^*) \\ &\le \frac{1}{n}h(\boldsymbol{U}_1)-\frac{1}{n}h(\boldsymbol{N}_2^*) \\ &= \frac{1}{2}\log\left[2\pi e\left(\sigma_X^2+\sigma_{T_0}^2+\sigma_{T_1}^2\right)\right]-\frac{1}{2}\log\frac{\mathbb{E}}{\tilde{B}_3}^2 G_n^{\text{opt}} \\ &= R_1^G+\frac{1}{2}\log\left(2\pi e G_n^{\text{opt}}\right) \end{aligned} \tag{23}$$

where the inequality follows from Theorem 5.2 and the fact that for a given covariance matrix, the joint Gaussian distribution maximizes the differential entropy.

Similarly, the sum-rate of $Q_{1,n}^*(\cdot)$ (conditioned on $\boldsymbol{Z}_1^*$) and $Q_{3,n}^*(\cdot)$ (conditioned on $\boldsymbol{Z}_3^*$) can be upper-bounded as follows:

$$\begin{aligned} R_2 &= \frac{1}{n}H(Q_{1,n}^*(\boldsymbol{X}+\boldsymbol{Z}_1^*)\,|\,\boldsymbol{Z}_1^*) \\ &\quad+\frac{1}{n}H(Q_{3,n}^*(b_3^*\boldsymbol{X}+b_4^*\tilde{\boldsymbol{W}}_1+b_5^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{Z}_3^*)\,|\,\boldsymbol{Z}_3^*) \\ &= \frac{1}{n}h(\boldsymbol{X}+\boldsymbol{N}_1^*)-\frac{1}{n}h(\boldsymbol{N}_1^*) \\ &\quad+\frac{1}{n}h(b_3^*\boldsymbol{X}+b_4^*\tilde{\boldsymbol{W}}_1^n+b_5^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{N}_3^*)-\frac{1}{n}h(\boldsymbol{N}_3^*) \\ &= \frac{1}{n}h(\tilde{\boldsymbol{W}}_2')-\frac{1}{n}h(\boldsymbol{N}_1^*)+\frac{1}{n}h(\boldsymbol{\Delta})-\frac{1}{n}h(\boldsymbol{N}_3^*) \end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{n}h(\boldsymbol{U}_2') - \frac{1}{n}h(\boldsymbol{N}_1^*) + \frac{1}{n}h(\boldsymbol{U}_2 - b_6\boldsymbol{U}_2') - \frac{1}{n}h(\boldsymbol{N}_3^*)\\
&\overset{(a)}{=} \frac{1}{n}h(\boldsymbol{U}_2') - \frac{1}{n}h(\boldsymbol{N}_1^*)\\
&\quad + \frac{1}{n}h(b_7\bar{\boldsymbol{B}}_2 + b_8\bar{\boldsymbol{B}}_3 + \bar{\boldsymbol{B}}_4) - \frac{1}{n}h(\boldsymbol{N}_3^*)\\
&= \frac{1}{2}\log\left[2\pi e\left(\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2\right)\right] - \frac{1}{2}\log\frac{\mathbb{E}\tilde{B}_2^2}{G_n^{\text{opt}}}\\
&\quad + \frac{1}{2}\log\left[2\pi e\left(b_7^2\mathbb{E}\bar{B}_2^2 + b_8^2\mathbb{E}\bar{B}_3^2 + \mathbb{E}\bar{B}_4^2\right)\right] - \frac{1}{2}\log\frac{\mathbb{E}\bar{B}_4^2}{G_n^{\text{opt}}}\\
&= R_2^G + \log\left(2\pi e G_n^{\text{opt}}\right) \qquad (24)
\end{aligned}
$$

where (a) follows from the Gram–Schmidt orthogonalization of $U_2', X^G, U_1$, and $U_2$.

*Remark:* Since the decoders only need to know $\tilde{\boldsymbol{W}}_2 = \boldsymbol{\Delta} + b_6^*\tilde{\boldsymbol{W}}_2'$ instead of $\tilde{\boldsymbol{W}}_2'$ and $\boldsymbol{\Delta}$ separately, we can actually further reduce $R_2$ to $\frac{1}{n}H(\tilde{\boldsymbol{W}}_2 \mid \boldsymbol{Z}_1^*, \boldsymbol{Z}_2^*, \boldsymbol{Z}_3^*)$. Since $G_n^{\text{opt}} \to \frac{1}{2\pi e}$ as $n \to \infty$, it follows from (23) and (24) that $R_1 \le R_1^G, R_2 \le R_2^G$ as $n \to \infty$.

The preceding results imply that for general i.i.d. sources, under the same distortion constraints, the rates required by our scheme are upper-bounded by the rates required for the quadratic Gaussian case. This further implies our scheme can achieve the whole Gaussian MD rate–distortion region as the dimension of the (optimal) lattice quantizers becomes large.

### C. Optimality and an Upper Bound on the Coding Rates

Define $\mathcal{Q}_{\text{out}}$ such that $(R_1, R_2, D_1, D_2, D_3) \in \mathcal{Q}_{\text{out}}$ if and only if

$$
\begin{aligned}
R_i &\ge \frac{1}{2}\log\frac{P_X}{D_i}, \qquad i = 1, 2\\
R_1 + R_2 &\ge \frac{1}{2}\log\frac{P_X}{D_3} + \frac{1}{2}\log\phi(D_1, D_2, D_3)
\end{aligned}
$$

where we have the expression of $\phi(D_1, D_2, D_3)$ at the bottom of the page, and $P_X = 2^{2h(X)}/2\pi e$ is the entropy power of $X$. It was shown by Zamir [11] that for i.i.d. sources with finite differential entropy, $\mathcal{Q}_{\text{out}}$ is an outer bound of the MD rate–distortion region and is asymptotically tight at high resolution (i.e., $D_1, D_2, D_3 \to 0$ along a straight line). Again, we only need consider the case $D_1 + D_2 - P_X \le D_3 \le (\frac{1}{D_1} + \frac{1}{D_2} - \frac{1}{P_X})^{-1}$. At high resolution, we can write

$$
\begin{aligned}
&\frac{1}{2}\log\phi(D_1, D_2, D_3)\\
&\qquad = \frac{1}{2}\log\frac{P_X}{(\sqrt{D_1 - D_3} + \sqrt{D_2 - D_3})^2} + o(1)
\end{aligned}
$$

where $o(1) \to 0$ as $D_1, D_2, D_3 \to 0$.

The following theorem says our scheme is asymptotically optimal at high resolution for general smooth i.i.d. sources.

*Theorem 5.3:* As $D_1, D_2$, and $D_3$ approach zero along a straight line, the region

$$
\begin{aligned}
R_i &\ge \frac{1}{2}\log\frac{P_X}{D_i} + \frac{1}{2}\log(2\pi e G_n^{\text{opt}}) + o(1), \quad i = 1, 2\\
R_1 + R_2 &\ge \frac{1}{2}\log\frac{P_X}{D_3} + \frac{1}{2}\log\frac{P_X}{(\sqrt{D_1 - D_3} + \sqrt{D_2 - D_3})^2}\\
&\quad + \frac{3}{2}\log(2\pi e G_n^{\text{opt}}) + o(1)
\end{aligned}
$$

is achievable using optimal $n$-dimensional lattice quantizers via successive quantization with quantization splitting.

*Proof:* See Appendix III. □

*Remark:*

1) As $D_1, D_2, D_3 \to 0$ and $n \to \infty$, the above region converges to the outer bound and thus is asymptotically tight.
2) The sum-rate redundancy of our MD quantization scheme (i.e., successive quantization with quantization splitting) is at most three times the redundancy of an optimal $n$-dimensional lattice quantizer in the high-resolution regime. It is easy to see from (23) and (24) that for the Gaussian source, this is true at all resolutions. Specifically, for scalar quantizers, we have $G_1^{\text{opt}} = \frac{1}{12}$, and thus the redundancy is $\frac{3}{2}\log\frac{\pi e}{6}$. This actually overestimates the sum-rate redundancy of our scheme in certain cases. It will be shown in the next section that for the scalar case, the redundancy is approximately twice the redundancy of a scalar quantizer at high resolution.
3) If the successive quantization with quantization splitting is replaced by time-sharing the quantization schemes for two vertices, then since vertex only requires two quantization operations, it can be shown that the redundancy of the time-sharing approach is at most twice the redundancy of an optimal $n$-dimensional lattice quantizer.

The following theorem gives a single-letter upper bound on the rates of our scheme at all resolutions as the dimension of the optimal lattices becomes large.

*Theorem 5.4:* There exists a sequence of lattice dimensions $n_1, n_2, \ldots, n_m, \cdots$ ($n_m \to \infty$ as $m \to \infty$) such that

$$
\begin{aligned}
&\limsup_{m\to\infty}\frac{1}{n_m}[h(\boldsymbol{X} + \boldsymbol{N}_1^*) - h(\boldsymbol{N}_1^*)]\\
&\quad \le h\left(X + N_1^G\right) - h\left(N_1^G\right)\\
&\limsup_{m\to\infty}\frac{1}{n_m}[h(b_1^*\boldsymbol{X} + b_2^*\tilde{\boldsymbol{W}}_2' + \boldsymbol{N}_2^*) - h(\boldsymbol{N}_2^*)]\\
&\quad \le h\left(X + b_2^*N_1^G + N_2^G\right) - h\left(N_2^G\right)
\end{aligned}
$$

$$
\phi(D_1, D_2, D_3) = \begin{cases} 1, & D_3 < D_1 + D_2 - P_X\\ \dfrac{P_X D_3}{D_1 D_2}, & D_3 > \left(\frac{1}{D_1} + \frac{1}{D_2} - \frac{1}{P_X}\right)^{-1}\\ \dfrac{(P_X - D_3)^2}{(P_X - D_3)^2 - [\sqrt{(P_X - D_1)(P_X - D_2)} - \sqrt{(D_1 - D_3)(D_2 - D_3)}]^2}, & \text{o.w.} \end{cases}
$$

and

$$\limsup_{m\to\infty}\frac{1}{n_m}\left[h\left(b_3^*\boldsymbol{X}+b_4^*\tilde{\boldsymbol{W}}_1^n+b_5^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{N}_3^*\right)-h(\boldsymbol{N}_3^*)\right]$$
$$\le h\left((b_3^*+b_1^*b_4^*+b_2^*b_4^*+b_5^*)X+(b_2^*b_4^*+b_5^*)N_1^G+b_4^*N_2^G+N_3^G\right)-h\left(N_3^G\right)$$

where

$$N_1^G\sim\mathcal{N}(0,\mathbb{E}\tilde{B}_2^2),N_1^G\sim\mathcal{N}(0,\mathbb{E}\tilde{B}_3^2),N_1^G\sim\mathcal{N}(0,\mathbb{E}\bar{B}_4^2)$$

and the generic source variable $X$ are all independent.

*Remark:* This theorem implies that as the dimension of the optimal lattices goes to infinity, the rates required by our scheme can be upper-bounded as

$$R_1\le h\left(X+b_2^*N_1^G+N_2^G\right)-h\left(N_2^G\right),$$
$$R_2\le h\left(X+N_1^G\right)-h\left(N_1^G\right)+h\left((b_3^*+b_1^*b_4^*+b_2^*b_4^*+b_5^*)X+(b_2^*b_4^*+b_5^*)N_1^G+b_4^*N_2^G+N_3^G\right)-h\left(N_3^G\right).$$

By comparing the above two expressions with (19) and (20), we can see that if $X$ is not Gaussian, then $R_1<R_1^G,R_2<R_2^G$.

*Proof:* See Appendix IV. □

## VI. A Scalar Quantization Scheme

In this section, we give a geometric interpretation of our MD quantization scheme when undithered scalar quantization is used in the proposed framework. This interpretation serves as a bridge between the information-theoretic description of the coding scheme [7] and the practical quantization operation. Furthermore, it facilitates a high-resolution analysis, which offers a performance comparison between the proposed quantization scheme and existing MD quantization techniques.

### A. The Geometric Interpretation

A classical scalar quantizer can be modeled to be composed of three components [63].

1) The *lossy encoder* is a mapping $q:\mathcal{R}\to\mathcal{I}$, where the index set $\mathcal{I}$ is usually taken as a collection of consecutive integers. Commonly, this lossy encoder is alternatively specified by a partition of $\mathcal{R}$, i.e., the boundary points of the partition segments.
2) The *lossy decoder* is a mapping $q^{-1}:\mathcal{I}\to\mathcal{R}'$, where $\mathcal{R}'\subset\mathcal{R}$ is the reproduction codebook.
3) The *lossless encoder* $\gamma:\mathcal{I}\to\mathcal{C}$ is an invertible mapping into a collection $\mathcal{C}$ of variable-length binary vectors. This is essentially the entropy coding of the quantization indices.

The successive quantization coding scheme in Fig. 8 is redrawn in terms of quantization encoder and decoder in Fig. 10. The lossless encoder $\gamma$ is not essential in this interpretation and is thus omitted. The lossy decoders in the receiver are mappings $q_1^{-1}:\mathcal{I}_1\to\mathcal{R}_1',q_2^{-1}:\mathcal{I}_2\to\mathcal{R}_2'$, and $q_3^{-1}:\mathcal{I}_1\times\mathcal{I}_2\to\mathcal{R}_3'$, respectively.

[7]Although the ECDQ-based MD scheme considered in the preceding section is certainly of practical value, we mainly use it as an analytical tool to establish the optimality of our scheme. In practice, it is more desirable to have an MD scheme based on undithered quantization.

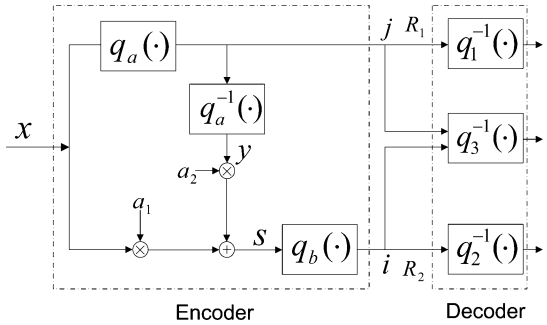


Fig. 10. Coding scheme using successive quantization in terms of quantization encoder and decoder.

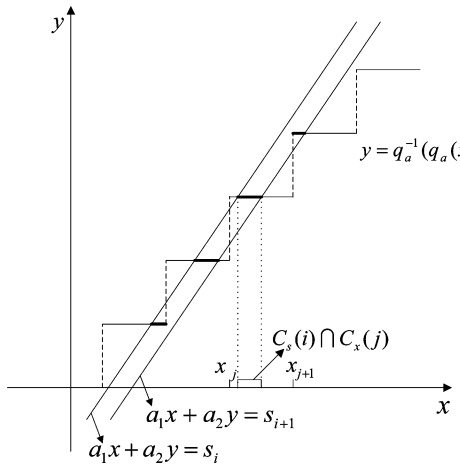


Fig. 11. The geometric interpretation of the partitions using successive quantization.

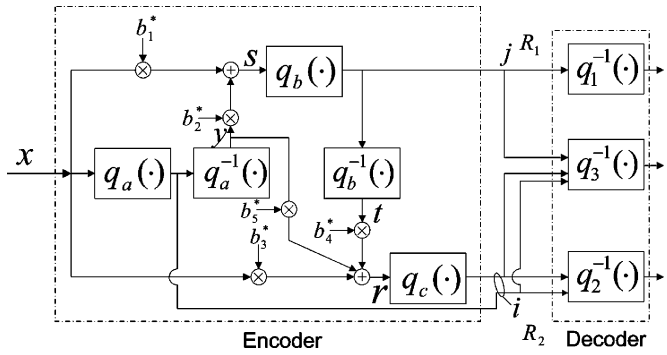


Fig. 12. Coding scheme using quantization splitting in terms of quantization encoder and decoder.

For simplicity, assume the lossy encoders $q_a$ and $q_b$ generate uniform partitions of $\mathcal{R}$, respectively, while the lossy decoder $q_a^{-1}$ takes the center points of the partition cells of $q_a$ as the reproduction codebook. Function $y=q_a^{-1}(q_a(x))$ is piecewise constant. A linear combination of $x$ and $y$ is then formed as $s=a_1x+a_2y$, which is mapped by $q_b$ to a quantization index $q_b(s)$. Consider a partition cell $i$, given by $(s_i,s_{i+1}]$, in the lossy encoder $q_b$.

In Fig. 11, this partition cell is represented on the $(x,y)$ plane. For operating points on the dominant face of the SEGC region, it is always true that $\sigma_{T_0}^2\le\sigma_{T_1}\sigma_{T_2}$, which implies $a_2\le 0$ [from (29)], and thus the slope of the line $a_1x+a_2y=s_i$ is always positive. It is clear that, given $q_b(s)=i$, $x$ can fall only into the several segments highlighted by the thicker lines in Fig. 11, i.e., into the set $C_s(i)=\{x:a_1x+a_2q_a(q_a^{-1}(x))\in(s_i,s_{i+1}]\}$. The information regarding $x$ is thus revealed to the lossy decoder $q_2^{-1}$. In the

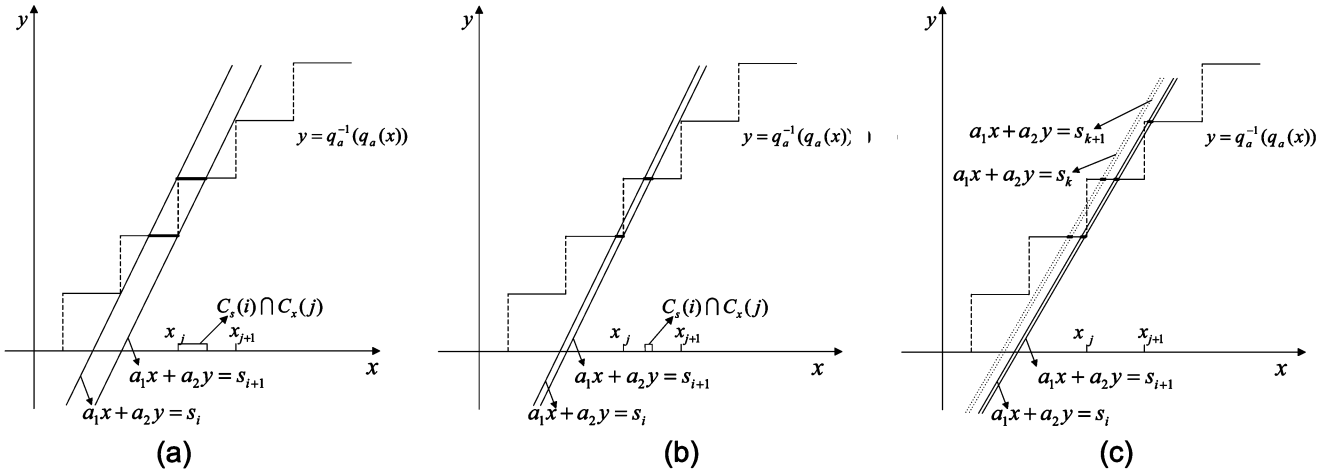


Fig. 13. Several special cases of the partition formed using successive quantization. (a) $a_1 = 2, a_2 = -1$, and the stepsize of $q_a$ is the same as that of $q_b$. (b) $a_1 = 2, a_2 = -1$, but the stepsize of $q_a$ is much larger than that of $q_b$. (c) When the stepsize of $q_a$ is much larger than that of $q_b$, by slightly varying $a_1$ and $a_2$, the two side distortions can be made equal.

lossy encoder $q_a$, the information is revealed to the lossy decoder $q_1^{-1}$ in the traditional manner that, when index $j$ is specified, $x$ is in the $j$th cell, which is $(x_j, x_{j+1}]$; denote it as $C_x(j) = (x_j, x_{j+1}]$. Jointly, the lossy decoder $q_3^{-1}$ has the information that $x$ is in the intersection of the two sets as $C_s(i) \cap C_x(j)$.

The successive quantization with quantization splitting scheme in Fig. 9 is redrawn in Fig. 12. It can be observed that $q_a, q_a^{-1}$ and $q_b$ play roles similar to those in Fig. 10. Let $s = b_1^* x + b_2^* q_a^{-1}(q_a(x))$ and define

$$C_s(j) = \{x : b_1^* x + b_2^* q_a^{-1}(q_a(x)) \in (s_j, s_{j+1}]\}$$

where $(s_j, s_{j+1}]$ is the $j$th partition cell in the lossy encoder $q_b$.

Notice the index $i = (i_a, i_c)$ has two components, one is the output of $q_a$, and the other is that of $q_c$. The lossy encoder $q_c$ and the lossy decoders $q_2^{-1}$ and $q_3^{-1}$ always have the exact output from $q_a$, which in effect confines the source to a finite range. Thus, we need to consider only the case for a fixed $q_a(x)$ value. It is obvious that when $q_a(x) = i_0$ is fixed, $q_a^{-1}(i_0) = y_{i_0}$. Consider the linear combination of $r = b_3^* x + b_4^* t + b_5^* q_a^{-1}(q_a(x))$, where $t = q_b^{-1}(q_b(s))$. It is similar to the linear combination of $s = b_1^* x + b_2^* y$, but with the additional constant term $b_5^* y_{i_0}$, when $i_0$ is given. It can be shown that this constant term in fact removes the conditional mean such that $E(r \,|\, q_a(x) = i_0) \approx 0$, and the lossy encoder $q_c$ is merely a partition of an interval near zero. Thus, with $q_a(x) = i_0$ given, $q_b, q_b^{-1}$ and $q_c$ essentially adopt the same roles as $q_a, q_a^{-1}$ and $q_b$, respectively, in Fig. 10. This implies that a similar geometric interpretation again holds for the additional components in Fig. 12, since $b_4^* = b_8 \leq 0$ and $b_3^* = b_7 - b_5 b_8 > 0$. Define

$$C_{xr}(i_a, i_c) = \{x : x \in (x_{i_a}, x_{i_a+1}], r \in (r_{i_c}, r_{i_c+1}]\}$$

where $(x_{i_a}, x_{i_a+1}]$ is the $i_a$th partition cell in the lossy encoder $q_a$ and $(r_{i_c}, r_{i_c+1}]$ is the $i_c$th partition cell in the lossy encoder $q_c$. Given the index pair $(i, j) = (i_a, i_c, j)$, the joint lossy decoder $q_3^{-1}$ is provided with information that $x \in C_s(j) \cap C_{xr}(i_a, i_c)$.

### *B. High-Resolution Analysis of Several Special Cases*

In what follows, the high-resolution performance of the proposed coding scheme using scalar quantization is analyzed under several special conditions. Of particular interest is the balanced case, where $R_1 = R_2 = R$ and two side distortions are equal, $D_1 = D_2$. For the sake of simplicity, we focus on the zero-mean Gaussian source. For this case, it can be shown [22] that the central and side distortion product remains bounded by a constant at fixed rate, which is $D_3 D_1 \geq \frac{1}{4}\sigma_x^4 2^{-4R}$, independent of the tradeoff between them. For the sake of simplicity, we focus on the zero-mean Gaussian source.

*1) High-Resolution Analysis for Successive Quantization:* Consider using the quantization method depicted in Fig. 10 to construct two descriptions, such that $D_1 = D_2$ with possibly unbalanced rates. For the case $\sigma_x^2 \gg D_1$ and $D_1 \gg D_3$ at high rate, it is clear that $\sigma_x^2 \gg \sigma_{T_1}^2 = \sigma_{T_2}^2 \gg \sigma_{T_0}^2$. Thus, $a_1 \approx 2$ and $a_2 \approx -1$ [from (28), (29)], which suggests that the slope of the line $a_1 x + a_2 y = s_i$ should be approximately 2 in this case.

Next we consider the three cases depicted in Fig. 13. In Fig. 13(a), $a_1 = 2, a_2 = -1$ are chosen. By properly choosing the thresholds and the stepsize, a symmetric (between the two descriptions) partition can be formed. Cells $C_x(\cdot)$ and cells $C_s(\cdot)$ both are intervals, with bins staggered by half the stepsize. This in effect gives the staggered index assignment of [66], [23]. By using this partition, the central distortion is reduced to $1/4$ of the side distortions. Though in this case the condition $D_1 \gg D_3$ does not hold, choosing $a_1 = 2, a_2 = -1$ indeed generates two balanced descriptions. The high-resolution performance of the partition in Fig. 13(a) is straightforward, being given by

$$D_1 = D_2 \approx \frac{1}{12}\Delta_a^2 \approx \frac{2\pi e}{12} 2^{-2R_1} \sigma_x^2$$

where the second equality is true when entropy coding is assumed, and $D_3 \approx \frac{1}{4} D_1$ (also see [65]).

In Fig. 13(b), the stepsize in $q_b$, which is denoted by $\Delta_b$, is chosen to be much smaller than that of $q_a$, which is denoted as $\Delta_a$; however, $a_1 = 2$ and $a_2 = -1$ are kept unchanged. In this case, the partition by $q_a$ is still uniform, and the performance of $q_1^{-1}$ is given by

$$D_1 \approx \frac{1}{12}\Delta_a^2 \approx \frac{2\pi e}{12} 2^{-2R_1} \sigma_x^2.$$

This differs from the previous case in that most of the cells $C_s(i)$ are no longer intervals, but rather the union of two noncontiguous intervals, when $\Delta_a \gg \Delta_b$; for a small portion of the $C_s$ cells, each of them can consist of three noncontiguous intervals, but when $\Delta_a \gg \Delta_b$, this portion is negligible and will be omitted in the discussion which follows. Furthermore, cell $C_s(i)$ approximately consists of two length-$\Delta_b/2$ intervals whose midpoints are $\frac{1}{2}\Delta_a$ apart. The distortion achieved by using this partition in the lossy decoder $q_2$ is $D_2 \approx (\frac{1}{4}\Delta_a)^2 = \frac{3}{4}D_1$. Intuitively, this says that the average distance of the points in the cell $C_s(i)$ from its reproduction codeword is approximately $\frac{1}{4}\Delta_a$, which is obviously true given the geometric structure of the cell $C_s(i)$. Note that $D_1$ and $D_2$ are not of equal value.

The rate of the second description is less straightforward, but consider the joint partition revealed to $q_3^{-1}$. This partition is almost uniform, while the rate of the output of $q_b$ after entropy coding is one bit less than that when the same partition is used in a classical quantizer, because each cell $C_s(i)$ consists of two local intervals instead of one as in the classical quantizer. Thus,

$$D_3 \approx \frac{1}{12}\Delta_b^2 \approx \frac{2\pi e}{12}2^{-2(R_2+1)}\sigma_x^2 = \frac{2\pi e}{48}2^{-2R_2}\sigma_x^2. \quad (25)$$

It follows that an achievable high-resolution operating point using scalar quantization is given by $D_1 = \frac{2\pi e}{12}2^{-2R_1}\sigma_x^2$, $D_2 = \frac{3}{4}D_1, D_3 = \frac{2\pi e}{48}2^{-2R_2}\sigma_x^2$.

In order to make $D_1 = D_2$ when $\Delta_a \gg \Delta_b$, the values of $a_2$ can be varied slightly. First, let $\Delta_a$ be fixed such that $D_1(\approx \frac{1}{12}\Delta_a^2 \approx \frac{2\pi e}{12}2^{-2R_1}\sigma_x^2)$ and $R_1$ are then both fixed. It is clear with stepsize $\Delta_b$ fixed, as $a_2$ decreases from $-1$, the distortion $D_2$ increases. A simple calculation shows that when $a_2 = -4/3, D_2 > D_1$; thus, the desired value of $a_2$ is in $(-4/3, -1)$, and we find this value to be $a_2 = -1.0445$. The detailed calculation is relegated to Appendix V. By using such a value, it can be shown that an achievable high-resolution operating point is by $D_1 = D_2 \approx \frac{2\pi e}{12}2^{-2R_1}\sigma_x^2$ and $D_3 \approx 0.8974 \cdot \frac{2\pi e}{48}2^{-2R_2}\sigma_x^2$. The rates $R_1$ and $R_2$ are usually not equal.

*2) Balanced Descriptions Using Quantization Splitting:* As previously pointed out, in the quantization splitting coding scheme $\sigma_{T_3}^2$ should be chosen to be $2\sigma_{T_0}\sigma_{T_1}$ when balanced descriptions are required; then $\sigma_{T_1} \gg \sigma_{T_0}$ implies $\sigma_{T_1}^2 \gg \sigma_{T_3}^2 \gg \sigma_{T_0}^2$. It follows that $b_1^* \approx 2, b_2^* = -1, b_3^* \approx 2, b_4^* \approx -1$, and $b_5^* \approx 3$. We make the following remarks assuming these values.

- The conditional expectation $E(r \,|\, q_a(x) = i)$ is approximately zero, which implies only the case in which $q_a^{-1}((q(x))) = 0$ needs to be considered. This is obvious from the geometric structure given in Fig. 13(b) and the values of $b^*$s.
- The partition formed by $q_c$ does not improve the distortion $D_1$ over $q_a$. This is because the slope of the line $b_3^*x + b_4^*t + b_5^*y_{i_0} = r_i$ on the $(x,t)$ plane is given in such a way that it almost aligns with the function $t = f(x)$. In such a case, the cell $C_{xr}(i_a, i_b)$ consists of segments from almost every cell $C_s(j)$ for which $C_s(j) \cap \{x : q_a(x) = i_a\} \neq \emptyset$. Intuitively, it is similar to letting the slope of $a_1x + a_2y = s_i$ have a slope of 1 in Fig. 13(b), such that the distortion $D_2$ does not improve much over $\sigma_x^2$ in the successive quantization case.

With these two remarks, consider constructing balanced descriptions using scalar quantization for $R_1 = R_2$ as follows. Chose $b_1^* = 2$ and $b_2^* = -1.0445$ such that, without the lossy encoder $q_c$, the distortions $D_1$ and $D_2$ are made equal. Denote the entropy rate of $q_a$ as $R_{1a}$ and that of $q_b$ as $R_2$. Let $b_3^* = 2, b_4^* = -1$, but $b_5^* = 2.9555$ such that $E(r \,|\, q_a(x) = i)$ is approximately zero. By doing this, $b_3^*x + b_4^*t + b_5^*y_{i_0} = r_i$ on the $(x,t)$-plane aligns with the function $t = f(x)$, and thus the remaining rate $R_1 - R_{1a}$ is used by $q_c$ to improve $D_3$, but $D_1$ and $D_2$ are not further improved. Since $q_a$ and $q_b$ are both operating on high resolution, assuming $R_1 - R_{1a}$ is also high, then $q_c$ partitions each $x \in C_s(j) \cap C_{xr}(i_a, i_c)$ into $2^{R_1 - R_{1a}}$ uniform segments, thus improving $D_3$ by a factor of $2^{-2(R_1 - R_{1a})}$.

Using this construction, we can achieve a balanced high-resolution operating point of $(R_1, R_1, D_1, D_1, D_3)$ without time-sharing, where

$$D_1 = \frac{2\pi e}{12}2^{-2R_{1a}}\sigma_x^2$$

and

$$D_3 \approx 0.8974 \cdot \frac{2\pi e\sigma_x^2}{48}2^{-2(2R_1 - R_{1a})}.$$

Thus, when $\sigma_x^2 \gg D_1 = D_2 \gg D_3$, the central and side distortion product is 2.596 dB away from the information-theoretic distortion product; in fact, for any sources with a smooth probability density function (pdf), this quantizer can achieve a granular distortion 2.596 away from the Shannon outer bound which is tight at high resolution. This is a better upper bound than the best known upper bound of the granular distortion using scalar quantization, which is 2.67 dB away from the information-theoretic distortion product [23], which was previously derived in [23] using MDSQ [20], [21] with systematic optimization of quantization thresholds.

Though we only considered scalar quantizers in this section, the extension of the coding schemes to undithered vector quantization is straightforward. It can be seen that in Figs. 10 and 12, the encoders do not need to introduce the index assignment component, which is the key component in the MDSQ and MDLVQ framework [20], [28], and thus, we circumvent the difficulty of designing the optimal index assignment. It is also worth mentioning that even when the distortions and rates are balanced, the proposed quantization system is not completely symmetric; this is quite different from MDSQ, and the effect of such asymmetry on practical image/video coding system is unknown.

## VII. Conclusion

We proposed a lattice quantization scheme which can achieve the whole Gaussian MD rate–distortion region. The proposed scheme is universal in the sense that it only needs the information of the first- and second-order statistics of the source; furthermore, the scheme is asymptotically optimal for all smooth sources at high resolution.

Our results, along with a recent work by Erez and Zamir [68], consolidate the link between MMSE estimation and lattice coding (quantization), or in a more general sense, the connection between Wiener and Shannon theories as illuminated by Forney [69], [70].

Although the linear MMSE structure is optimal in achieving the Gaussian MD rate–distortion region as the dimension of the (optimal) lattice quantizers goes to infinity, it is not optimal for finite-dimensional lattice quantizers since the distribution of quantization errors is no longer Gaussian. Using nonlinear structure to exploit the higher order statistics may result in better performance. We also want to point out that many other standard quantizers can be easily incorporated in our framework, although in this paper, to facilitate the theoretic analysis, we have mainly focused on the lattice quantizers.

## Appendix I
## Gram–Schmidt Orthogonalization for Random Vectors

Let $\mathcal{H}_v$ denote the set of all $n$-dimensional,[8] finite-covariance-matrix, zero-mean, real random (column) vectors. $\mathcal{H}_v$ becomes a Hilbert space under the inner product mapping

$$\langle \boldsymbol{X}, \boldsymbol{Y} \rangle = \mathbb{E}(\boldsymbol{X}\boldsymbol{Y}^T) : \mathcal{H}_v \times \mathcal{H}_v \to \mathcal{R}^{n\times n}.$$

For $\boldsymbol{X}_1^M = (\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M)^T$ with $\boldsymbol{X}_i \in \mathcal{H}_v, i = 1, \ldots, M$, the Gram–Schmidt orthogonalization proceeds as follows:

$$\boldsymbol{B}_1 = \boldsymbol{X}_1$$
$$\boldsymbol{B}_i = \boldsymbol{X}_i - \hat{K}_{i-1}\boldsymbol{X}_1^{i-1}, \qquad i = 2, \ldots, M$$

where $\hat{K}_{i-1} \in \mathcal{R}^{n\times(i-1)n}$ is a matrix satisfying $\hat{K}_{i-1}K_{\boldsymbol{X}_1^{i-1}} = K_{\boldsymbol{X}_i\boldsymbol{X}_1^{i-1}}$. When $K_{\boldsymbol{X}_1^{i-1}}$ is invertible, we have $\hat{K}_{i-1} = K_{\boldsymbol{X}_i\boldsymbol{X}_1^{i-1}}K^{-1}_{\boldsymbol{X}_1^{i-1}}$. Here $K_{\boldsymbol{X}_1^{i-1}}$ is the covariance matrix of $(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_i)^T$ and

$$K_{\boldsymbol{X}_i\boldsymbol{X}_1^{i-1}} = \mathbb{E}[\boldsymbol{X}_i(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{i-1})^T].$$

Again, a sequential quantization system can be constructed with $\boldsymbol{X}_1$ as the input to generate a zero-mean random vector $\tilde{\boldsymbol{X}}_1^M = (\tilde{\boldsymbol{X}}_1, \ldots, \tilde{\boldsymbol{X}}_M)^T$ whose covariance matrix is also $K_{\boldsymbol{X}_1^M}$. Assume $K_{\boldsymbol{B}_i} = \mathbb{E}\boldsymbol{B}_i\boldsymbol{B}_i^T$ is nonsingular for $i = 2, \ldots, M$. Let $Q_{i,n}(\cdot)$ be an $n$-dimensional lattice quantizer, $i = 1, 2, \ldots, L-1$. The dither $\boldsymbol{Z}_i$ is an $n$-dimensional random vector, uniformly distributed over the basic cell of $Q_{i,n}, i = 1, 2, \ldots, M-1$. Suppose $(\boldsymbol{X}_1, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{M-1})$ are independent, and $\mathbb{E}\boldsymbol{Z}_i\boldsymbol{Z}_i^T = K_{\boldsymbol{B}_i}, i = 1, 2, \ldots, M$. Define

$$\tilde{\boldsymbol{X}}_1 = \boldsymbol{X}_1 \tag{26}$$
$$\tilde{\boldsymbol{X}}_i = Q_{i-1,n}\left(\hat{K}_{i-1}\tilde{\boldsymbol{X}}_1^{i-1} + \boldsymbol{Z}_{i-1}\right) - \boldsymbol{Z}_{i-1}, \quad i = 2, \ldots, M. \tag{27}$$

It is easy to show that $\boldsymbol{X}_1^M$ and $\tilde{\boldsymbol{X}}_1^M$ have the same covariance matrix.

Suppose $K_{\boldsymbol{B}_i}$ is singular for some $i$, say $K_{\boldsymbol{B}_i}$ is of rank $k$ with $k < n$. For this type of degenerate case, the quantization operation should be carried out in the nonsingular subspace of $K_{\boldsymbol{B}_i}$. Let $K_{\boldsymbol{B}_i} = U\Lambda U^T$ be the eigenvalue decomposition of $K_{\boldsymbol{B}_i}$. Without loss of generality, assume $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_k, 0, \ldots, 0\}$, where $\lambda_i < 0$ for all $i = 1, 2, \ldots, k$. Define $\Lambda_k = \mathrm{diag}\{\lambda_1, \ldots, \lambda_k\}$. Now replace the $n$-dimensional quantizer $Q_{i-1,n}(\cdot)$ in (27) by a $k$-dimensional quantizer $Q_{i-1,k}(\cdot)$ and replace the dither $\boldsymbol{Z}_{i-1}$ by a dither $\tilde{\boldsymbol{Z}}_{i-1}$ which is a $k$-dimensional random vector, uniformly distributed over the basic cell of $Q_{i-1,k}$ with $\mathbb{E}\tilde{\boldsymbol{Z}}_{i-1}\tilde{\boldsymbol{Z}}_{i-1}^T = \Lambda_k$. Let

$$[\tilde{\boldsymbol{X}}_i]_{1,k} = Q_{i-1,k}\left(\left[U^T\hat{K}_{i-1}\tilde{\boldsymbol{X}}_1^{i-1}\right]_{1,k} + \tilde{\boldsymbol{Z}}_{i-1}\right) - \tilde{\boldsymbol{Z}}_{i-1}$$

and we have

$$\tilde{\boldsymbol{X}}_i = U\begin{pmatrix}[\tilde{\boldsymbol{X}}_i]_{1,k} \\ \left[U^T\hat{K}_{i-1}\tilde{\boldsymbol{X}}_1^{i-1}\right]_{k+1,n}\end{pmatrix}$$

where $[U^T\hat{K}_{i-1}\tilde{\boldsymbol{X}}_1^{i-1}]_{1,k}$ is a column vector containing the first $k$ entries of $U^T\hat{K}_{i-1}\tilde{\boldsymbol{X}}_1^{i-1}$ and $[U^T\hat{K}_{i-1}\tilde{\boldsymbol{X}}_1^{i-1}]_{k+1,n}$ is a column vector that contains the remaining entries of $U^T\hat{K}_{i-1}\tilde{\boldsymbol{X}}_1^{i-1}$.

## Appendix II
## Linear Estimator Coefficients and Statistics of the Innovation Process

$$\alpha_i = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_i}^2}, \qquad i = 1, 2$$

$$\beta_i = \frac{\sigma_X^2\sigma_{T_1}\sigma_{T_2}}{\sigma_{T_i}\left(\sigma_{T_1} + \sigma_{T_2}\right)\left(\sigma_X^2 + \sigma_{T_0}^2\right)}, \qquad i = 1, 2$$

$$a_1 = \frac{\sigma_{T_1}^2 + \sigma_{T_1}\sigma_{T_2}}{\sigma_{T_0}^2 + \sigma_{T_1}^2} \tag{28}$$

$$a_2 = \frac{\sigma_{T_0}^2 - \sigma_{T_1}\sigma_{T_2}}{\sigma_{T_0}^2 + \sigma_{T_1}^2} \tag{29}$$

$$b_1 = \frac{\sigma_{T_2}^2 + \sigma_{T_3}^2 + \sigma_{T_1}\sigma_{T_2}}{\sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2} \tag{30}$$

$$b_2 = \frac{\sigma_{T_0}^2 - \sigma_{T_1}\sigma_{T_2}}{\sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2} \tag{31}$$

$$b_3 = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2} \tag{32}$$

$$b_4 = \frac{\sigma_X^2 + \sigma_{T_0}^2 - \sigma_{T_1}\sigma_{T_2}}{\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2} \tag{33}$$

$$b_5 = \frac{\sigma_{T_2}^2 + \sigma_{T_3}^2 + \sigma_{T_1}\sigma_{T_2}}{\sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2} \tag{34}$$

$$b_6 = \frac{\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_2}^2}{\sigma_X^2 + \sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2} \tag{35}$$

$$b_7 = \frac{\sigma_{T_3}^2}{\sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2} \tag{36}$$

$$b_8 = \frac{\sigma_{T_3}^2\left(\sigma_{T_0}^2 - \sigma_{T_1}\sigma_{T_2}\right)}{\sigma_{T_0}^2\left(\sigma_{T_1} + \sigma_{T_2}\right)^2 + \sigma_{T_3}^2\left(\sigma_{T_0}^2 + \sigma_{T_1}^2\right)} \tag{37}$$

$$\mathbb{E}B_2^2 = \sigma_{T_0}^2 + \sigma_{T_1}^2 \tag{38}$$

$$\mathbb{E}B_3^2 = \frac{\sigma_{T_0}^2(\sigma_{T_1} + \sigma_{T_2})^2}{\sigma_{T_0}^2 + \sigma_{T_1}^2} \tag{39}$$

$$\mathbb{E}\tilde{B}_2^2 = \sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2 \tag{40}$$

$$\mathbb{E}\tilde{B}_3^2 = \frac{\sigma_{T_0}^2\left(\sigma_{T_1} + \sigma_{T_2}\right)^2 + \sigma_{T_3}^2\left(\sigma_{T_0}^2 + \sigma_{T_1}^2\right)}{\sigma_{T_0}^2 + \sigma_{T_2}^2 + \sigma_{T_3}^2} \tag{41}$$

[8] This condition is introduced just for the purpose of simplifying the notations.

$$\mathbb{E}\bar{B}_2^2=\frac{\sigma_X^2\left(\sigma_{T_0}^2+\sigma_{T_2}^2+\sigma_{T_3}^2\right)}{\sigma_X^2+\sigma_{T_0}^2+\sigma_{T_2}^2+\sigma_{T_3}^2}$$

$$\mathbb{E}\bar{B}_3^2=\frac{\sigma_{T_0}^2\left(\sigma_{T_1}+\sigma_{T_2}\right)^2+\sigma_{T_3}^2\left(\sigma_{T_0}^2+\sigma_{T_1}^2\right)}{\sigma_{T_0}^2+\sigma_{T_2}^2+\sigma_{T_3}^2}$$

$$\begin{aligned}\mathbb{E}\bar{B}_4^2&=\frac{\sigma_{T_3}^2\left(\sigma_X^2+\sigma_{T_0}^2+\sigma_{T_2}^2\right)}{\sigma_X^2+\sigma_{T_0}^2+\sigma_{T_2}^2+\sigma_{T_3}^2}-b_7^2\mathbb{E}\bar{B}_2^2-b_8^2\mathbb{E}\bar{B}_3^2\\&=\frac{\sigma_{T_3}^2\left(\sigma_X^2+\sigma_{T_0}^2+\sigma_{T_2}^2\right)}{\sigma_X^2+\sigma_{T_0}^2+\sigma_{T_2}^2+\sigma_{T_3}^2}\\&\quad-\frac{\sigma_X^2\sigma_{T_3}^4}{\left(\sigma_X^2+\sigma_{T_0}^2+\sigma_{T_2}^2+\sigma_{T_3}^2\right)\left(\sigma_{T_0}^2+\sigma_{T_2}^2+\sigma_{T_3}^2\right)}\\&\quad-\frac{\sigma_{T_3}^4\left(\sigma_{T_0}^2-\sigma_{T_1}\sigma_{T_2}\right)^2}{\left[\sigma_{T_0}^2\left(\sigma_{T_1}+\sigma_{T_2}\right)^2+\sigma_{T_3}^2\left(\sigma_{T_0}^2+\sigma_{T_1}^2\right)\right]\left(\sigma_{T_0}^2+\sigma_{T_2}^2+\sigma_{T_3}^2\right)}.\end{aligned}\tag{42}$$

## Appendix III
## Proof of Theorem 5.3

It is easy to verify that as $D_1, D_2, D_3 \to 0$ along a straight line, we have $\frac{\sigma_{T_0}^2}{D_3} \to 1$, $\frac{\sigma_{T_0}^2+\sigma_{T_i}^2}{D_3} \to 1$, and $\frac{\sigma_{T_i}}{\sqrt{D_i-D_3}} \to 1$, $i=1,2$. Let

$$\sigma_{T_3}^2 \in \left[0, M\left(\frac{D_3}{D_1}(\sqrt{D_1-D_3}+\sqrt{D_2-D_3})^2+D_2\right)\right]$$

where $M$ is a fixed large number. Clearly, $\sigma_{T_3}^2 \to 0$ as $D_1, D_2, D_3 \to 0$.

For the MD quantization scheme shown in Fig. 9, we have

$$\begin{aligned}R_1&=\frac{1}{n}H(Q_{2,n}^*(b_1^*\boldsymbol{X}+b_2^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{Z}_2^{*n})\,|\,\boldsymbol{Z}_2^*)\\&=\frac{1}{n}h(\boldsymbol{X}+b_2^*\boldsymbol{N}_1^*+\boldsymbol{N}_2^*)-\frac{1}{n}h(\boldsymbol{N}_2^*)\\&=\frac{1}{n}h(\boldsymbol{X}+b_2^*\boldsymbol{N}_1^*+\boldsymbol{N}_2^*)-\frac{1}{2}\log\frac{\mathbb{E}\tilde{B}_3^2}{G_n^{\mathrm{opt}}}\\R_2&=\frac{1}{n}H(Q_{1,n}^*(\boldsymbol{X}+\boldsymbol{Z}_1^*)\,|\,\boldsymbol{Z}_1^*)\\&\quad+\frac{1}{n}H(Q_{3,n}^*(b_3^*\boldsymbol{X}+b_4^*\tilde{\boldsymbol{W}}_1+b_5^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{Z}_3^*)\,|\,\boldsymbol{Z}_3^*)\\&=\frac{1}{n}h(\boldsymbol{X}+\boldsymbol{N}_1^*)-\frac{1}{n}h(\boldsymbol{N}_1^*)\\&\quad+\frac{1}{n}h(b_3^*\boldsymbol{X}+b_4^*\tilde{\boldsymbol{W}}_1^n+b_5^*\tilde{\boldsymbol{W}}_2'+\boldsymbol{N}_3^*)-\frac{1}{n}h(\boldsymbol{N}_3^*)\\&\le\frac{1}{n}h(\boldsymbol{X}+\boldsymbol{N}_1^*)-\frac{1}{n}h(\boldsymbol{N}_1^*)\\&\quad+\frac{1}{n}h(b_7\bar{\boldsymbol{B}}_2+b_8\bar{\boldsymbol{B}}_3+\bar{\boldsymbol{B}}_4)-\frac{1}{n}h(\boldsymbol{N}_3^*)\\&=\frac{1}{n}h(\boldsymbol{X}+\boldsymbol{N}_1^*)-\frac{1}{2}\log\frac{\mathbb{E}\tilde{B}_2^2}{G_n^{\mathrm{opt}}}\\&\quad+\frac{1}{2}\log\left[2\pi e\left(b_7^2\mathbb{E}\bar{B}_2^2+b_8^2\mathbb{E}\bar{B}_3^2+\mathbb{E}\bar{B}_4^2\right)\right]-\frac{1}{2}\log\frac{\mathbb{E}\bar{B}_4^2}{G_n^{\mathrm{opt}}}.\end{aligned}$$

Since

$$\frac{1}{n}h(\boldsymbol{X}+\boldsymbol{N}_1^*)=h(X)+o(1)$$

and

$$\frac{1}{n}h(\boldsymbol{X}+b_2^*\boldsymbol{N}_1^*+\boldsymbol{N}_2^*)=h(X)+o(1)$$

as $D_1, D_2, D_3 \to 0$ along a straight line, it follows that

$$\begin{aligned}R_1&=h(X)-\frac{1}{2}\log\frac{\mathbb{E}\tilde{B}_3^2}{G_n^{\mathrm{opt}}}+o(1)\\&=\frac{1}{2}\log\frac{P_X(D_2+\sigma_{T_3}^2)}{D_3(\sqrt{D_1-D_3}+\sqrt{D_2-D_3})^2+\sigma_{T_3}^2D_1}\\&\quad+\frac{1}{2}\log\left(2\pi eG_n^{\mathrm{opt}}\right)+o(1),\\R_2&\le h(X)-\frac{1}{2}\log\frac{\mathbb{E}\tilde{B}_2^2}{G_n^{\mathrm{opt}}}\\&\quad+\frac{1}{2}\log\left[2\pi e\left(b_7^2\mathbb{E}\bar{B}_2^2+b_8^2\mathbb{E}\bar{B}_3^2+\mathbb{E}\bar{B}_4^2\right)\right]\\&\quad-\frac{1}{2}\log\frac{\mathbb{E}\bar{B}_4^2}{G_n^{\mathrm{opt}}}+o(1)\\&=\frac{1}{2}\log\frac{P_X}{D_2+\sigma_{T_3}^2}\\&\quad+\frac{1}{2}\log\frac{D_3(\sqrt{D_1-D_3}+\sqrt{D_2-D_3})^2+\sigma_{T_3}^2D_1}{D_3(\sqrt{D_1-D_3}+\sqrt{D_2-D_3})^2}\\&\quad+\log\left(2\pi eG_n^{\mathrm{opt}}\right)+o(1).\end{aligned}$$

So we have

$$\begin{aligned}R_1+R_2&\le\frac{1}{2}\log\frac{P_X^2}{D_3(\sqrt{D_1-D_3}+\sqrt{D_2-D_3})^2}\\&\quad+\frac{3}{2}\log\left(2\pi eG_n^{\mathrm{opt}}\right)+o(1).\end{aligned}$$

When $\sigma_{T_3}^3=0$, there is no quantization splitting and the quantizer $Q_{3,n}^*(\cdot)$ can be removed. In this case, we have

$$\begin{aligned}R_1&=\frac{1}{2}\log\frac{P_XD_2}{D_3(\sqrt{D_1-D_3}+\sqrt{D_2-D_3})^2}\\&\quad+\frac{1}{2}\log\left(2\pi eG_n^{\mathrm{opt}}\right)+o(1),\\R_2&=\frac{1}{2}\log\frac{P_X}{D_2}+\frac{1}{2}\log\left(2\pi eG_n^{\mathrm{opt}}\right)+o(1).\end{aligned}$$

When $\sigma_{T_3}^3=M[\frac{D_3}{D_1}(\sqrt{D_1-D_3}+\sqrt{D_2-D_3})^2+D_2]$, we have

$$\begin{aligned}R_1&=\frac{1}{2}\log\frac{P_X}{D_1}+\frac{1}{2}\log\left(2\pi eG_{2,n}^*\right)+\epsilon(M)+o(1),\\R_2&=\frac{1}{2}\log\frac{P_XD_1}{D_3(\sqrt{D_1-D_3}+\sqrt{D_2-D_3})^2}\\&\quad+\log\left(2\pi eG_n^{\mathrm{opt}}\right)-\epsilon(M)+o(1)\end{aligned}$$

where $\epsilon(M)\to 0$ as $M\to\infty$. Therefore, the region

$$\begin{aligned}R_1&\ge\frac{1}{2}\log\frac{P_X}{D_1}+\frac{1}{2}\log\left(2\pi eG_n^{\mathrm{opt}}\right)+\epsilon(M)+o(1)\\R_2&\ge\frac{1}{2}\log\frac{P_X}{D_2}+\frac{1}{2}\log\left(2\pi eG_n^{\mathrm{opt}}\right)+o(1)\\R_1+R_2&\ge\frac{1}{2}\log\frac{P_X^2}{D_3(\sqrt{D_1-D_3}+\sqrt{D_2-D_3})^2}\\&\quad+\frac{3}{2}\log\left(2\pi eG_n^{\mathrm{opt}}\right)+o(1)\end{aligned}$$

is achievable.

By symmetry, the region

$$\begin{aligned} R_1 &\geq \frac{1}{2}\log\frac{P_X}{D_1} + \frac{1}{2}\log\left(2\pi e G_n^{\text{opt}}\right) + o(1) \\ R_2 &\geq \frac{1}{2}\log\frac{P_X}{D_2} + \frac{1}{2}\log\left(2\pi e G_n^{\text{opt}}\right) + \epsilon(M) + o(1) \\ R_1 + R_2 &\geq \frac{1}{2}\log\frac{P_X^2}{D_3(\sqrt{D_1 - D_3} + \sqrt{D_2 - D_3})^2} \\ &\quad + \frac{3}{2}\log\left(2\pi e G_n^{\text{opt}}\right) + o(1) \end{aligned}$$

is achievable via the other form of quantization splitting. The desired result follows by combining these two regions and choosing $M$ large enough.

## Appendix IV
## Proof of Theorem 5.4

We shall only give a heuristic argument here. The rigorous proof is similar to that of Theorem 3 in [38] and thus is omitted.

It is well known that the distribution of the quantization noise converges to a white Gaussian distribution in the divergence sense as the dimension of the optimal lattice becomes large [38]. So we can approximate $\boldsymbol{N}_i^*$ by $\boldsymbol{N}_i^G$, where $\boldsymbol{N}_i^G$ is a zero-mean Gaussian vector with the same covariance as that of $\boldsymbol{N}_i^*$, $i = 1, 2, 3$. Therefore, for large $n$, we have

$$\begin{aligned} &\frac{1}{n}h(\boldsymbol{X} + \boldsymbol{N}_1^*) - \frac{1}{n}h(\boldsymbol{N}_1^*) \\ &\quad \approx \frac{1}{n}h(\boldsymbol{X} + \boldsymbol{N}_1^G) - \frac{1}{n}h(\boldsymbol{N}_1^G) \\ &\quad = h\left(X + N_1^G\right) - h\left(N_1^G\right), \\ &\quad h\left(b_1^*\boldsymbol{X} + b_2^*\tilde{\boldsymbol{W}}'_2 + \boldsymbol{N}_2^*\right) - \frac{1}{n}h\left(\boldsymbol{N}_2^*\right) \\ &\quad = \frac{1}{n}h\left(\boldsymbol{X} + b_2^*\boldsymbol{N}_1^* + \boldsymbol{N}_2^*\right) - \frac{1}{n}h\left(\boldsymbol{N}_2^*\right) \\ &\quad \approx \frac{1}{n}h\left(\boldsymbol{X} + b_2^*\boldsymbol{N}_1^G + \boldsymbol{N}_2^G\right) - \frac{1}{n}h\left(\boldsymbol{N}_2^G\right) \\ &\quad = h\left(X + b_2^*N_1^G + N_2^G\right) - h\left(N_2^G\right) \end{aligned}$$

and

$$\begin{aligned} &\frac{1}{n}h\left(b_3^*\boldsymbol{X} + b_4^*\tilde{\boldsymbol{W}}_1^n + b_5^*\tilde{\boldsymbol{W}}'_2 + \boldsymbol{N}_3^*\right) - \frac{1}{n}h\left(\boldsymbol{N}_3^*\right) \\ &\quad = \frac{1}{n}h\left((b_3^* + b_1^*b_4^* + b_2^*b_4^* + b_5^*)\boldsymbol{X}\right. \\ &\qquad \left. + (b_2^*b_4^* + b_5^*)\boldsymbol{N}_1^* + b_4^*\boldsymbol{N}_2^* + \boldsymbol{N}_3^*\right) - \frac{1}{n}h\left(\boldsymbol{N}_3^*\right) \\ &\quad \approx \frac{1}{n}h\left((b_3^* + b_1^*b_4^* + b_2^*b_4^* + b_5^*)\boldsymbol{X}\right. \\ &\quad \left. + (b_2^*b_4^* + b_5^*)\boldsymbol{N}_1^G + b_4^*\boldsymbol{N}_2^G + \boldsymbol{N}_3^G\right) - \frac{1}{n}h\left(\boldsymbol{N}_3^G\right) \\ &\quad = h\left((b_3^* + b_1^*b_4^* + b_2^*b_4^* + b_5^*)X\right. \\ &\qquad \left. + (b_2^*b_4^* + b_5^*)N_1^G + b_4^*N_2^G + N_3^G\right) - h\left(N_3^G\right). \end{aligned}$$

## Appendix V
## The Calculation of Scalar Operating Point Using Successive Quantization

Observe in Fig. 13(c) that the value $a_2$ is slightly different from $-1$, such that a portion of the $C_s$ cells consist of three length-$\frac{1}{2}\Delta_b$ intervals which are approximately $\frac{-a_2}{2}\Delta_a$ apart (denote the set of this first class of cells as $C'_s$), while the other $C_s$ cells consist of only two length-$\frac{1}{2}\Delta_b$ intervals which are also $\frac{-a_2}{2}\Delta_a$ apart (denote the set of this second class of cells as $C''_s$); the ratio between the Lebesgue measure of these two sets is a function of $a_2$, which is approximately $\frac{-3-3a_2}{4+3a_2}$. Here we again ignore the cells $C_s$ whose constituent segments are at the border of $q_a$ partition cells, which is a negligible portion when $\Delta_a \gg \Delta_b$. The average distortion for each first class cell $C_s$ is approximately $\frac{2}{3}(\frac{-a_2}{2}\Delta_a)^2$, while the average distortion for each second class cell $C_s$ is approximately $(\frac{1}{2}\cdot\frac{-a_2}{2}\Delta_a)^2$. Thus, the distortion $D_2$ can be approximated as

$$\begin{aligned} D_2 \approx &(-3 - 3a_2)\cdot\frac{2}{3}\left(\frac{-a_2}{2}\Delta_a\right)^2 + (4 + 3a_2)\left(\frac{1}{2}\frac{-a_2}{2}\Delta_a\right)^2 \\ = &\frac{-1}{16}(5a_2 + 4)a_2^2{\Delta_a}^2. \end{aligned} \tag{43}$$

Notice that $-3 - 3a_2 + 4 + 3a_2 = 1$; thus, $(-3 - 3a_2)$ is the probability of a random $C_s$ cell being a first class cell. Letting $D_1 = D_2 = \frac{1}{12}{\Delta_a}^2$, we can solve for $a_2$; the only real solution to this equation is $a_2 = -1.0445$. The distortion $D_3$ is approximately $\frac{1}{12}(\frac{1}{2}\Delta_b)^2$, by using an almost uniform partition of stepsize $\frac{1}{2}\Delta_b$. To approximate the entropy rate for $q_b$, consider the rate contribution from the first class $C_s$ cells, namely

$$\begin{aligned} R'_2 &= -\sum_{C_s(i)\in C'_s} p\left(q_2^{-1}(i)\right)\frac{3}{2}\Delta_b \log_2\left(p\left(q_2^{-1}(i)\right)\frac{3}{2}\Delta_b\right) \\ &\approx (3 + 3a_2)\log_2\left(\frac{3}{2}\Delta_b\right) \\ &\quad - \sum_{C_s(i)\in C'_s} p\left(q_2^{-1}(i)\right)\frac{3}{2}\Delta_b \log_2\left(p\left(q_2^{-1}(i)\right)\right) \end{aligned} \tag{44}$$

where $p(x)$ is the pdf of the source. Similarly, the rate contribution from the second class $C_s$ cells is

$$\begin{aligned} R''_2 &= -\sum_{C_s(i)\in C''_s} p\left(q_2^{-1}(i)\right)\frac{2}{2}\Delta_b \log_2\left(p\left(q_2^{-1}(i)\right)\frac{2}{2}\Delta_b\right) \\ &\approx -(4 + 3a_2)\log_2\left(\frac{2}{2}\Delta_b\right) \\ &\quad - \sum_{C_s(i)\in C''_s} p\left(q_2^{-1}(i)\right)\frac{2}{2}\Delta_b \log_2\left(p\left(q_2^{-1}(i)\right)\right). \end{aligned} \tag{45}$$

Thus, the rate $R_2$ can be approximated as

$$\begin{aligned} R_2 &\approx R'_2 + R''_2 \\ &\approx -\log_2(\Delta_b) + (3 + 3a_2)\log_2\left(\frac{3}{2}\right) \\ &\quad - \sum_{C_s(i)\in C'_s} p\left(q_2^{-1}(i)\right)\frac{3}{2}\Delta_b \log_2\left(p\left(q_2^{-1}(i)\right)\right) \\ &\quad - \sum_{C_s(i)\in C''_s} p\left(q_2^{-1}(i)\right)\frac{2}{2}\Delta_b \log_2\left(p\left(q_2^{-1}(i)\right)\right). \end{aligned} \tag{46}$$

When $q_a(\cdot)$ is high resolution, $p(q_2^{-1}(i))$ is approximately equal to $p(x)$, for any $x \in C_s(i)$, and thus equal to $p(q_3^{-1}(i,\cdot))$.

Using this approximation and taking $\frac{1}{2}\Delta_b$ as $\delta x$, the last two terms in (46) can be approximated by an integral, which is in fact $h(p)$, the differential entropy of the source. It follows that

$$\begin{aligned} R_2 &\approx R_2' + R_2'' \\ &\approx -\log_2(\Delta_b) + (3+3a_2)\log_2\left(\frac{3}{2}\right) + h(p) \end{aligned} \quad (47)$$

where $h(p) = \frac{1}{2}\log(2\pi e\sigma_x^2)$ for the Gaussian source. Thus,

$$D_3 \approx \frac{1}{12}\left(\frac{1}{2}\Delta_b\right)^2 \approx 0.8974 \cdot \frac{2\pi e\sigma_x^2}{48} 2^{-2R_2}.$$

## ACKNOWLEDGMENT

The authors wish to thank Prof. Ram Zamir and two anonymous reviewers for their valuable comments and suggestions.

## REFERENCES

[1] H. Witsenhausen, "On source networks with minimal breakdown degradation," *Bell Syst. Tech. J.*, vol. 59, no. 6, pp. 1083–1087, July/Aug. 1980.

[2] J. Wolf, A. Wyner, and J. Ziv, "Source coding for multiple descriptions," *Bell Syst. Tech. J.*, vol. 59, no. 8, pp. 1417–1426, Oct. 1980.

[3] L. Ozarow, "On a source coding problem with two channels and three receivers," *Bell Syst. Tech. J.*, vol. 59, no. 10, pp. 1909–1921, Dec. 1980.

[4] H. S. Witsenhausen and A. D. Wyner, "Source coding for multiple descriptions, II: A binary source," *Bell Syst. Tech. J.*, vol. 60, pp. 2281–2292, Dec. 1981.

[5] A. A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 851–857, Nov. 1982.

[6] R. Ahlswede, "The rate-distortion region for multiple descriptions without excess rate," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 6, pp. 721–726, Nov. 1985.

[7] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 4, pp. 502–521, Jul. 1987.

[8] H. S. Witsenhausen and A. D. Wyner, "On team guessing with independent information," *Math. Oper. Res.*, vol. 6, pp. 293–304, May 1981.

[9] T. Berger and Z. Zhang, "Minimum breakdown degradation in binary source coding," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 6, pp. 807–814, Nov. 1983.

[10] R. Ahlswede, "On multiple descriptions and team guessing," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 4, pp. 543–549, Jul. 1986.

[11] R. Zamir, "Gaussian codes and Shannon bounds for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2629–2635, Nov. 1999.

[12] F. W. Fu, R. W. Yeung, and R. Zamir, "On the rate-distortion region for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 48, no. 7, pp. 2012–2021, Jul. 2002.

[13] H. Feng and M. Effros, "On the rate loss of multiple description source codes," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 671–683, Feb. 2005.

[14] L. Lastras-Montaño and V. Castelli, "Near sufficiency of random coding for two descriptions," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 681–695, Feb. 2006.

[15] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.

[16] S. S. Pradhan, R. Puri, and K. Ramchandran, "$n$-channel symmetric multiple descriptions—Part I: $(n, k)$ source–channel erasure codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, pp. 47–61, Jan. 2004.

[17] P. Ishwar, R. Puri, S. S. Pradhan, and K. Ramchandran, "On compression for robust estination in sensor networks," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun.Jul. 2003, p. 193.

[18] J. Chen and T. Berger, "Robust distributed source coding," *IEEE Trans. Inf. Theory*, submitted for publication.

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

[20] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 821–834, May 1993.

[21] V. A. Vaishampayan and J. Domaszewicz, "Design of entropy-constrained multiple-description scalar quantizers," *IEEE Trans. Inf. Theory*, vol. 40, no. 1, pp. 245–250, Jan. 1994.

[22] V. A. Vaishampayan and J. C. Batllo, "Asymptotic analysis of multiple-description quantizers," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 278–284, Jan. 1998.

[23] C. Tian and S. S. Hemami, "Universal multiple description scalar quantizer: Analysis and design," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2089–2102, Sep. 2004.

[24] J. Balogh and J. A. Csirik, "Index assignment for two-channel quantization," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2737–2751, Nov. 2004.

[25] T. Y. Berger-Wolf and E. M. Reingold, "Index assignment for multichannel communication under failure," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2656–2668, Oct. 2002.

[26] N. Gortz and P. Leelapornchai, "Optimization of the index assignments for multiple description vector quantizers," *IEEE Trans. Commun.*, vol. 51, no. 3, pp. 336–340, Mar. 2003.

[27] P. Koulgi, S. L. Regunathan, and K. Rose, "Multiple description quantization by deterministic annealing," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 2067–2075, Aug. 2003.

[28] V. A. Vaishampayan, N. Sloane, and S. Servetto, "Multiple description vector quantization with lattice codebooks: Design and analysis," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1718–1734, Jul. 2001.

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

[30] V. K. Goyal, J. A. Kelner, and J. Kovačević, "Multiple description vector quantization with a coarse lattice," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 781–788, Mar. 2002.

[31] C. Tian and S. S. Hemami, "Optimality and sub-optimality of multiple description vector quantization with a lattice codebook," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2458–2468, Oct. 2004.

[32] Y. Frank-Dayan and R. Zamir, "Dithered lattice-based quantizers for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. 48, no. 1, pp. 192–204, Jan. 2002.

[33] M. T. Orchard, Y. Wang, V. A. Vaishampayan, and A. R. Reibman, "Redundancy rate-distortion analysis of multiple description coding using pairwise correlating transforms," in *Proc. IEEE Int. Conf. Image Processing*, Santa Barbara, CA, Oct. 1997, vol. 1, pp. 608–611.

[34] Y. Wang, M. T. Orchard, and A. R. Reibman, "Optimal pairwise correlating transforms for multiple description coding," in *Proc. Int. Conf. Image Processing*, Chicago, IL, Oct. 1998, pp. 679–683.

[35] S. S. Pradhan and K. Ramchandran, "On the optimality of block orthogonal transforms for multiple description coding of Gaussian vector sources," *IEEE Signal Process. Lett.*, vol. 7, no. 4, pp. 76–78, Apr. 2000.

[36] V. K. Goyal and J. Kovačević, "Generalized multiple description coding with correlating transforms," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2199–2224, Sep. 2001.

[37] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizer," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 428–436, Mar. 1992.

[38] ——, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, Jul. 1996.

[39] ——, "Information rates of pre/post filtered dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1340–1353, Sep. 1996.

[40] R. Zamir, S. Shamai (Shitz), and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, Jun. 2002.

[41] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*. Upper Saddle River, NJ: Prentice-Hall, 2000.

[42] T. Guess and M. K. Varanasi, "An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 173–187, Jan. 2005.

[43] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.

[44] A. A. El Gamal and E. van der Meulen, "A proof of Martons coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–122, Jan. 1981.

[45] X. Zhang, J. Chen, S. B. Wicker, and T. Berger, "Successive coding in multiuser information theory," *IEEE Trans. Inf. Theory*, submitted for publication.

[46] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Probl. Control Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.

[47] M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.
[48] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications, CISM Courses and Lectures*, G. Longo, Ed. Vienna/New York: Springer-Verlag, 1978, vol. 229, pp. 171–231.
[49] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
[50] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Elec. Eng., Cornell Univ., Ithaca, NY, May 1978.
[51] T. Berger, K. Housewright, J. Omura, S. Y. Tung, and J. Wolfowitz, "An upper bound on the rate-distortion function for source coding with partial side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 6, pp. 664–666, Nov. 1979.
[52] R. Ahlswede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Information Theory*. Budapest, Hungary: Hungarian Acad. Sci., 1973, pp. 23–52.
[53] H. Liao, "Multiple access channels," Ph.D. dissertation, Dept. Elec. Eng., Univ. Hawaii, Honolulu, 1972.
[54] B. Rimoldi and R. Urbanke, "Asynchronous Slepian-Wolf coding via source-splitting," in *Proc. IEEE Int. Symp. Information Theory*, Ulm, Germany, Jun./Ju. 1997, p. 271.
[55] T. P. Coleman, A. H. Lee, M. Médard, and M. Effros, "Low-complexity approaches to Slepian-Wolf near-lossless distributed data compression," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3546–3561, Aug. 2006.
[56] J. Chen and T. Berger, "Successive Wyner-Ziv coding scheme and its application to the quadratic gaussian CEO problem," *IEEE Trans. Inf. Theory*, submitted for publication.
[57] A. B. Carleial, "On the capacity of multiple-terminal communication networks," Ph.D. dissertation, Stanford Univ., Stanford, CA, Aug. 1975.
[58] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple- access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 364–375, Mar. 1996.
[59] A. J. Grant, B. Rimoldi, R. L. Urbanke, and P. A. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 873–890, Mar. 2001.
[60] B. Rimoldi, "Generalized time sharing: A low-complexity capacity-achieving multiple- access technique," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2432–2442, Sep. 2001.
[61] R. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[62] H. Feng and M. Effros, "On the achievable region for multiple description source codes on gaussian sources," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 195.
[63] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.
[64] R. Zamir, private communication.
[65] C. Tian and S. S. Hemami, "A new class of multiple description scalar quantizers and its application to image coding," *IEEE Signal Process. Lett.*, vol. 12, no. 4, pp. 329–332, Apr. 2005.
[66] S. D. Servetto, K. Ramchandran, V. A. Vaishampayan, and K. Nahrstedt, "Multiple description wavelet based image coding," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 813–826, May 2000.
[67] W.-Y. Chan, S. Gupta, and A. Gersho, "Enhanced multistage vector quantization by joint codebook design," *IEEE Trans. Commun.*, vol. 40, no. 11, pp. 1693–1697, Nov. 1992.
[68] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2293–2314, Oct. 2004.
[69] G. D. Forney Jr., "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proc. 41st Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2003, pp. 430–439.
[70] ——, "Shannon meets Wiener II: On MMSE estimation in successive decoding schemes," in *Proc. 42nd Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2004.